BRIEF TEST OF ADULT COGNITION BY TELEPHONE (BTACT)

Technical Report Psychology Department Brandeis University January 21, 2005

Patricia A. Tun Memory and Cognition Laboratory, Volen National Center for Complex Systems, and Margie E. Lachman Life-span Development Laboratory

The Brief Test of Adult Cognition by Telephone (BTACT) was developed in response to the need for a brief but reliable test that covers important areas of adult cognition. It can be administered easily over the telephone to adults from a broad range of ages and educational backgrounds. The BTACT is being administered by telephone in 2003-2005 to the MIDUS II (Midlife in the U.S.) national sample of 7,000 adults aged 35-85 (the longitudinal follow-up study of the John D. and Catherine T. MacArthur Research Network on Successful Midlife Development (MIDMAC) Survey of Midlife in the United States (MIDUS)). This research is supported by the National Institute on Aging, Grant No. PO1 AG20166. Also, we appreciate the support of Arthur Wingfield and the Memory and Cognition Laboratory, and the Lifespan Development Laboratory, as well as the technical assistance of Elisabeth Fracalossi and Douglas van Hollen in developing the protocol for scoring.

Test of Adult Cognition by Telephone (BTACT)

Background

The Brief Test of Adult Cognition by Telephone (BTACT) was designed to tap areas of cognitive function that are sensitive to the effects of aging. These areas include episodic verbal memory (Craik & Anderson, 1999), working memory span and executive function (Baddeley, 1986; 1996), reasoning (Miller & Lachman, 2000; Schaie, 1996), and speed of processing (Meyerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, 1996; Verhaeghen & Salthouse, 1997). The tests include the recording of response latencies, which afford a measure of speed of processing. The use of latency data adds an extra dimension to the cognitive measures, providing greater sensitivity to subtle individual differences in speed of processing that may not be revealed by accuracy measures alone (e.g. Salthouse, 1996).

The proposed cognitive battery was informed by the results of previous telephone interviews (Herzog and Wallace, 1997*,* Kawas, Karagiozis, Resau, Corrada, & Brookmeyer, 1995; Nesselroade, Pederson, McClearn, Plomin, & Bergeman, 1988) that have been used successfully in surveys with older adults (see Lachman & Spiro, 2002). Other cognitive batteries such as the Mini Mental Status Exam (MMSE) have been successfully adapted for telephone administration. The Telephone Interview for Cognitive Status (Brandt, Spencer, & Folstein, 1988), adapted from the MMSE, has been used successfully as a screening instrument for dementia. The large-scale HRS/AHEAD study (Herzog, Rodgers, & Kulka, 1983; Herzog and Wallace, 1997) demonstrated the feasibility of a telephone survey of cognitive function in adults over the age of 70. Our battery incorporates similar tests from the domains that are affected in normal aging (verbal memory and working memory), and draws upon the methods developed by the AHEAD study for handling standardization of testing, non-response, and missing data. Importantly, the AHEAD study found no significant differences in performance between respondents tested by telephone and face-to-face assessments (Herzog & Rodgers, 1998). Also, the TELECOG (Tennestedt, Lachman, & Salthouse, 1999), a computerized telephone test that uses voice recognition to assess memory and attentional switching in adults, has shown similar performance in person and over the telephone.

The primary contribution of the BTACT has been the development of a brief and reliable test battery that can be administered by lay interviewers over the telephone in survey research, in order to assess individual differences in cognition among well-functioning middle-aged and older adults. Our selection of tests was motivated by the desire to choose cognitive tests that would (1) assess a range of cognitive abilities that are of particular interest in terms of current cognitive aging theory, and (2) be sensitive to normal age-related change in midlife and later life, based on empirical studies. In addition, the tests should (3) have established psychometric properties, (4) be easily administered in a telephone interview by lay interviewers, and (5) be administered in a short period of time. We also were interested in including measures of speed of processing and reaction time sensitive enough to detect age and health-related differences in broad community samples.

To this end we have adapted tests that will provide estimates of episodic verbal memory, working memory span, executive function, reasoning, and speed of processing. Our measures focus on the more biologically-based fluid intelligence (mechanics), which is driven by basic neurological processes, rather than crystallized intelligence (pragmatics), which is more heavily influenced by education (Baltes, Staudinger, & Lindenberger, 2000; Willis & Schaie, 1999). We have carried out pilot work on adults ranging in age from 20 to 85 confirming that these measures can be used effectively over the telephone with this age range, in less than 20 minutes.

Components of the BTACT

Verbal Memory – Immediate

Immediate and delayed episodic memory for verbal materials are tested using free recall of a word list drawn from the Rey Auditory-Verbal Learning Test (RAVLT; Rey, 1964; Lezak, 1995;
E.M. Taylor, 1959). Previous research has demonstrated recall to be sensitive to age across the lifespan (Geffen, et. al., 1990), including differences in middle-age and later life, as well as effects of gender and education (Bleecker, et al., 1988; Ivnik, et al., 1990). We administer one trial of the test for immediate recall, with a second delayed trial at the end of the session that permits us to assess forgetting. The participant is instructed to listen carefully to a list of 15 words read aloud at a rate of one second per word, and then to recall as many words as possible. A maximum of one minute is allowed for recall. If the participant stops before the minute is up, the interviewer gives one prompt before discontinuing the test. No feedback is given regarding correct responses, repetitions, or errors.

In scoring the accuracy measure, one point is given for each correct response. Also noted are repetitions of words previously reported, and intrusions of words that were not on the study list, which can provide an estimate of failures of source monitoring (Tun, 1989). As an auxillary measure of interest, it is also possible to measure the total time taken for recall, calculated from the onset of the last word on the study list to the onset of the last correct item reported by the participant. (See Appendix A for procedures in calculating latencies). This total time, divided by the number of items correctly recalled, can provide an estimate of recall efficiency, in terms of seconds per word recalled (Tun, 1989).

Previous authors have reported retest reliability on the five-trial version of the RAVLT to be on the order of .55 (Spreen & Strauss, 1991) after a one-year interval. Reliability for word list recall on the WMS is .79-.80 for ages 16-89. Factor analysis has shown the RAVLT to have a verbal learning and memory loading with the verbal subtests of the Wechsler Memory Scale (Spreen & Strauss, 1991).

Working Memory Span

We also obtain an estimate of working memory, a system responsible for both holding active material in awareness (storage) and manipulating information (Baddeley, 1986; 1996) that shows declines in adulthood (e.g. Baddeley, 1986; Miyake & Shah, 1999; Reuter-Lorenz, et al. 2001; Wingfield, Tun, & Rosen, 1995). Working memory is believed by current theorists (e.g.

Engle, Tuholski, Laughlin, & Conway, 1999) to be associated with fluid intelligence. Our measure of working memory capacity will be the backward digit span test from the WAIS-III (1997), which has been standardized with reliability coefficients from .87 to .93, and is significantly correlated with longer measures of working memory such as the reading span (Waters & Caplan, 2003). Backward digit span performance has been reported to show estrogenic effects (Carlson & Sherwin, 1998) as well as age-related changes (WMS, 1997).

For this span measure, the listener hears increasingly longer series of digits, ranging from two to eight digits, and attempts to repeat them in the reverse order from which they were heard. There are two chances to complete each level, and the score is the longest string that is repeated exactly in reverse order. The experimenter reads aloud each set of digits in list intonation, at a rate of one per second, beginning with a set size of two digits and progressing to eight digits. The test is discontinued when the participant misses both trials at a set size. The accuracy score is the largest set size that was correctly reproduced.

Verbal Fluency

The BTACT includes a test of category fluency, which is commonly used as an index of executive function (Kozora & Cullum, 1995; Lezak, 1995). The central executive component of working memory is especially sensitive to age-related changes in adulthood, (Baddeley, 1996; Mayr & Kliegl, 2000) and is linked with frontal lobe function (e.g. Baldo, Shimamura, Delis, Kramer, & Kaplan, 2001; Butler, Rorsman, Hill & Tuma, 1993; Grady & Craik, 2000; West, 1996). We will present a classic verbal fluency test, which requires one to rapidly generate new members of a class of words from a semantic category (Drachman & Leavitt, 1972), while monitoring performance to avoid repetitions. Verbal fluency tests have been demonstrated to be sensitive to aging effects (Bryan & Luszcz, 2000a, 2000b; Rabbit, 1999; Tombaugh, Kozak, & Rees, 1999), as well as education (Ruff, Light, Parker, & Levin, 1996), gender (Auriacombe, et al., 2001), and estrogenic effects (Barrett-Connor & Kritz-Silverstein, 1993), but to be stable across ethnic groups (Kempler, Teng, Dick, Taussig, & Daviss, 1998).

Respondents are given one minute to generate as many items as possible from the category "animals." The score is the number of unique words produced in 60 seconds. We also consider the number of repetitions of the same word, or intrusion errors. In addition to the number of unique words produced, we have found it useful to do a qualitative analysis of responses, in which repetitions of the same word may show age-related deficits in monitoring output (Perlmuter, et al., 1987). Verbal fluency protocols may also be analyzed in terms of clustering patterns (Troyer, et al., 1997), although Salthouse, Atkinson & Berish (2003) recently reported similar negative correlations with age between the number of unique words and the number of clusters, and no correlation with age for the number of words per cluster. Bryan et al. (1997) have reported test-retest reliability for category fluency of .73 for young adults and .75 for older adults.

Attention-Switching/Reaction Time

As an additional measure of speed of processing, we have added a simple two-choice response task. The red/green task is a variant of a traditional speeded classification task that has been used by many researchers, (e.g. Kramer, Hahn, & Gopher, 1999; Rogers & Monsell, 1995; Salthouse, T., 2000; Salthouse, Fristoe, McGuthry & Hambrick, 1998; Tennestedt, Lachman, & Salthouse, 1999). This speeded test allows us to assess simple attention and reaction time in blocked conditions, as well as task-switching in an alternating condition. Task-switching has been shown to be especially effortful for older adults (e.g. Kramer et al., 1999; Kray & Lindenberger, 2000), who show more extensive frontal lobe activation than young adults during such tasks (e.g. DiGirolamo et al., 2001), but little is known about functioning in midlife.

The test involves baseline measures that give an estimate of processing speed under simple and complex conditions, as well as an alternating task that assesses executive functions including task-switching and inhibitory control. The simple form of the task ("normal") requires one of two familiar responses: saying "stop" when one hears the spoken word "red," and "go" to the word "green". The "reverse" condition requires inhibiting this familiar response and giving the reverse, incompatible response ("stop" to "green", "go" to "red"). Finally, in an alternating condition the participant is given occasional cues to switch back and forth between these two modes of response, providing an estimate of task-switching ability. Such task-switching procedures have been widely studied and found to be sensitive to age, as well as to health and other changes in frontal lobe function (Baddeley, 1996; Cepeda, Kramer, & Gonzalez de Sather, 2001; Kray & Lindenberger, 2000). Participants practice each condition before test trials begin. They receive 20 normal and 20 reverse baseline trials, followed by a mixed block of 32 trials (four sets each of normal and reverse condition, with the number of items in each set varying from four to six to make the switch less predictable).

The measures obtained are both accuracy and latency of response in each condition. Previous studies have reported high rates of accuracy, with errors averaging from .02 (Rogers & Monsell, 1995) to .04 (Salthouse et al., 1998) across the adult age span. We found similar accuracy of performance in Pilot Study II, with error rates on all conditions less than 3-4% across all ages. Reliability is typically high for switching tasks, e,g, Salthouse et al. reported a split half reliability of .94. for their subjects on a similar task. We calculated test-retest reliability for a sample of 36 subjects (see Pilot Study II), and found reliabilities of .66 for the baselines and .77 for the task-switching condition.

For latency analyses we prefer the use of medians to eliminate the effect of outliers, following previous researchers (e.g. Cepeda, et al., 2001). However, we also calculate means and standard deviations in order to assess intra-individual variability of performance (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000; Nesselroade & Salthouse, 2000). Among the measures of interest is an estimate of the "cost" of task-switching at the global level, in which mixed switching trials are compared to the simpler blocked baseline condition; this typically shows age-related increases in cost. In addition, we calculate the local cost within the alternating condition, by comparing the first trial after a switch to subsequent trials in which no switch is required. Finally, a compatibility cost compares normal trials with reverse trials, to assess the effect of incompatible responses (Cepeda et al., 2001).

Reasoning

We also assess abstract reasoning, a classic measure of fluid intelligence that our pilot work found to be sensitive to age differences in midlife as well as to menopausal status. Research findings have consistently reported steady decline across adulthood in comprehension of number series (e.g. Horn & Masunaga, 2000; Salthouse, 1992) and letter series (Miller & Lachman, 2000).

The measure of inductive reasoning is a number series completion task (similar to Salthouse & Prill, 1987). Although in most series test the participant views the entire series at once, similar patterns of performance have been shown for series that were presented successively, one item at a time, as we will present them (Salthouse & Prill, 1987). Participants will be given a series of numbers such as "3, 6, 9, 12, 15", and asked to respond with the item that best continues the sequence (e.g. "18"). The experimenter reads aloud each number in a set in list intonation, instructing the participant to say "okay" as a signal that he/she is ready to go on to the next number. At the end of the set the experimenter will say, "And the next number is...?" as a signal to give the final response.

Each participant receives a total of five series, representing three levels of difficulty. There is one series at the simplest level one, in which all elements are associated with the adjacent elements by the same type of relation (e.g. 3, 6, 9, 12, 15); one problem at the more complex level two, in which the pattern consists of two alternating relations, one applying to the odd-numbered elements and the other to the even-numbered elements (e.g. 63, 91, 65, 94, 67…97); and three problems at level three, in which the adjacent elements are related by a quantity that systematically varies; the pattern is evident only from the more abstract second-order relations (e.g. 2, 4, 7, 11, 16…22). Accuracy scoring is determined by the number of sets correctly completed at each level, for a total of five. Previous reports have noted good reliability on reasoning tasks (Schaie, 1996) although the tests had larger number of items.

Speed of processing

As a measure of speed of processing we have included a backward counting task. The participant is given 30 seconds to count backwards from 100 by ones, as quickly as possible. The score is the total number of correct numbers reported. We also note the number of errors and skipped numbers.

This test was designed primarily to measure speed of self-generated processing. In previous research with older adults, the HRS/AHEAD study (Herzog & Wallace, 1997) used a backward counting task but only measured accuracy. Participants were required to count backwards from 20 to 10 for 10 consecutive numbers, with a second chance if a respondent made an error. These studies reported 90-97% accuracy on the first try for their older samples of adults. The AHEAD study also reported that backward counting had a correlation of $r=.29$ with serial 7s, their working memory measure, suggesting that the need to keep track also places demands on memory (Ofstedal, McAuley, & Herzog, 2002).

When used in the BTACT with participants from a wide age range, the primary interest is in speed of backward counting rather than accuracy, which should be very high. Good reliability of .95 has been reported for latencies on a similar test of backward counting by threes (Salthouse, Atkinson, & Berish, 2003). Our pilot work showed good test-retest reliability of $r=.87$, $p<.05$ on a sample of 36 participants ranging in age from 24 to 80, (mean age= 55.8, SD= 16.9) tested at a 6 month interval. We also found that this test was quite sensitive to age, with a correlation with age of $r= -.63$, $p<.05$, consistent with previous studies using backward counting tasks (e.g. Maylor & Wing, 1996). .

Verbal memory- delayed

At the end of each session we obtain a measure of delayed episodic memory for the recall list studied earlier. Participants are asked to recall as many words as possible from the original list. The difference between performance on the immediate list and the delayed list provides an estimate of forgetting. Scoring is carried out as for the immediate memory test.

General Procedure

The BTACT was designed to be administered by telephone, although it may also be administered face-to-face. We are sensitive to the issue of individual differences in hearing ability, although this has not presented particular problems in a previous large-scale phone survey of older adults (Herzog & Wallace, 1997). Similarly, hearing issues have been very rare in the pilot work we have carried out in developing the BTACT. Nevertheless, at the beginning of each interview we verify that the participant can repeat a short string of numbers (3, 8, 5) before proceeding further. Interviewers may also wish to obtain a self-rating of hearing ability, e.g. using a five-point rating scale ("poor" to "excellent"). Cognitive data would be excluded from the analysis for any participant who has difficulty understanding instructions, even after they are repeated. We also verify that the participant is a native English speaker, or if not, at what age he/she began speaking English.

In order to obtain the best cognitive performance from all of the ages tested, experimenters should be sensitive to time-of-day effects based on circadian rhythms. For example, adults over 45 may show their best performance before noon, and younger adults in the afternoon. (Hasher, Zacks, & May, 1999). Before beginning the cognitive battery, the purpose is explained briefly and participants are asked if they would like to continue. We avoid the use of the word "test", and use non-threatening terms such as "exercise" or "task" so as to minimize anxiety about performance. Experimenters make it very clear that no one is expected to get all of the answers correct, and participants should simply do their best. The overall tone should be supportive and non-judgmental (see Herzog & Rodgers, 1999). (See the experimenter instructions found in the BTACT, Appendix B)

An important feature of the cognitive protocol is the provision for responses of both "Don't know" and "Refuse to answer." This allows respondents the option not to answer a question if they so choose. For the age groups to be tested, we expect relatively low rates of refusal.

Trained interviewers read the instructions for each section to participants. We have chosen to use this method rather than a recorded session, because it provides greater sensitivity in responding to questions or uncertainties that the participant may have (Herzog & Rodgers, 1999), as well as to random events such as coughs or sneezes that might disrupt timing of events that require precise timing. Before each section the instructions are given clearly, along with a practice item, and the participant will be probed for understanding and have an opportunity to ask questions. The cognitive testing session is recorded as a computer sound files for later scoring of response accuracy and latencies, using sound-editing software (see Appendix A). For best results in recording latencies, we recommend that the interviews be conducted over land-line telephones rather than cellular telephones, which can produce more variable delays in transmission and processing.

Pilot Study I

In the course of developing the BTACT, we first carried out a validation study using an earlier version of the test. The primary goal for the validation study was to compare the results of the telephone administration with the more standard face-to-face administration, and to determine whether the same patterns of age differences would be found with both modes of assessment. The cognitive battery initially included the following tasks: verbal memory, as assessed by immediate and delayed paired-associate learning (Wechsler, 1997b); working memory, assessed by backward digit span (Wechsler, 1997a); verbal fluency and executive function (Drachman & Leavitt, 1972), assessed by simple verbal fluency tests (initial letter and category naming) and an alternating version (alternating letter and category generation) (Wechsler, 1997a); and reasoning, assessed by a 10 item number series test (after Salthouse & Prill, 1987). Latencies on the verbal memory and reasoning tasks reflected speed of processing. Participant responses were recorded on a Macintosh G3 computer using Sound Edit software, for later transcription and scoring of both accuracy and latencies of responses.

Forty-six participants ages 20 to 79 (M=49.2, SD=17.4) were tested both over the telephone and face-to-face, with intervals of 2-4 weeks between tests. Although we expected some practice effects, we counterbalanced order over the two modes of test administration. Twenty-six participants were tested first over the telephone, and 20 were tested first face-to-face. The mean education level was 16.0 (SD=2.6, range 11-22 years), and the group included 30 females and 16 males. The two groups (phone first and face-to-face first) did not differ significantly in age, education, or gender.

Overall, our findings suggested that administering the tests by phone did not show differences from face-to-face administration. These findings are consistent with previous authors who have looked at effects of mode of testing, and have found no significant differences between telephone and face-to-face testing even in older adults (Debanne et al., 1997; Herzog and Rodgers, 1988; Herzog & Wallace, 1997; Kawas et al., 1995; Nesselroade et al., 1988). The Karolinska Institute Twins Study in Sweden (Nesselroade et al, 1988) assessed cognitive functioning in older adults by telephone, and found that taking the standard versions of established cognitive tests and shortening them to half the original length only minimally compromised the overall reliability of the instrument. Our measures showed satisfactory psychometric properties of reliability and validity, and were sensitive to individual differences in age. Some measures, such as verbal fluency and reasoning, were also sensitive to educational level. Additional support for the reliability and validity of these measures is that they derive from well-validated tests such as the WAIS and the WMS, and standard neuropsychological testing.

Table 1 presents means for each test for the phone administration and the face-to-face administration, as well as the correlation between the two modes of testing, and the correlation between performance and age. The pattern of performance across age is shown in Figure 1, which shows standardized scores across tertiles of age: young (20-44 years), middle-aged (46-58 years), and older (61-79 years) adults. We will focus primarily on findings for the measures that were retained in the final revised version of the BTACT.

Paired-associate learning showed correlations of .55 for immediate and .67 for delayed tests between the two modes of administration. In the revised version of the BTACT we use word list recall as a measure of verbal episodic memory, which should show similar effects of mode.

Backward span performance did not differ significantly across phone and face-to-face modes of testing (see Table 1). Previous studies have shown test-retest reliability of .87 to .93 (WMS, 1997). Our pilot work confirmed that the backward span test was sensitive to age differences, beginning in middle age (see Figure 1). Support for backward span as a valid index of working memory comes from the fact that our findings are quite similar to other studies that have used composites of several working memory measures (e.g. N-back, subtract 2-span) in terms of the pattern of correlations between working memory, age, fluency, and verbal memory (Bryan, Luszcz & Pointer, 1999; Hultsch et al., 1998). Also, Engle et al. (1999) have recently shown a strong association between working memory and fluid intelligence, which is consistent with our finding of a significant correlation between performance on the backward digit span and the reasoning tests ($r=.43$, $p<.01$).

We administered three speeded verbal fluency tests commonly used in neuropsychological testing, that required participants to generate words based on an initial letter (F), a semantic category (animals), or alternating between a letter (S) and a category (foods) (cf. Borkowski, Benton, & Spreen, 1967; Lezak, 1995; Parkin et al., 1995). In order to keep the BTACT brief, in the revised version we retained only the category fluency task. Mode of testing was not significantly different for the category test or the alternating test. Scores were comparable to previous studies using these measures (e.g. Debanne et al., 1997; Ruff et al., 1996; Tombaugh, Kozak, & Rees, 1999; Zec et al., 1999). Previous studies have reported test-retest reliabilities of .74 for fluency tests (Bryan et al., 1997; Ruff et al., 1996).

Table 1.
<u>Summary Data for Cognitive Tests</u>

Test	**Phone Test**	**Face-to-Face Test**	**r-Face with Phone**	**r- Age**

PA Immediate	4.3 (3.0)	4.4 (2.7)	.55***	-.53***
PA Delayed	5.0 (2.7)	4.7 (3.0)	.67***	-.52***
Backward Digit Span	5.1 (2.4)	5.2 (2.5)	.85***	-.70***
VFL Letter	16.9 (4.3)	18.5 (6.2)	.73***	-.34*
VFL Category	22.3 (4.8)	22.8 (6.5)	.53***	-.41**
VFL Alternating	15.8 (5.2)	17.1 (4.3)	.71***	-.47**
Number Series	5.1 (2.4)	5.2 (2.5)	.85***	-.70***

*** p < .001, ** p < .01, * p < .05

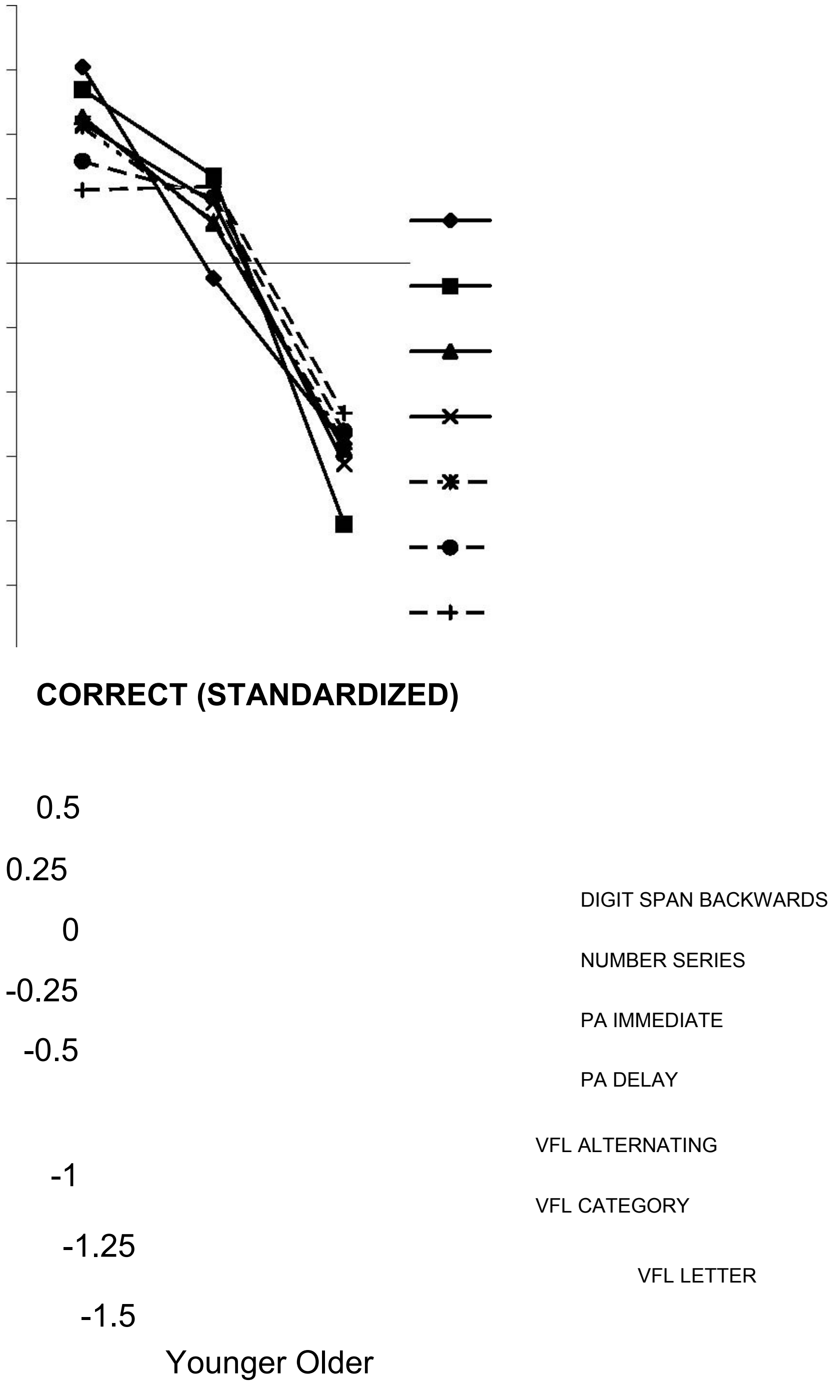
As predicted, we found that the category fluency task was sensitive to age (r=-.41), with performance declines begining in middle age for this sample. Age-related declines in these tasks have been attributed to declines in switching efficiency (DiGirolamo et al., 2001; Kramer et al., 1999), which is important in the semantic fluency and alternating task but not the letter fluency task (Troyer, Moscovitch, & Winocur, 1997). We also found significant relationships with education and category fluency (r=.35, p<.05), although our data were consistent with previous studies (Auriacombe, et al., 2001; Tombaugh et al., 1999) in showing that semantic fluency was associated more with age than with education. Also, category fluency showed a significant correlation with backward digit span, supporting the role of working memory during retrieval (Rosen & Engle, 1997).

Figure 1 Age Related Differences on Cognitive Tests (standardized scores)

1

0.75

-0.75



A microanalysis of verbal fluency performance also proved to be of interest. First, we found a significant positive age association with the number of repetitions that participants made for letter fluency, consistent with previous reports that older adults have more difficulty monitoring output (Perlmuter, et al., 1987). Also, the temporal pattern of output was of interest: we found larger age differences in the later phase of the fluency task, which has been suggested reflects a more effortful controlled search as opposed to semiautomatic, rapid search processes early in the phase (Fernaeus & Almkvist, 1998). This is consistent with the theory that older adults are particularly affected by effortful cognitive processes (Craik & Anderson, 1999).

The original number series test included 10 items, which was reduced to five items in the revised version. Number series performance did not differ across mode of testing, with a correlation of .85 between the two tests. Previous reports have noted good reliability on reasoning tasks (Schaie, 1996). As with previous reports (e.g. Schaie, 1996), we found that performance declined with age, with performance differences beginning in middle age (see Figure 1). Performance was significantly correlated with backward digit span ($r=.43$, $p<.01$), as would be predicted if both tasks draw on fluid abilities and working memory. Consistent with previous work (Salthouse & Prill, 1987), age differences increased across the three levels of difficulty of reasoning problems, with correlations with age of $r=-.41$ for Level 1, $r=-.52$ for Level 2, and $r=-.62$ for Level 3.

In summary, data from the pilot study confirm the equivalence of phone testing and face-to-face testing for this cognitive battery, and show that the tests used have acceptable psychometric properties. Older adults were able to understand instructions and carry out the tests over the phone. The tests were sensitive to individual differences in age and education, generating patterns of results that were consistent with previous findings and with our predictions. The pattern of intercorrelations among measures was consistent with previous studies such as the Berlin Aging Study (Baltes & Mayer, 1999) and the Victoria Longitudinal Study (Hultsch et al., 1998), which have used composites of measures for memory, speed, reasoning, and fluency. Although these factors were found to be intercorrelated, previous work suggests that each contributes a significant amount of reliable, ability-specific variance (Lindenberger & Baltes, 1994), with differential trajectories and sensitivity likely by age and education.

Pilot Study II

Based on our initial pilot data, we refined and revised the BTACT, by adding a more complete assessment of processing speed while keeping the test under 20 minutes in length for most respondents. In addition, we dropped the paired associate test and substituted word list recall. Thus, the current version includes measures of episodic verbal memory (word list recall), working memory span (backward span), executive function (verbal fluency), reasoning (number series), speed of processing (backward counting), and attention-switching and reaction time (switching task). The latter two measures were added in order to capture age changes in processing speed, which we believe to be of special significance in assessing adult cognitive abilities. Word list recall was substituted for paired associate recall, as a more age-sensitive measure.

We carried out pilot testing on this revised version using a probability sample of adults from the Greater Boston area, in order to achieve a more representative sample than the convenience sample tested in phase I. We tested 56 adults ranging in age from 23 to 80, including 23 men and 33 women, all of whom had been tested approximately one year earlier on face-to-face tests of cognition. For purposes of analysis we divided these into a young group (aged 23-39), a middle-aged group (aged 40-59), and an older group (aged 60-80). Table 2 gives characteristics of each of these groups in terms of educational backgrounds, vocabulary performance on the Shipley Vocabulary Test , and self-ratings of health and hearing (each on a five-point scale with one representing "poor" and five as "excellent" as compared to others of the same age). There were no significant differences between the groups in education, vocabulary, self-rated health, or self-rated hearing.

Table 2.
Demographic information for Pilot II participants

	Younger (23-39)	Middle (40-59)	Older (60-80)
N	17	19	20
Gender	6 M, 11 W	9 M, 10 W	8 M, 12 W

Means and (SD) Age 32.82 *(5.11)* 49.89 *(5.71)* 69.85 *(5.49)* **Education** 16.00 *(2.32)* 15.84 *(2.29)* 15.00 *(2.72)* **(in years) Vocabulary** 18.00 *(1.94)* 18.26 *(1.76)* 18.47 *(1.81)* **(0-20) Health Rating** 4.06 *(1.09)* 4.26 *(0.81)* 4.35 *(0.75)* Poor (1) – Excellent (5) **Hearing Rating** 4.35 *(0.79)* 4.21 *(1.13)* 4.00 *(0.79)* Poor (1) – Excellent (5)

Figure 2.
<u>Relationship of age and BTACT performance (standardized scores)</u>

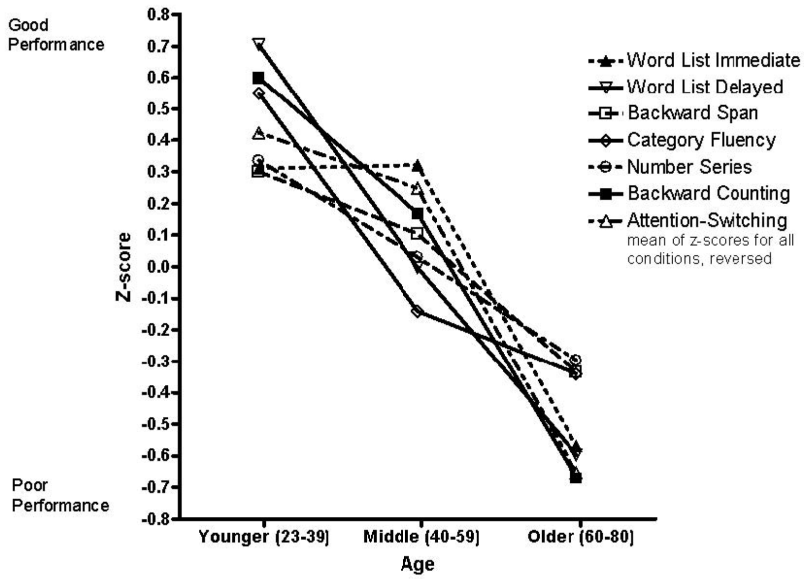


Figure 2 shows the relationship between age and the BTACT measures for these age groups. For the purpose of this figure, scores on each of the tests have been standardized, with higher scores representing good performance and lower scores representing poorer performance. As the figure shows, all of the measures showed age differences , although the pattern of effects varied from test to test. The largest age differences were shown for the speeded tasks, backward counting and attention-switching, as well as for category fluency and delayed word list recall, all of which showed lower levels from young to middle-aged to older groups. Backward span and number series showed more gradual declines with age, while immediate word list recall did not show differences until after middle-age. Data for these tests are presented in Table 3.

Table 3.
<u>Means and standard deviations of BTACT scores (items correct) by age group</u>

	Word List	Word List	Back.	Category	Number	Back.	Attention-

	Imm. Correct	Del. Correct	Digit Span	Fluency	Series (% correct)	Counting	Switching (latencies in msec)
Younger	7.76 *(2.05)*	6.06 *(2.38)*	4.69 *(1.54)*	23.88 (*4.53)*	.63 *(.26)*	49.6 *(10.91)*	896.17 *(111.34)*
Middle	7.79 *(1.90)*	4.32 *(1.83)*	4.44 *(.98)*	20.05 (*4.34)*	.54 *(.32)*	44.8 *(9.13)*	929.84 *(172.09)*
Older	5.90 *(1.92)*	2.85 *(2.16)*	3.90 *(1.12)*	18.94 *(6.54)*	.45 *(.26)*	35.60 *(8.61)*	1091.58 *(124.51)*

Effects of age and education

Table 4 presents the correlations among age, education, and the BTACT measures. As shown by the table, we found significant negative relationships with age and performance on the immediate and delayed word list recall, category fluency, number series, and backward counting, with a marginal age effect on backward digit span. Also, age was related to significantly longer latencies on both the baseline condition and the task-switching condition of the red/green attentional task, demonstrating age-related slowing.

Although education was correlated with some of the measures, the significant correlation between age and education reflects the lower level of education achieved by the older group, and therefore we carried out partial correlations to control for the effect of education. After statistically controlling for the effect of education, the effect of age remained significant for all measures except the number series. The age association with category fluency was reduced to $r = -.31$, which is consistent with previous reports on the role of education in this task (Ruff et al., 1996). Category fluency performance depends not only on speed but on vocabulary, a crystallized ability. After partialling out education, the correlations between age and latencies on the attention task were reduced to $r = .39$ for the normal baseline, $r = .35$ for the reverse baseline, $r = .46$ for the switch trials, and $r = .28$ for the no-switch trials , perhaps reflecting a general tendency for more highly-educated individuals to perform at higher levels of functioning. A significant effect of partialling out education was seen also on the number series task, where the age effect was reduced to $r = -.21$.

As shown in correlation Table 4, both immediate and delayed word list recall showed significant correlations with age. In addition, a combined measure of total word recall (immediate plus delayed) had a correlation of $r = -.49$ with age, and a derived measure of forgetting over the 15 minute session (immediate – delayed) was greater with age, $r = .26$, $p < .05$. Partialling out the effects of education had very little effect on these associations.

Table 4.

Correlations of BTACT measures with age and education; inter-correlations of BTACT measures

	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	1.00											
2. Education	-.22*	1.00										
3. Word list (immediate)	-.38**	.19	1.00									
4. Word list (delayed)	-.51**	.09	.69**	1.00								
5. Backward Digit Span	-.22	.13	.38**	.42**	1.00							
6. Category Fluency	-.35**	.28*	.34**	.64**	.12	1.00						
7. Number Series	-.26*	.33**	.15	.21	.31*	.36**	1.00					

8. Backward Counting	-.51**	.03	.29*	.45**	.44**	.27*	.38**	1.00				
9. Attention-Switching: Normal Baseline	.43**	-.24*	-.26*	-.19	-.19	-.19	-.27*	-.42**	1.00			
10. Attention-Switching: Reverse Baseline	.40**	-.32**	-.28*	-.26*	-.38**	-.18	-.41**	-.57**	.67**	1.00		
11. Attention-Switching: Experimental Switch	.48**	-.21	-.40**	-.32**	-.43**	-.18	-.40**	-.59**	.67**	.68**	1.00	
12. Attention-Switching: Experimental No Switch	.31*	-.20	-.33**	-.23*	-.31*	-.20	-.35**	-.52**	.72**	.74**	.69**	1.00

*p<.05. **p<.01. (One-tailed significance)

Word list recall: Both immediate and delayed word list recall showed significant age effects, with the effect being greater for the delayed trials. On the immediate trials the oldest age group scored lower than the younger and middle groups, while on the delayed there was a significant drop from the young to the middle group, as well as from the middle to the older group. We also calculated a forgetting score (immediate – delayed), on which the younger group showed significantly less forgetting over the interval than the middle and older groups. Repetitions of studied items were few (Y=0.40 items, M=0.90, O=1.17) and the groups did not differ. Neither did the groups differ in the number of extralist intrusions (Y=0.29, M=0 .75, O=0.20).

Backward digit span: Although there was an age-related decline in backward digit span, it did not reach significance with this sample.

Category fluency: There was a significant age decline in the number of unique responses produced in the category fluency task, with younger adults producing more than either middle-aged or older adults. Repetitions were few in number, and did not differ significantly between groups, (Y= .29, M=.55, O=.11).

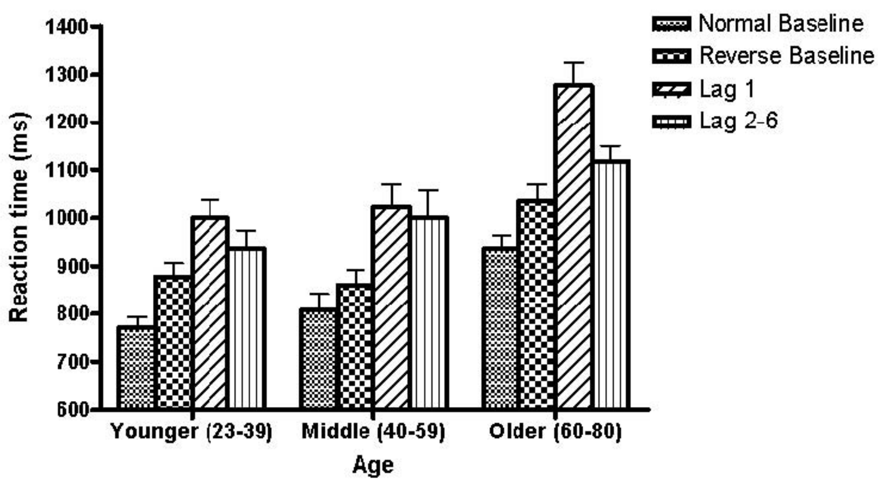
Number series: Although number series showed a significant negative correlation with age, the downward trend did not result in significant differences among these groups. Age group differences were largest on item one, for which the older group scored significantly lower than the middle-aged and younger groups. Overall means for the items were: Item one, M= .38, SD= .49; item two, M= .82, SD=.39; item three, M= .53, SD= .50; item four, M= .43, SD= .50; item five, M= .50, SD= .54.

Backward counting: Backward counting produced robust age effects, with the older group producing fewer correct responses than the younger and middle groups. Errors were very few in number (means of 0.10, 0.50,.and 1.06 for the young, middle and older groups, respectively), and the groups did not differ in number of errors.

<u>Latency performance</u>

Figure 3 shows performance by three age groups (young, 23-39; middle-aged, 40-59; older, 60-80) on the attention-switching task. Latencies (in milliseconds) are shown for four types of trials: (1) normal baseline (blocked); (2) reverse baseline (blocked); and in the mixed task-switching condition, trials that were (3) “switch trials”, i.e. the first trial after a switch was required, and (4) “non-switch trials”, i.e. subsequent trials that did not require switching between normal and reverse. Data were lost from one participant due to equipment failure, and data were dropped for this task for one middle-aged participant who was apparently distracted. Latencies showed a robust pattern of age associations across the three groups. We see an increase in latency from the easiest blocked normal condition, to the blocked reverse condition. Latencies increase in the task-switching condition, especially for the switch trials that immediately follow a cue to switch response mode, on which age differences were especially pronounced.

Figure 3.
Age-related differences in reaction time on the attention-switching task.



As the figure shows, there are positive correlations of age with latency for all conditions, and means are significantly longer for older adults than for middle-aged or younger adults. However, the increase is most pronounced for the switch trials, consistent with the hypothesis that task-switching becomes more difficult with age. Age was associated with greater attention-switching costs at both the global level and the local level. First, age was associated with a significant global cost, $r = .50$, $p < .01$, calculated as the proportional increase in latency from the blocked baseline conditions, to the task-switching condition. ((Task-switching trials – blocked baseline trials)/blocked baseline trials). On the figure, reading from left to right across the bars for each group, this reflects the increase in latency from the first two blocked task bars on the left to the third and fourth alternating task bars. In addition, age was also associated with a significantly increase in the local cost within the task-switching condition, $r = .50$, $p < .05$. That is, the older adults’ times were differentially slower for the first response that followed a switch, relative to the subsequent run of responses that did not require switching. This is termed a local cost, within the task-switching condition, and it showed a significant correlation of $r = .50$ with age, $p < .05$. (Calculated as (switch trials – non-switch trials)/ non-switch trials). This may be seen in the figure as the increase in latency from the last bar in each group, to the third bar in each group. Older adults showed significantly greater costs than middle-aged adults, who showed a greater cost than young adults. The cost of incompatibility—i.e., moving from a compatible (normal) response to an incompatible (reverse) response—was not associated with age ($p < .05$). All age groups showed high accuracy on the attentional task, with error rates of 4% or less, and the groups did not differ significantly on accuracy.

Construct validity: Intercorrelations among cognitive measures

The retesting of this group of participants at an interval of approximately one year allows us to examine the pattern of intercorrelations among cognitive measures. At Time 1, participants were tested face-to-face using several paper-and-pencil measures. Time 2 testing was conducted over the phone, using the auditory tests from the BTACT. Table 5 shows the correlations among the Time 1 measures and the Time 2 BTACT measures. The correlation table shows good convergence among the measures that would be expected to assess the same constructs.

The measures of episodic verbal memory from the BTACT, the immediate and delayed word list recall, showed correlations of r=.45 and r=.54, respectively, with recall of categorizable word lists from Time 1. Backward digit span, the BTACT measure of working memory span, showed correlations of r=.40 with the forward digit span and r= .35 with the backward digit span administered at Time 1.

Category fluency, the BTACT measure of executive function, correlated significantly with digit symbol substitution, r= .36, another measure that requires executive control, as well as with recall at Time 1 testing, r=.52. Importantly, category fluency was the only BTACT measure to correlate significantly with the Adjusted Ratio of Clustering from the Time 1 categorizable word list, r= .27, as would be predicted: clustering is a useful strategy in producing items on the fluency task, just as it is on the Time 1 list of categorizable words. By contrast, clustering is not particularly useful on the BTACT word list recall task of unrelated items, and the immediate and delayed word list recall scores showed no correlation with the clustering measure. Also, consistent with predictions based on previous research (Ruff et al., 1996), category fluency showed significant correlations with vocabulary scores from Time 1, r= .28.

Number series performance also showed a significant correlation with vocabulary scores, r=.33, which are related to education. In addition, number series performance also correlated with more fluid cognitive measures from Time 1 testing including backward digits, r=.35, digit symbol substitution, r=.33, and and word recall, r=.37. Also, Table 4 showed correlations between number series performance and speeded measures on the BTACT such as the backward counting task, r=.38, and the attention-switching latencies, r=-.27 to r=-.41. Overall, this suggests that although number series involves unique elements of reasoning, it also depends heavily on speeded processing and memory in order to hold the digits while determining the pattern.

The pattern of correlations in Table 5 suggests that all of the BTACT measures share some dependence on both memory (note correlations with word list recall) and on speed of processing, (note correlations with digit symbol substitution, which was particularly strong for the BTACT speeded measures such as backward counting, r=.50, and response latencies on the attention switching task, r=-.36 to r=-.51.). This would be predicted based on the fact that the auditory presentation used throughout the BTACT, unlike reading, does not allow looking back, and the listener must process stimuli presented at a fairly rapid pace.

Table 5.

Correlations between BTACT measures (listed vertically) and Time 1 measures (on horizontal)

	Shipley Vocab	**Digit Symbol Substitution**	**Forward Digit Span**	**Backward Digit Span**	**Word List Recall**	**Word List Recall Adjusted Ratio of Clustering**	**PIC Internal Control**	**PIC External Control**
Word list imm.	-.07	.52**	.15	.17	.45**	.03	.26*	-.09

Word list del.	.10	.54**	.13	.13	.54**	.02	.21	-.01
Back. Digit Span	.23*	.35**	.40**	.35**	.38**	-.01	.20	-.13
Category Fluency	.28*	.36**	.06	.08	.52**	.27*	.21	-.18
Number Series	.33*	.33**	.27*	.35**	.37**	.07	.26*	-.41**
Backward Counting	.05	.50**	.31**	.26*	.41**	-.03	.19	-.16
Attention-Switching: Normal Baseline	.17	-.36**	.12	.13	-.25*	.02	-.07	.10
Attention-Switching: Reverse Baseline	-.12	-.47**	-.12	-.07	-.31*	.05	-.20	.33**
Attention-Switching: Experimental Switch	.15	-.51**	-.22	-.19	-.32**	.11	-.14	.14
Attention Switching: Experimental No Switch	.05	-.44**	-.06	-.19	-.29*	-.04	.00	.15

*p<.05. **p<.01. (One-tailed significance)

Face-phone validity

Additionally, we are currently investigating the phone-face validity of the most current version of the BTACT in an ongoing study. In this experiment, participants first complete the BTACT face-to-face with an experimenter. The participant is contacted one to two weeks later by the same experimenter to complete an identical version of the experiment, but in this instance over the telephone. The current N=25, including 11 younger participants (undergraduates at Brandeis University aged 18-22 years) and 14 older participants (community-dwelling volunteers aged 67 to 82 years). (Data were dropped from 4 participants due to non-completion, not following directions, and technical difficulties). As illustrated by Table 6, the correlations between scores attained on the face version of the test and the phone version of the test are consistently high, showing good validity between the tests administered face-to-face and over the telephone.

Table 6.
Correlations between phone vs. face versions of the BTACT subtests

	Correlation Phone vs. Face
Word List Immediate	.73**
Word List Delayed	.88**
Backward Digit Span	.57**
Category Fluency	.67**
Number Series	.56**
Backward Counting	.95**
Attention-Switching: Normal Baseline	.80**
Attention-Switching: Reverse Baseline	.83**
Attention-Switching: Experimental Switch	.65**
Attention-Switching: Experimental No Switch	.75**

*p<.05. **p<.01. (Two-tailed significance)

Further directions

We are currently in the process of validating an alternate version of the BTACT, which will be useful in studies with multiple assessments. This version, in addition to presenting an alternate form of the tasks from the BTACT, uses a computerized interface through which the experimenter navigates, entering responses for each task as the participant makes them. This will, we hope, reduce some of the work of scoring, especially in scoring the reaction time-relevant measures such as the red/green attention-switching task.

Conclusion

In summary, we have presented here a brief cognitive battery for assessing cognitive function across the adult lifespan, that can be easily administered over the telephone. In our pilot work these tests have demonstrated good psychometric properties, and have produced a good range of performance in participants who vary widely in age, education, and socio-economic background. We believe that the BTACT makes a contribution to the field of cognitive research by extending the tools available to researchers in order to test a wider range of participants, including those who may be unable or unwilling to come into a laboratory or to participate in other research settings. Moreover, the BTACT goes beyond other telephone batteries in that it provides sensitive assessments of reaction time and speed of processing, important for detecting subtle age and health-related differences in large community samples. We hope that the instrument will provide a useful tool for collecting data in survey research and many other areas.

References

Auriacombe, S., Fabrigoule, C., Lafont, S., Amieva, H., Jacqmin-Gadda, H., & Dartigues, J. (2001). Letter and category fluency in normal elderly participants: a population-based study. *Aging, Neuropsychology and Cognition, 8*, 98-108.

Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon Press.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*, 5-28.

Baldo, J.V., Shimamura, A.P., Delis, D.C., Kramer, J., Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society, 7*, 586-596.

Baltes, P. & Mayer, K. U. (1999). *The Berlin Aging Study: Aging from 70 to 100*. New York: Cambridge University Press.

Baltes, P., Staudinger, U.M., & Lindenberg, U. (1999). Lifespan psychology: theory and application to intellectual functioning. *Annual Review of Psychology, 50*, 471-507.

Barrett-Connor, E., & Kritz-Silverstein, D. (1993). Estrogen replacement therapy and cognitive function in older women. *Journal of the American Medical Association, 269*, 26372641.

Bleecker, M.L., Bolla-Wilson, K., Agnew, J., & Meyers, D.A.(1988). *Age-related sex differences in verbal memory. Journal of Clinical Psychology, 44*, 403- 411.

Borkowski, J.G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia, 5*, 135-140.

Brandt, J., Spencer, M., & Folstein, M. (1988). The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 1*, 111-117.

Bryan, J. & Luszcz, M.A. (2000a). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology, 22*, 40 55.

Bryan, J. & Luszcz, M.A. (2000b). Measures of fluency as predictors of incidental memory among older adults. *Psychology and Aging, 15*, 483-489.

Bryan, J., Luszcz, M.A., & Pointer, S. (1999). Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies. *Aging, Neuropsychology, and Cognition, 6*, 273-287.

Bryan, J., Luszcz, M.A., & Crawford, J.R. (1997). Verbal knowledge and speed of information processing as mediators of age differences in verbal fluency performance among older adults. *Psychology and Aging, 12*, 473-478.

Carlson, L.E., & Sherwin, B.B. (1998). Steroid hormones, memory, and mood in a healthy elderly population. *Psychoneuroendocrinology, 23*, 583-603.

Cepeda, N.J., Kramer, A., & Gonzalez de Sather, J.C.M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology, 37*, 715-70.

Craik, F. I. M., & Anderson, N. (1999). Applying cognitive research to problems in aging. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII* (pp. 583-616). New York: Academic Press.

Debanne, S.M., Patterson, M.B., Dick, R., Riedel, T.M., Schnell, A., & Rowland, D.Y. (1997). Validation of a telephone cognitive assessment battery. *Journal of the American Geriatric Society, 45*, 1352-1359.

DiGirolamo, G.J., Kramer, A.F., Cepeda, N.J., et al., (2001). General and task-specific frontal lobe recruitment in older adults during executive processes: A fMRI investigation of task-switching. *Neuroreport, 12*, 2065-2071.

Drachman, D.A., & Leavitt, T. (1972). Memory impairment in the aged: Storage versus retrieval deficit. *Journal of Experimental Psychology, 93*, 302-308.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Fernaeus, S., & Almkvist, O. (1998). Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology, 20*, 137-143.

Geffen, G., Moar, K.J., O'Hanlon, A.P., Clark, CR., & Geffen, L.B. (1990). Performance measures of 16- to 86-year-old males and females on the Auditory Verbal Learning Test. *Clinical Neuropsychologist, 4*, 45-63.

Grady, C.L., & Craik, F.I. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology, 10*, 224-231.

Herzog, A.R., & Rodgers, W.L. (1998). Cognitive performance measures in survey research on older adults. In N. Schwarz, D.C. Park, B. Knauper, & S. Sudman (Eds.). *Cognition, aging and self-reports* (pp. 327-340). Philadelphia, PA: Psychology Press.

Herzog, A.R., Rodgers, W.L., & Kulka, R.A. (1983). Interviewing older adults—a comparison of telephone and face-to-face modalities. *Public Opion Quarterly, 47*, 405-418

Herzog, A.R., & Wallace, R.B. (1997). Measures of cognitive functioning in the AHEAD study [Special Issue]. *Journal of Gerontology: Social Sciences, 52B*, 37-48.

Horn, J.L., & Masunaga, H. (2000). New directions for research into aging and intelligence: The development of expertise (pp. 125-159). In T.J. Perfect & E.A. Maylor (Eds.), *Models of cognitive aging*. London: Oxford University Press.

Ivnik, R.J., Malec, J.F., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1990). The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. *Psychological Assessment, 2*, 304-312.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged*. New York: Cambridge University Press.

Kawas, C., Karagiozis, H., Resau, L, Corrada, M., & Brookmeyer, R. (1995). Reliability of the Blessed Telephone Information-Memory-Concentration Test. *Journal of Geriatric Psychiatry and Neurology, 8*, 238-242.

Kempler, D., Teng, E.L., Dick, M., Taussig, I.M., Daviss, D.S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society, 4*, 531-538.

Kramer, A., Hahn, S., Cohen, N. J. et al., (1999). Ageing, fitness, and neurocognitive function. *Nature, 400*, 418-419.

Kramer, A., Hahn, S., & Gopher, D. (1999). Task coordination and aging: explorations of executive control processes in the task switching paradigm. *Acta Psychologica, 101*, 339-378.

Kray, J. & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15*, 126-147.

Kozora, E., & Cullum, C.M. (1995). Generative naming in normal aging: total output and qualitative changes using phonemic and semantic constraints. *Clinical Neuropsychologist, 9*, 313320.

Lachman, M. E., & Spiro, A. (2002) *Critique of cognitive measures in the Health and Retirement Study (HRS) and the Asset and Health Dynamics among the Oldest Old (AHEAD) study*. HRS Data Monitoring Board Report.

Lezak, M.D. (1995). *Neuropsychological assessment*, Third edition. New York: Oxford University Press.

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 9*, 339-355.

Maylor, E. A., & Wing, A. M. (1996). Age differences in postural stability are increased by additional cognitive demands. *Journal of Gerontology, Psychological Sciences, 51B*, P143-154.

Mayr, U. & Kliegl, R. (2000). Complex semantic processing in old age. *Psychology and Aging, 15*, 29-43

Meiran, N., Gotler, A., & Perlman, A. (2001). Old age is associated with a pattern of relatively intact and relatively impaired task switching. *Journal of Gerontology: Psychological Sciences, 56B*, 88-102.

Meyerson, J., Hale, S., Wagstaff, D., Poon, L.W., & Smith, G.A. (1990). The information-loss model: A mathematical theory of age-related slowing. *Psychological Review, 97*, 475-487.

Miller, L. S., & Lachman, M. E. (2000). Cognitive performance and the role of control beliefs in midlife. *Aging, Neuropsychology, and Cognition, 7*, 69-85.

Miyake, A. & Shah, P. (1999). *Models of working memory: mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press.

Nesselroade, J.R., Pederson, N.L., McClearn, G.B., Plomin, R., & Bergeman, C.S. (1988). Factorial and criterion validities of telephone-accessed cognitive ability measures. Age and gender comparisons in adult twins. *Research on Aging, 10*, 220-234.

Ofstedal, M.B., McAuley, G.F., & Herzog, A.R. (2002). *HRS/AHEAD Documentation Report*. Ann Arbor, MI: University of Michigan.

Parkin, A. J., Walter, B. M., & Hunkin, N. M. (1995). Relationships between normal aging, frontal lobe function, and memory for temporal and spatial information. *Neuropsychology, 9*, 304-312.

Perlmuter, L.C., Tun, P., Sizer, N., McGlinchey, R.E., & Nathan, D.M. (1987). Age and diabetes related changes in verbal fluency. *Experimental Aging Research, 13*, 9-14.

Rabbitt, P. (1999). *Methodology of frontal and executive function*. Hove, UK: Psychology Press.

Reuter-Lorenz, P.A., Marshuetz, C., Jonides, J., Smith, E.E., Hartley, A., Koeppe, R. (2001). Neurocognitive ageing of storage and executive processes. *European Journal of Cognitive*

Psychology, 13, 257-278.

Rey, (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rogers, R.D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Rosen, V.M., & Engle, R.W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211-227.

Ruff, R.M., Light, R.H., Parker, S.B., & Levin, H.S. (1996). Benton Controlled Oral Word Association Test: Reliability and update norms. *Archives of Clinical Neuropsychology, 14*, 329338.

Salthouse, T.A. (1992). Reasoning and spatial abilities. In F.I.M. Craik & T.A. Salthouse (eds.), *Handbook of aging and cognition* (pp. 167-211). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428.

Salthouse, T.A. (1998). Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26*, 153-166.

Salthouse, T.A., (2000). Aging and measures of processing speed. *Biological Psychology, 54*, 35-54.

Salthouse, T.A., Atkinson, T.M., & Berish, D.E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566-594.

Salthouse, T.A., Fristoe, N., McGuthry, K.E., & Hambrick, D.Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13*, 445-461.

Salthouse, T. A., & Prill, K.A. (1987). Inferences about age impairments in inferential reasoning. *Psychology and Aging, 2*, 43-51.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.

Spreen, O. & Straus, E. (1991). *A compendium of neuropsychological tests: administration, norms, and commentary.* London: Oxford University Press.

Taylor, E.M., (1959). *Psychological appraisal of children with cerebral deficits*. Cambridge, MA: Harvard University Press.

Tennestedt, S., Lachman, M.E., & Salthouse, T. (1999). Telecog: A brief cognitive telephone assessment. Unpublished manuscript, Watertown, MA: New England Research Institutes.

Tombaugh, T.N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical*

Neuropsychology, 14, 167-177.

Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components f verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*, 138-146.

Tun, P.A. (1989). Adult age differences in processing expository and narrative test. *Journal of Gerontology: Psychological Sciences, 44*, P9-P15.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13(2)*, 230-241.

Verhaeghen, P., & Salthouse, T.A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear effects and structural models. *Psychological Bulletin, 122*, 231-249.

Waters, G.S. & Caplan, D. (2003).The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*, 550-564.

Wechsler, D. (1997). *Weschler Adult Intelligence Scale - Third Edition* (WAIS). New York: Psychological Corporation.

Wechsler, D. (1997). *Weschler Memory Scale – Third Edition*. New York: Psychological Corp.

West, R.L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*, 272-292.

Willis, S.L. & Schaie, K.W. (1999). Intellectual Functioning in Midlife. In S.L. Willis & J.D. Reid (Eds.), *Life in the middle: psychological and social development in middle age* (pp. 233247). San Diego, CA: Academic Press, Inc.

Wingfield, A., Tun, P., & Rosen, MJ. (1995). Age differences in veridical and reconstructive recall of syntactically and randomly segmented speech. *Journals of Gerontology: Psychological Sciences, 50*, 257-266.

Zec, R.F., Landreth, E.S., Fritz, S., et al. (1999). A comparison of phonemic, semantic, and alternating word fluency in Parkinson's Disease. *Archives of Clinical Neuropsychology, 14*, 255264.

Appendix A. Scoring latencies:

In addition to accuracy on the attention-switching test, we calculate response latencies in milliseconds from the onset of the stimulus (e.g. "red") to the onset of the participant's response (e.g. "stop"). These are taken from the recorded sound files using a computerized sound-editing system (e.g. Sound Studio or Cool Edit Pro) that digitizes the sound track and visualizes the waveform of individual responses along a time line. Although this method is labor-intensive, we have found that it is best to manually determine latencies, due to cases in which participants make false starts, self-corrections, or extraneous noises or comments (e.g. coughs, "hmm-m") that would be taken as responses by the computer if a voice-activated relay were used. Once scorers have been trained, we have found good interscorer reliability of .97 and higher.

Appendix B. Brief Test of Adult Cognition by Telephone (BTACT)