

# Executive function on the Psychology Experiment Building Language tests

Brian J. Piper · Victoria Li · Massarra A. Eiwaz · Yuliyana V. Kobel · Ted S. Benice · Alex M. Chu · Reid H. J. Olsen · Douglas Z. Rice · Hilary M. Gray · Shane T. Mueller

© Psychonomic Society, Inc. 2011

**Abstract** The measurement of *executive function* has a long history in clinical and experimental neuropsychology. The goal of the present report was to determine the profile of behavior across the lifespan on four computerized measures of executive function contained in the recently developed Psychology Experiment Building Language (PEBL) test battery <http://pebl.sourceforge.net/> and evaluate whether this pattern is comparable to data previously obtained with the non-PEBL versions of these tests. Participants ( $N = 1,223$ ; ages, 5–89 years) completed the PEBL Trail Making Test (pTMT), the Wisconsin Card Sort Test (pWCST; Berg, *Journal of General Psychology*, 39, 15–22, 1948; Grant & Berg, *Journal of Experimental Psychology*, 38, 404–411, 1948), the Tower of London (pToL), or a time estimation task (Time-Wall). Age-related effects were found over all four tests, especially as age increased from young childhood through adulthood. For several tests and measures (including pToL and pTMT), age-related slowing was found as age increased in adulthood. Together, these findings indicate that the PEBL tests provide valid and versatile new research tools for measuring executive functions.

**Electronic supplementary material** The online version of this article (doi:10.3758/s13428-011-0096-6) contains supplementary material, which is available to authorized users.

B. J. Piper (✉) · V. Li · M. A. Eiwaz · Y. V. Kobel · T. S. Benice · A. M. Chu · R. H. J. Olsen · D. Z. Rice · H. M. Gray  
Department of Behavioral Neuroscience,  
Oregon Health & Science University,  
Portland, OR 97239, USA  
e-mail: piperb@ohsu.edu  
e-mail: psy391@gmail.com

S. T. Mueller  
Applied Research Associates,  
Fairborn, OH, USA

**Keywords** Age · Children · Adolescents · Elderly

The personal computer has become a ubiquitous tool in the neurobehavioral sciences over the last three decades. Test administration to human participants has particularly benefited from this technological advance because the instructions can readily be presented in a standardized fashion, potentially in multiple languages, large volumes of objective data can be efficiently collected with lower probabilities of experimenter error, and tests can be administered by various individuals after brief training. Relative to their paper-and-pencil counterparts, computerized tests can also be more easily adapted to a neuroimaging environment (Kubo et al., 2009) and require less manual effort for scoring and data analysis. The measurement of executive functions has especially benefitted from computerized test developments (Conners, 1985; Greenberg & Waldmant, 1993; Gur et al., 2010; Wild, Howieson, Webbe, Seelye, & Kaye, 2008). Executive functions are adaptive, goal-directed actions that allow an individual to override automatic or established behaviors. Tasks that involve set shifting, response inhibition, and working memory, especially those that require solving a novel problem, are thought to provide indices of executive function (Garon, Bryson, & Smith, 2008).

However, two challenges continue to face those interested in adopting computerized behavioral testing. First, although the price for individual tests may be reasonable, many researchers prefer to measure a broad array of sensory, motor, and higher-order cognitive functions using a battery approach (Piper, Acevedo, Craytor, Murray, & Raber, 2010; Wild et al., 2008). Some of the better studied batteries (e.g., the Cambridge Neuropsychological Test Automated Battery (CANTAB; Fray & Robbins, 1996) not only charge for the initial setup with specialized

equipment, but also have substantial annual or per-use license fees to keep the software operational. While this price is partially understandable to offset the resources needed for test development, excessive costs may limit the potential of smaller laboratories or investigators in developing nations from being full participants in the research process. Second, the computer code that underlies the commercial tests either may not be readily available or may be insufficiently documented so that other researchers can interpret the operations or verify their correctness.

The Psychology Experiment Building Language (PEBL) was developed to overcome both of these limitations. The software is freely downloadable (<http://pebl.sourceforge.net/>) and has been documented so that others with moderate programming skills can modify the individual tests to meet their experimental needs (Mueller, 2010b). The current PEBL battery (version 0.6) consists of approximately 50 tests, including many classic ones in experimental psychology and behavioral neurology (Mueller, 2010a, 2010c). The objective of the present study was to report on findings from a subset of PEBL battery tests that assess several core capacities of executive function across the lifespan, including attention, planning, decision making, and cognitive flexibility.

Characterization of age-related performance profiles can be useful for several purposes, including providing fundamental data to help determine the etiology of these changes, enabling diagnostic tests of specific impairments to be developed, and, eventually, identifying interventions for both younger and older individuals that may optimize neurocognitive functioning. Many neurobehavioral capacities improve in young children as they gain maturity, including fine motor function, reaction time, sustained attention, and working memory (Kail & Salthouse, 1994; Piper, 2011). As age progresses, there is often an age-related loss in processing speed and the emergence of generalized cognitive reduction, with deficits in visuospatial skills, working memory, and executive function (see Mahncke, Bronstone, & Merzenich, 2006; Park, 2000). In the present study, we sought to establish whether these general aging effects would occur across a set of four distinct and complementary behavioral tests. Each of the four tests employed, including their historical antecedents, is described below. Three of these tests are versions of some of the most widely used measures in clinically orientated behavioral neuropsychological research, while the fourth has seen limited prior use. To provide some context, we will first review the historical origins of the tests we completed.

The Trail Making Test (TMT) was originally part of the Army Individual Test Battery (1944) but was later adopted into the Halstead–Reitan Neuropsychological Test Battery (Reitan, 1955). Traditionally, the experimenter used a

stopwatch to record how long participants took to connect dots on paper that were either numbered (Trails Part A: 1–2–3–4–5) or alternated between numbers and letters (Trails Part B: 1–A–2–B–3). The TMT is simple to administer and is used as an index of visual attention (Trails A) and cognitive flexibility (Trails B). The TMT has also been employed as a sensitive indicator of brain damage (Davidson, Gao, Mason, Winocur, & Anderson, 2008; Perriñez et al., 2007; Reitan, 1955, 1958; Stuss et al., 2001). In addition, the TMT has been employed with a wide variety of other populations, including those with age-associated memory impairment (Hänninen et al., 1997), alcoholics (O’Leary, Radford, Chaney, & Schau, 1977), and children with temporal lobe epilepsy (Guimarães et al., 2007).

The Wisconsin Card Sorting Test (WCST) is another prevalent measure of executive function in contemporary neuropsychological practice and research. This test was originally conceptualized by Berg (1948; Grant & Berg, 1948). The original design of the task involved physically placing cards in one of four piles on the basis of the characteristics of the stimuli. The rule for correctly sorting the stimuli changes regularly, and the ability to switch strategies based on the shape, color, or number of stimuli is recorded. A response in which the earlier rule is incorrectly employed is a perseverative error. Like the TMT, the WCST has been used with various clinical conditions, including schizophrenics (Shad, Tamminga, Cullum, Haas, & Keshavan, 2006) and children with attention deficit hyperactivity disorder (Tsuchiya, Oki, Yahara, & Fujieda, 2005).

The Tower of London (ToL) is a test of planning in which colored disks or balls on pegs are moved individually from an initial state to match a goal state. Optimal performance involves forming, retaining, and implementing a plan to make as few moves as possible. This test was originally developed as a simplified version of the Tower of Hanoi by Shallice (1982). The cognitive and neurophysiological substrates of ToL performance have been frequently and thoroughly examined (Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999; Ward & Allport, 1997). Both lesion and neuroimaging findings have identified the prefrontal cortex as integral in performing the ToL, as well as the TMT and WCST (Davidson et al., 2008; Kubo et al., 2009; Schall et al., 2003; Zakzanis, Mraz, & Graham, 2005).

The “Time-Wall” task is a recently developed nonverbal decision-making test modeled after a task originally included in the Unified Tri-Services Cognitive Performance Assessment Battery, which was used by the military of the United States for personnel testing (Perez, Masline, Ramsey, & Urban, 1987; Snyder & Rice, 1990). The objective of the original Time-Wall task was to assess the ability to estimate the time at which a target, moving vertically at a fixed rate, will have traveled a specified distance. Thus, it draws on skills relating to both motion perception (Sekuler,

Watamaniuk, & Blake, 2002) and prediction. An interesting aspect of performance on Time-Wall type tasks is that this skill appears to be a stable characteristic that does not improve (Jerison, Crannell, & Pownall, 1957), even with extensive practice (Perez et al., 1987).

Although factor-analytic studies indicate that *executive function* is not a simple, unitary process (Fisk & Sharp, 2004; Miyake et al., 2000), performance across the lifespan on the aforementioned tests generally exhibits a U-shaped association with age. Table 1 provides an overall framework for the present endeavor by outlining the contributions of age to executive function, using the non-PEBL test versions. The Reitan Neuropsychological Laboratory, the originator and distributor of the Reitan TMT, has recognized this developmental profile and has constructed different versions of the test for preadolescent (ages, 9–14 years) versus older (ages, 15+) participants. Among adults, the time to complete Part B increases with age (Yeudall, Reddon, Gill, & Stefanyk, 1987; Tombaugh, 2004). Similarly, the number of perseverative errors decreases with age in children, is stable from ages 20 to 60, and is elevated at senescence on the WCST (Chelune & Baer, 1986; Fisk & Sharp, 2004; Hartman, Bolton, & Fehnel, 2001; Huizinga, Dolan, & Van der Molan, 2006; Strauss, Sherman, & Spreen, 2006). Young adults were most efficient at solving of the CANTAB ToL, relative to either children or older adults (DeLuca et al. 2003).

To demonstrate the validity of the PEBL executive function tests, we set out to determine whether the tests demonstrate the expected U-shaped relationship between age and behavior, with a progression during childhood, optimal performance (i.e., lowest scores/errors) during late-adolescence/early-adulthood, and then a regression during senescence.

## Method

Attendees of the Oregon Museum of Science and Industry were first asked their age and handedness and then were asked to complete a short computer-administered task that typically lasted from 3 to 12 min. The tests were implemented using the PEBL platform (all scripts are in the [supplemental materials](#)) and were typically identical or slightly adapted from versions distributed in version 0.5 of the PEBL Test Battery (exact methods are described below, and the scripts are available as [supplemental materials](#)). Testing was performed on one of ten personal computers running the Microsoft Windows operating system. The minimum age (5–7 years) to participate was based on the complexity of each test and preliminary testing. Children participating in an experiment were tested without the assistance of their parents or guardians, who instead were encouraged to take part in the study on a separate

**Table 1** Performance across the lifespan on measures of executive function

First Author	Age (N)	Findings
Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003	8–13 (108)	TMT <sup>a</sup> : B > A for completion times for each age group, no age effects
Yeudall et al., 1987	15–40 (225)	TMT <sup>a</sup> : B > A for completion times for each age group, age effects for B time but not A
Tombaugh, 2004	18–89 (911)	TMT <sup>a</sup> : B > A for completion times for each age group, gradual increase in B time starting at ages 65–69
Chelune and Baer, 1986	6–12 (105)	WCST <sup>a</sup> : PE decreased with age but was indistinguishable from adult norms at ages 10–12
Fisk and Sharp, 2004	20–85 (95)	WCST <sup>a</sup> : $r = .24$ between age and PE
Hartman et al., 2001	younger (85) and older (76) adults	WCST <sup>a</sup> : older adults (mean age 70) made more PE and completed fewer categories than college students
Haaland et al., 1987	64–87 (75)	WCST <sup>a</sup> : healthy older adults (ages 80–87) completed fewer categories and made more errors relative to ages 64–69
Huizinga et al., 2006	7–21 (384)	WCST <sup>b</sup> : PE and categories completed decreased across ages ToL <sup>b</sup> : additional moves decrease with age up to 15, 15- and 21-year-olds equivalent
Korkman et al., 2001	5–12 (800)	ToL <sup>a</sup> : improvement from ages 5 to 9
Malloy-Diniz et al., 2008	5–8 (371)	ToL <sup>a</sup> : improvement across ages tested
Lehto et al., 2003	8–13 (108)	ToL <sup>b</sup> : improvement across ages tested
DeLuca et al., 2003	8–64 (194)	ToL <sup>b</sup> : optimal performance at ages 20–29 compared with ages 8–10 or 50–64

*Note.* TMT, Halstead–Reitan Trail Making Test; WCST, Wisconsin Card Sorting Test; PE, perseverative errors; ToL, Tower of London

<sup>a</sup>physical or <sup>b</sup>computerizedversion

workstation while their children completed the test. Testing was completed in a semiprivate area with partitions between computer stations to limit any visual distractions or viewing of the monitors at the adjacent station. The instructions were displayed and read by the experimenter to each participant. All procedures were approved by the Institutional Review Board of Oregon Health and Science University. Each participant completed only one test, and so, consequently, we have no direct correlational measures between performance on different tests. During the data collection period (May–September 2010), data were obtained using eight different PEBL tests, of which four are reported here and two are available elsewhere (Berteau-Pavy & Piper, 2011; Piper & Miller, 2011).

#### PEBL Trail Making Test (pTMT)

The participants ( $N = 384$ ; ages, 5–76 years; 51.3% female; 13.0% left-handed) completed a computerized version modeled generically after the Halstead–Reitan Trail Making Test. The PEBL version uses an automated algorithm to generate the problems, rather than using the specific set of layouts in the Halstead–Reitan test. The test administered was slightly modified from the one contained in version 0.5 of the PEBL battery, so that the same five test forms were used for all participants (instead of being generated randomly). The instructions below were displayed prior to testing:

In this experiment, your goal is to click on each circle, in sequence, as quickly as you can. When you click on the correct circle, its number will change to boldface, and a line will be drawn from the previous circle to the new circle. On some trials, the circles will be numbered from 1 to 25, and you should click on them in numerical order (1–2–3–4). On other trials, the circles will have both numbers (1 to 13) and letters (A through L), and you should click on them in an alternating order (1–A–2–B–3–C). If you click the wrong circle, no line will be drawn. The trial will continue until you have successfully clicked on all of the circles in the correct order. After the display appears, you can examine the circles as long as you want. Timing will not begin until you click on the first circle, which is labeled '1' on every trial.

The pTMT contained ten trails and alternated between five trials with only numbers (Part A) and five trials with alternating numbers and letters (Part B; see Fig. 1a). Each Part A trial had a corresponding Trail B (an isomorphic problem, mirrored along the vertical axis) with an equal distance to connect all the items. The primary dependent measure was the total time to complete each part. The B:A ratio and accuracy, defined as the minimum number of clicks necessary to complete each trial divided by the number made, were also calculated.

#### PEBL Wisconsin (Berg) Card Sorting Test (pWCST)

Participants ( $N = 246$ ; ages, 7–89 years; 45.5% female; 10.7% left-handed) completed a card sorting task (Fig. 2a) modeled after Berg (1948) and described more fully elsewhere (Lyvers & Tobias-Webb, 2010). The instructions were as follows:

You are about to take part in an experiment in which you need to categorize cards based on the pictures appearing on them. To begin, you will see four piles. Each pile has a different number, color, and shape. You will see a series of cards and need to determine which pile each belongs to.... The correct answer depends upon a rule, but you will not know what the rule is. But, we will tell you on each trial whether or not you were correct. Finally, the rule may change during the task, so when it does, you should figure out what the rule is as quickly as possible and change with it. Press any key to begin.

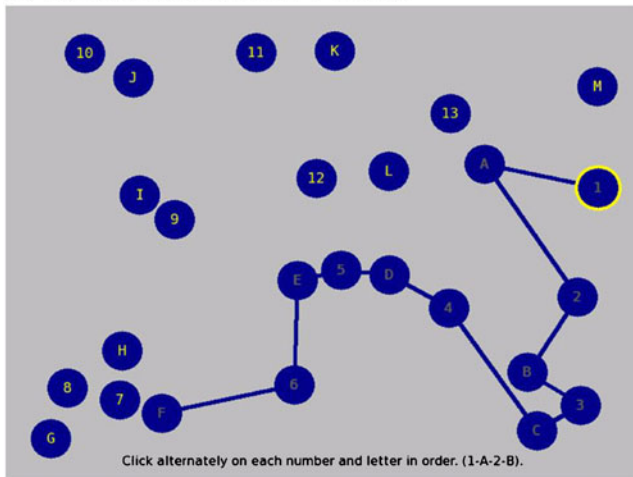
After each trial, feedback of “correct!” or “incorrect” was displayed for 500 ms. The maximum number of trials was 128 (i.e., two decks of 64 cards) but could be shorter (100) on the basis of optimal category completions. The rule (color, shape, or number) could switch as quickly as every tenth trial. The primary dependent measure was the percentage of the total number of trials with perseverative errors. A perseverative error was defined as an incorrect response to a shifted or new category that would have been correct for the immediately preceding category. Response time was also obtained for correct and incorrect decisions for each participant, although excessively short (<100 ms) or long (>10 s) trial times were excluded prior to calculating the mean for each participant.

#### PEBL Tower of London (pTOL)

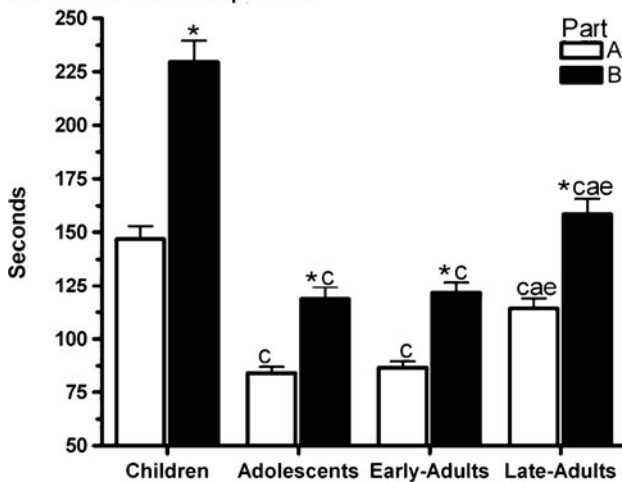
The participants ( $N = 325$ ; ages, 6–82 years; 44.0% female; 12.3% left-handed) completed eight trials with stimuli based on set A from Phillips et al. (1999). The instructions were as follows:

You are about to perform a task called the 'Tower of London'. Your goal is to move a pile of disks from their original configuration to the configuration shown on the top of the screen. You can only move one disk at a time, and you cannot move a disk onto a pile that has no more room (indicated by the size of the grey rectangle). To move a disk, click on the pile you want to move a disk off of, and it will move up into the hand. Then, click on another pile, and the disk will move down to that pile.

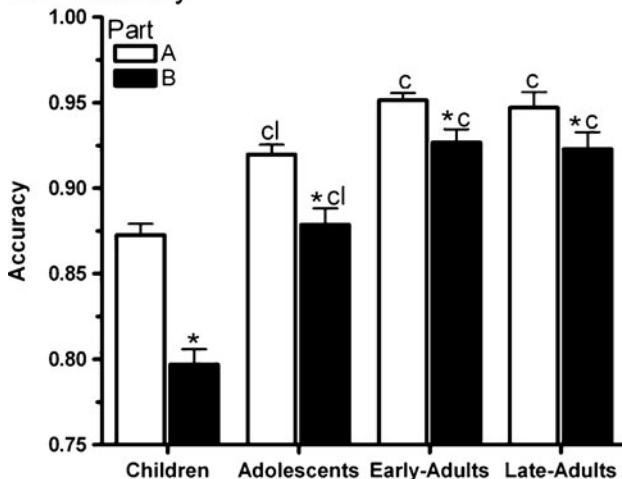
## A. Screen shot from Part B.



## B. Time to complete.



## C. Accuracy



Notably, unlike some versions of the ToL, the version we tested placed no restrictions on the height of the pegs or the number of moves allowed to solve the problem. The

**Fig. 1** PEBL Trail Making Test behavior in children (ages, 5–12 years;  $N = 166$ ), adolescents (ages, 13–19 years;  $N = 84$ ), early adults (ages, 20–49 years;  $N = 95$ ), and late adults (ages, 50–76 years;  $N = 39$ ). Screen shot from a trial in Part B (a), completion time (b), and accuracy (c). \* $p < .0005$  versus Part A; <sup>a</sup> $p < .05$  versus adolescents; <sup>c</sup> $p < .05$  versus early adults; <sup>e</sup> $p < .0005$  versus children; <sup>l</sup> $p < .05$  versus late adults

primary dependent measure was the total number of extra moves across the seven trials (moves made minus 48, the minimum necessary to solve the problems), although the total time was also recorded (Fig. 3).

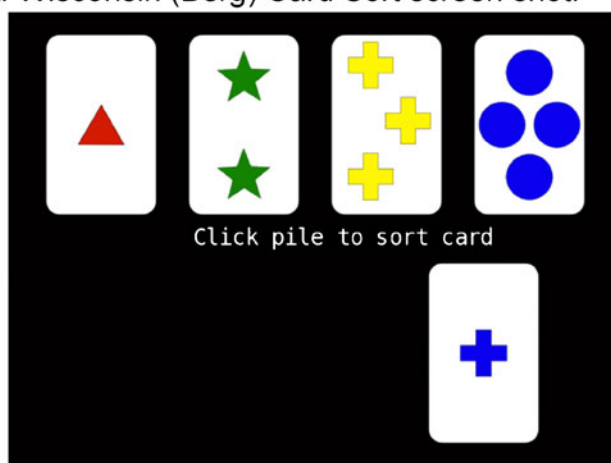
## Time-Wall

The participants ( $N = 268$ ; ages, 5–79 years; 47.8% female; 12.6% left-handed) completed a time estimation task based on Perez et al. (1987). The females in this sample were older ( $25.2 \pm 1.8$  years) than the males ( $19.9 \pm 1.3$  years),  $t(236.6) = 2.42$ ,  $p < .05$ , so sex differences were evaluated with the sample stratified into age categories. The instructions, slightly modified from Snyder and Rice (1990), were as follows:

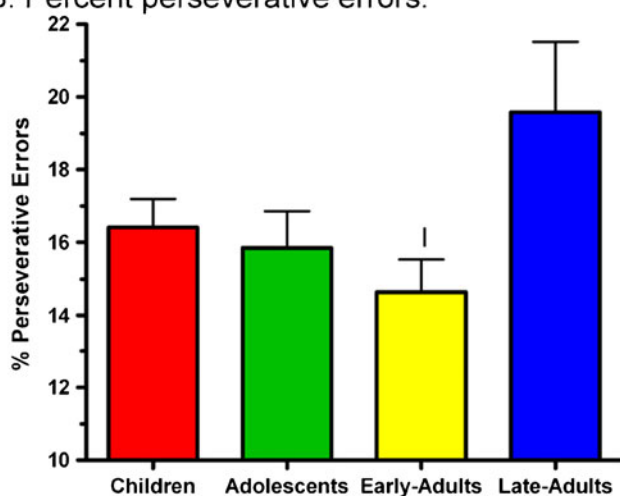
This is an experiment to see how well you can estimate the speed of a moving square target. The target will always start at the top of the screen and descend at a constant rate toward the bottom. After the target is two-thirds of the way down, it will pass behind a wall and become invisible. Your task is to press a button at the exact moment the moving target would pass through the notch marked at the very bottom of the display. In making this judgement, you are not to count or use any other rhythm method to facilitate your judgement. Instead, follow the target with your eyes and imagine it continuing straight down behind the wall to the notch. After you have pressed the button, you will receive feedback as to where the target actually was and whether you over or underestimated the time interval. When you are ready, press a key on the keyboard and the next target shall emerge from the top.

The participants underwent a brief practice, followed by 18 trials on which the correct completion time ranged from 2,000 to 9,200 ms ( $M = 5,822.4$  ms,  $SEM = 558.4$ ). Figure 4a shows a screenshot from Time-Wall. The primary dependent measure was inaccuracy, defined as the absolute value of the participant's response time minus the correct time divided by the correct time for that trial. Since the vast majority of responses on tests of this type are too early (Jerison & Arginteanu, 1958; Jerison et al., 1957), the percentage of trials on which the response time was greater than the correct time was determined. These two values map roughly onto precision and bias, where optimal Time-Wall performance would result in a smaller

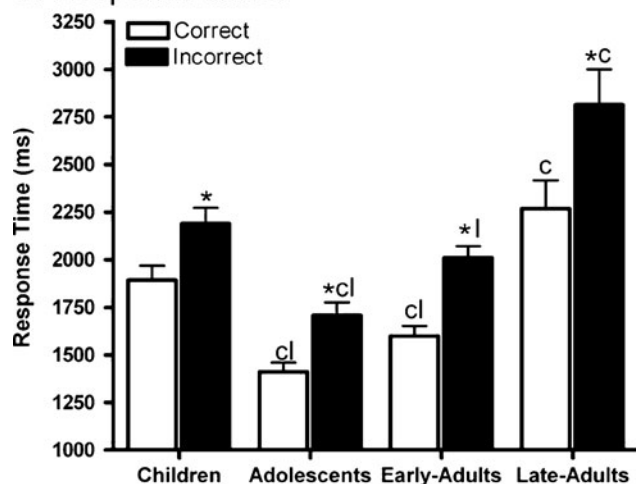
## A. Wisconsin (Berg) Card Sort screen shot.



## B. Percent perseverative errors.



## C. Response times.



values for inaccuracy (ideally, 0.00), and unbiased performance would result in a proportion of late responses close to 50%.

**Fig. 2** Wisconsin Card Sorting (Berg, 1948) performance in children (ages, 7–12 years;  $N = 71$ ), adolescents (ages, 13–19 years;  $N = 63$ ), early adults (ages, 20–49 years;  $N = 81$ ), and late adults (ages, 50–89 years;  $N = 30$ ). **a** Screen shot: The lower card is placed into one of the four piles on the basis of similarity of shape, color, or number. **b** Percentages of perseverative errors by age ( $^l p < .05$  vs. late adults). **c** Response times on correct and incorrect trials by age ( $^* p < .0005$  vs. correct;  $^c p < .05$  vs. children;  $^l p < .05$  vs. late adults)

## Statistical analysis

All analyses were conducted using SPSS, version 16.0 (SPSS Inc., Chicago, IL). Mixed ANOVAs were completed where applicable, with age divided into four groups: children (7 [or whatever the lower limit was for that test] to 12 years), adolescents (13–19 years), early adults (20–49 years), and late adults (50+ years). If Mauchly's sphericity test was significant on repeated measures ANOVAs, results of the Greenhouse–Geisser were reported, with the corresponding reduction in the degrees of freedom. Pearson correlation coefficients were completed among measures on the same tests. Mean data are presented with the *SEM*, and nonlinear regressions depict the 95% confidence intervals. Effect sizes are expressed as partial  $\eta^2$  for ANOVAs and Cohen's *d* for two-group comparisons.

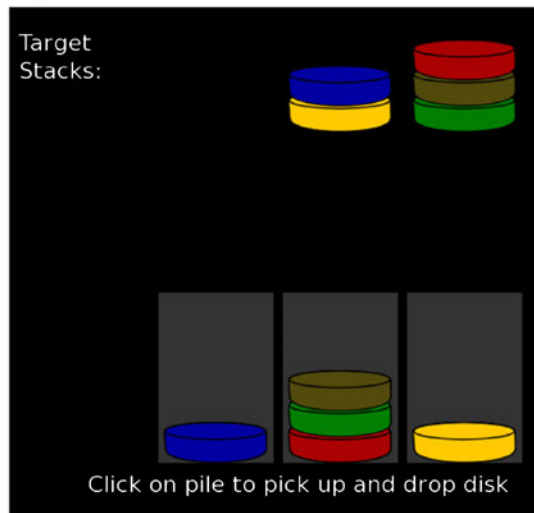
## Results

## PEBL Trail Making Test (pTMT)

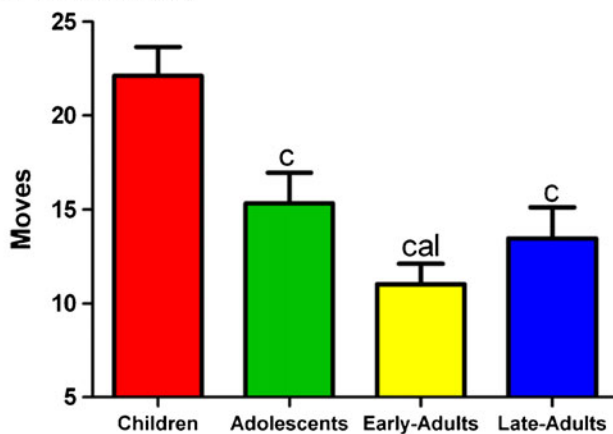
The total time to complete the pTMT was first analyzed with a mixed (within: part, A vs. B; between: age group, 5–12, 13–19, 20–49, vs. 50–76) ANOVA that identified main effects of part,  $F(1,380) = 251.7$ ,  $p < .0005$ ,  $\eta^2 = .40$ , and age,  $F(3,380) = 42.1$ ,  $p < .0005$ ,  $\eta^2 = .25$ , and an age  $\times$  part interaction,  $F(3,380) = 24.0$ ,  $p < .0005$ ,  $\eta^2 = .16$ . Figure 1b shows that, as was anticipated, all ages took longer to complete Part B, relative to A. The duration for children to finish each part was larger than that for all older groups. Late adults had greater A and B times, as compared with adolescents and early adults. Examination of the ratio of Part B: Part A time showed an age effect,  $F(3,380) = 9.92$ ,  $p < .0005$ ,  $\eta^2 = .07$ , with the ratio for children ( $1.566 \pm 0.029$ ) being higher ( $p \leq .01$ ) than that for adolescents ( $1.396 \pm 0.0268$ ,  $d = 0.54$ ), early adults ( $1.394 \pm 0.0195$ ,  $d = 0.58$ ), or late adults ( $1.398 \pm 0.047$ ,  $d = 0.50$ ).

A mixed ANOVA showed main effects of part,  $F(1,380) = 110.6$ ,  $p < .0005$ ,  $\eta^2 = .23$ , and age,  $F(3,380) = 47.93$ ,  $p < .0005$ ,  $\eta^2 = .28$ , and a part  $\times$  age interaction,  $F(3,380) = 14.72$ ,  $p < .0005$ ,  $\eta^2 = .10$ , on accuracy. For all ages, accuracy was lower for Part B, relative to A. Children had lower accuracy, relative to all the older groups, for both

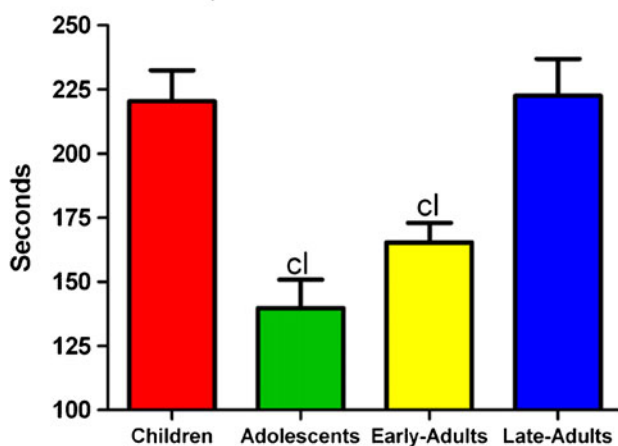
## A. Example Tower of London trial.



## B. Extra-moves.



## C. Time to completion.



**Fig. 3** Tower of London performance in children (ages, 6–12 years;  $N = 118$ ), adolescents (ages, 13–19 years;  $N = 51$ ), early adults (ages, 20–49 years;  $N = 99$ ), and late adults (ages, 50–82 years;  $N = 56$ ). Screen shot (a), extra moves (b), and time (c). <sup>a</sup> $p < .05$  versus adolescents; <sup>c</sup> $p < .01$  versus children; <sup>l</sup> $p < .001$  versus late adults

parts. Finally, late adults had greater accuracy than did adolescents (Fig. 1b). The time to complete Part B was strongly correlated with Part A time (Pearson's  $R(384) = .88$ ,  $p < .0005$ ) and Part B Accuracy ( $R(384) = -.61$ ,  $p < .0005$ ). Similarly, accuracy on Part A was highly associated with that on Part B,  $R(384) = .77$ ,  $p < .0005$ .

## PEBL Wisconsin (Berg) Card Sort Test (pWCST)

Overall, the males in this sample were younger ( $21.8 \pm 1.3$  years) than the females ( $29.8 \pm 1.9$  years),  $t(207.0) = 3.46$ ,  $p \leq .001$ . However, since there were no appreciable effects of sex, analyses did not incorporate this variable. The percentage of perseverative errors was analyzed with a one-way ANOVA, which revealed an age effect,  $F(3,242) = 2.86$ ,  $p < .05$ ,  $\eta^2 = .03$ , with late adults committing more errors than did early adults (Fig. 2b). Table 2 shows that late adults, but not adolescents or early adults, required fewer trials to complete their first category than did children ( $d = 0.61$ ). A larger proportion of early adults, as compared with all the other ages, completed the WCST in less than the full 128 trials. Late adults had fewer correct responses than did either adolescents ( $d = 0.58$ ) or early adults ( $d = 0.65$ ), but their scores were not significantly different from those for the children.

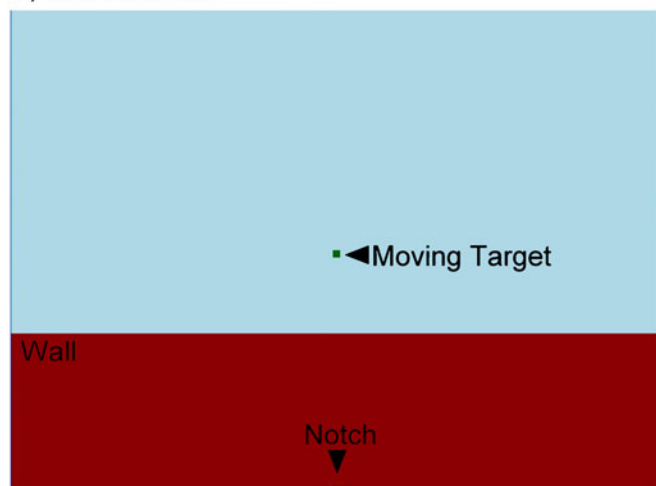
Response time was analyzed with a mixed (within: response type, correct vs. incorrect; between: age), which revealed main effects of age,  $F(3,241) = 21.05$ ,  $p < .0005$ ,  $\eta^2 = .21$ , and response,  $F(1,241) = 234.9$ ,  $p < .0005$ ,  $\eta^2 = .49$ , and an age  $\times$  response interaction,  $F(3,241) = 4.35$ ,  $p \leq .005$ ,  $\eta^2 = .05$ . Figure 2c shows that incorrect responses took longer for all ages. Children's correct responses were slower than those of adolescents ( $d = 0.91$ ) and early adults ( $d = 0.52$ ) but were faster than those of late adults ( $d = 0.51$ ). Similarly, children's incorrect responses took longer than those of adolescents ( $d = 0.58$ ) but were more rapid than those of late adults ( $d = 0.71$ ).

The percent correct was highly inversely related to the percentage of nonperseverative errors,  $R(246) = -.81$ . The percentage of perseverative errors was moderately associated with the percentage of total errors,  $R(246) = .52$ ,  $p < .0005$ , and weakly with percentage of unique errors,  $R(246) = .23$ ,  $p < .0005$ , correct response time,  $R(245) = .23$ ,  $p < .0005$ , and incorrect response time,  $R(245) = .16$ ,  $p < .05$ .

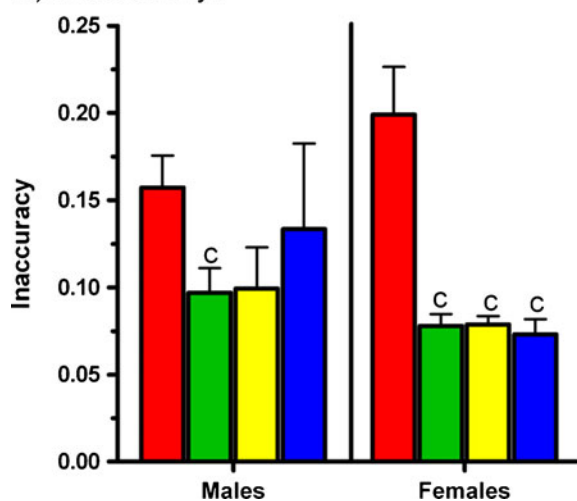
## PEBL Tower of London (pTOL)

The number of extra moves showed a main effect of age,  $F(3,321) = 12.78$ ,  $p < .0005$ ,  $\eta^2 = .11$ , with children making more moves than all other groups ( $d = 0.43$ – $0.80$ ). Furthermore, early adults made fewer moves than did either adolescents ( $d = 0.35$ ) or late adults ( $d = 0.37$ ; Fig. 2b). Further analysis noted the same relative age

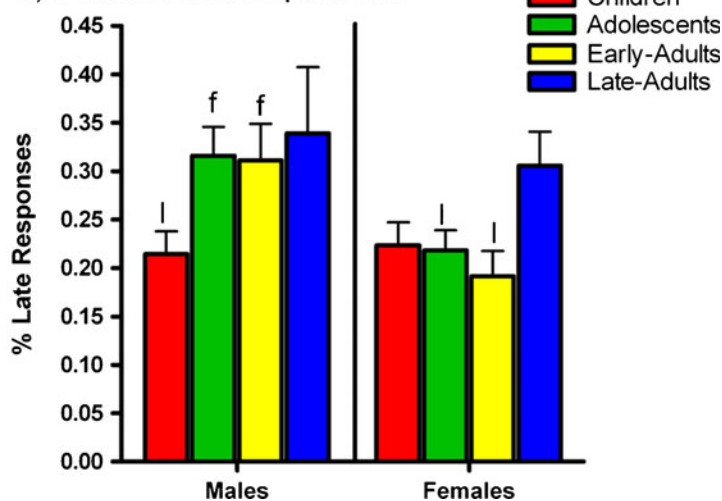
## A) Screenshot.



## B) Inaccuracy.



## C) Percent late responses.



**Fig. 4** Time-Wall screenshot (a, with target, wall, and notch labeled) and performance in children (ages, 5–12 years;  $N = 105$ ), adolescents (ages, 13–19 years;  $N = 73$ ), early adults (ages, 20–49 years;  $N = 59$ ),

and late adults (ages, 50–79 years;  $N = 32$ ). Inaccuracy (b) and percentages of late responses (c) by sex ( $^c p < .05$  vs. children;  $^f p \leq .05$  vs. late adults;  $^f p < .05$  vs. females)

pattern when examining the early (2–4) versus the later (5–8) trials (data not shown). The time to complete all trials was examined with an ANOVA that revealed a main effect of age,  $F(3,324) = 10.78$ ,  $p < .0005$ ,  $\eta^2 = .09$ .

Children and late adults each took longer than either adolescents or early adults but were indistinguishable from each other (Fig. 2c). There was a moderate association between extra moves and time,  $R(324) = .54$ ,  $p < .0005$ .

**Table 2** Wisconsin (Berg) Card Sort Test behavior in children (ages, 7–12;  $N = 71$ ), adolescents (ages, 13–19;  $N = 63$ ), early adults (ages, 20–49;  $N = 82$ ), and late adults (ages, 50–89;  $N = 30$ )

	Children	Adolescents	Early Adults	Late Adults
Complete 1st category (trials)	28.5 (2.6) <sup>a</sup>	25.1 (2.7)	23.3 (2.6)	19.0 (2.3)
Trials (% with <128)	1.4%	6.3% <sup>c</sup>	30.5% <sup>ba</sup>	3.3%
% Correct	68.1 (1.5)	71.9 (1.7) <sup>a</sup>	73.2 (1.6) <sup>a</sup>	63.8 (2.6)
% Nonperseverative errors	15.5 (1.4)	12.2 (1.3)	12.2 (1.3)	16.6 (2.6)
% Unique errors	2.1 (0.4)	1.5 (0.3)	1.0 (0.3)	1.6 (0.4)

<sup>a</sup>  $p < .01$  versus late adults; <sup>b</sup>  $p < .0005$  versus children; <sup>c</sup>  $p < .0005$  versus early adults



## Time-Wall

On average, time estimations were  $289.1 \pm 33.4$  ms early, with  $25.6\% \pm 1.1\%$  of the trials having a late response. The average inaccuracy was  $12.3\% \pm 0.8\%$ . A sex  $\times$  age ANOVA on percentage of late responses revealed a sex effect,  $F(1,261) = 6.66, p \leq .01, \eta^2 = .03$ , an age effect,  $F(3,261) = 3.02, p < .05, \eta^2 = .03$ , and a trend in the interaction,  $F(3,261) = 2.39, p = .069$ . Late adult females had more late responses than did either adolescents ( $d = 0.62$ ) or early adults ( $d = 0.74$ ). Among males, children had fewer late responses than did late adults ( $d = 0.63$ ). There was a sex difference favoring males at the adolescent ( $d = 0.65$ ) and early adult ( $d = 0.68$ ; Fig. 1b) ages. There was only an age effect,  $F(3,259) = 10.75, p < .0005, \eta^2 = .11$ , for inaccuracy. Girls were more inaccurate than all the older females ( $d = 0.89\text{--}0.92$ ), but boys showed a difference only relative to the adolescents ( $d = 0.53$ ). The correlation of the full-length version (18 trials) with an even shorter version (10 trials) was quite high for percentage of late trials,  $R(269) = .88, p < .005$ , and inaccuracy,  $R(267) = .97, p < .0005$ , indicating that future researchers could adopt the shortened 10-trial version of the test.

## Discussion

Our primary objective was to evaluate the validity of the PEBL tests. The prediction, based on Table 1, was of a nonlinear relationship between age and performance on measures of executive function. This was partially supported, particularly for the pTMT and the pWCST. These findings also suggest that these PEBL tests show substantial similarities with their more traditional test counterparts. The secondary hypothesis of a dissimilar profile across executive function tests was supported, especially at younger ages. Prior research with a college-aged sample completing multiple tests determined that executive function is not a unitary process (Miyake et al., 2000), and the present dataset corroborates that conclusion. The findings with each of these four measures, as well as some comments regarding their predecessor versions, are discussed below.

### pTMT

The original Halstead–Reitan TMT (henceforth referred to as the Reitan TMT for simplicity) and the pTMT, despite substantial procedural overlap, also differ in some key details. First, the Reitan TMT requires that the experimenter be constantly vigilant to notice any participant errors in connecting the dots and to imme-

diately redirect them, whereas the pTMT reduces this source of variance considerably by automatically recording the number of errors and by requiring that the correct target be identified before continuing. Second, the total distance in each pair (e.g., A1 vs. B1) is equivalent in the pTMT. In contrast, the total distance in the Reitan TMT is 57 cm longer for Part B than for A (Gaudino, Geisler, & Squires, 1995), the overall trail pattern follows a distinctly different path, and this longer distance artificially inflates the completion time between Parts B and A. However, despite removing this bias, the B version took significantly longer than the A for each age group on the pTMT. Third, PEBL's use of algorithmic generation of TMT problems allows multiple tests of each form to be completed, thereby increasing reliability. Fourth, the Reitan TMT includes separate versions for different ages (<15 vs. 15+), whereas a single test was easily completed across a wide age range in the present endeavor. In a separate study (Piper et al., 2010), we discovered that having separate test versions for different ages can be an inadvertent source of variance. Fourth, and perhaps most important, the Reitan TMT is completed using a pencil, whereas the pTMT involves a computer mouse. The possibility certainly exists that individuals who have greater computer experience would complete the pTMT more rapidly, but utilization of the B:A ratio eliminates this factor. More specifically, although a U-shaped function with the B completion time was observed, with an improvement during childhood and a slight, but significant, regression for late adults (Fig. 1b; see Tombaugh, 2004, for similar findings in adults), the B:A ratio removed any differences between late adults and early adults or adolescents. Further efforts at test development could also incorporate touch screen monitors with the PEBL battery to further diminish any preexisting individual differences in computer experience.

A secondary measure on the pTMT was accuracy [correct clicks/(correct + incorrect)]. The finding that all age groups had reduced accuracy on Part B, relative to A, is not unexpected. However, the higher accuracy of the late adults on Part B, relative to both children and adolescents, is remarkable, since this indicates that late adults may adopt a strategy of more careful deliberation prior to responding. This behavioral pattern, coupled with the generalized slowing among the elderly (Mahncke et al., 2006), may have had a differential impact on the performance of the remaining tests (described below).

### pWCST

The primary dependent measure on the pWCST, as well as the traditional (physical cards) WCST, is the percentage of

perseverative errors. The contribution of age to perseverative errors was rather subtle, particularly when placed within the context of robust age differences in response times. As compared with early adults, the response time for late adults was approximately 700 ms longer (41.8%) for correct responses and 1,100 ms longer for (55.0%) for incorrect responses. Children, adolescents, and early adults had equivalent performance, but late adults made more perseverative errors than did early adults. Examination of the secondary measures (Table 1) revealed that late adults differed from other age groups in a no uniform fashion. Hartman et al. (2001) have quite clearly documented that working memory, the limited capacity to temporarily store and process information, is essential for an individual to complete the WCST, because they must store information from completed sorts and process new information in order to identify the subsequent rule. The finding that older adults had fewer correct responses than did adolescents or early adults might suggest that the longer response times on each trial, when multiplied by the number of trials necessary to complete each category, would place much higher demands on working memory and that this factor, either independently or in conjunction with cognitive inflexibility, may account for age differences in this group. Overall, the present pattern of age differences with the pWCST is strongly congruent with prior reports (Haaland, Vranes, Goodwin, & Garry, 1987; Strauss et al., 2006).

There are a wide variety of existing computerized card sorting tests that share conceptual similarities. This includes a test within the CANTAB (the Intra-Extra Dimensional Set Shift, the instrument utilized by Huizinga et al., 2006), and the Wisconsin Card Sorting Test distributed by Psychological Assessment Resources (PAR, 2003). The PARWCST and the pWCST are each based on the procedures outlined in Berg (1948). However, there is at least one difference in that the feedback provided is auditory in the PAR test but visual in the pWCST. A direct comparison where participants complete the short (64-item) forms of both tests is needed to further clarify the extent that both instruments are measuring the same constructs.

#### pTOL

The number of extra moves is consistent with the developmental profile hypothesized a priori. More specifically, children made more moves relative to adolescents, who, in turn, made more moves than did early adults but were indistinguishable from late adults. These age differences are even more informative when placed within the context of completion times, with adolescents showing the shortest times of any age

group. Together, the uncoupling of moves and time may indicate inferior planning and decision-making efficiency in adolescents. This pattern of responding prior to fully considering all options could also be consistent with elevated adolescent impulsivity on the pToL.

At least two other neuropsychological batteries contain ToL-type tests that are worthy of mention. The ToL freely provided by Davis and colleagues within the Colorado Assessment Tests (<http://www.catstests.com/>) includes both visual and audio instructions (Davis & Keller, 1998). The CANTAB Stockings of Cambridge is essentially a ToL task and offers detailed normative data (DeLuca et al. 2003; Luciana, Collins, Olson, & Schissel, 2009) that parallel the present findings and is highly engaging for participants. One potential advantage of the pToL is its versatility. In addition to the trial structure used in the present study, nine others of different lengths and complexities (Phillips et al., 1999; Shallice, 1982) are available, and it can be further adapted to test new problem sets with minimal changes.

#### Time-Wall

In comparison with the other measures contained in this report, Time-Wall is a relatively obscure test. However, a few general findings are worthy of mention. An adult like response style was observed among adolescents for inaccuracy. The general response approach for all ages was to respond too early. Importantly, there was no evidence of inferior performance by late adults. In fact, late adult females were more likely than either adolescents or early adults to have late, rather than early, responses. These findings might suggest that the generalized cognitive slowing among the elderly (Mahncke et al., 2006) may confer some slight benefit on this simple decision-making test, which is quite different from the profile with the pWCST and pTMT.

#### Summary and conclusions

Overall, many studies examining changes in cognitive skill between childhood and senescence have been completed previously (e.g., Kail & Salthouse, 1994; Luciana et al., 2009; Mahncke et al., 2006; Park, 2000). However, the earlier data are often procured using unreleased special-purpose software and measures or with proprietary or obsolete software that is not available to many behavioral neuroscientists and neuropsychologists, especially those working in settings where costly software licenses are not an easily justifiable expense. This lifespan study may also provide useful normative behavior of age-related changes

on standardized tests (Appendix Tables 3, 4, 5, 6). The benefits of characterizing normative behavior are two-fold: First, this provides a previously unknown profile of age-related changes across standardized neurocognitive measures, and second, it does so for tasks whose software is freely available for use and modification by neurobehavioral and clinical researchers.

In conclusion, the studies reported here may serve as a cornerstone for further investigations into executive function across the lifespan, and this represents a first major step in providing norms for a set of data collection tools that are freely available, verifiable, modifiable, and redistributable. Together, these norms both bolster the tenets of the scientific method, by enabling better transparency,

replication, and exchange of information, and provide important normative data for using the software in applied testing contexts that proprietary tests have not reached because of their costs.

**Acknowledgements** We would like to thank Jacob Raber, Ph.D., Andrea Middleton, and the staff at the Oregon Museum of Science and Industry and the OHSU Center for Diversity and Multicultural Affairs, as well as William Cameron, Ph.D., and the OHSU Methamphetamine Abuse Research Center, for their support of this project. This research was supported by Public Health Service Grant 1 UL1 RR024120-01, the National Institute of Drug Abuse (L30 DA027582-01 and 1P50DA018165), the National Institute of Environmental Health Sciences (T32 ES007060), the Lewis and Clark HHMI Undergraduate Science Education Program, and the N.L. Tartar Trust.

## Appendix

**Table 3** Percentiles (10, 50, and 90) by age for the Psychology Experiment Building Language Trail Making Test for the time (seconds) to complete Part A and Part B and the B to A ratio

Age	N	Part A	Part B	B:A
		10 (50) 90	10 (50) 90	10 (50) 90
5–6	20	169.7 (265.0) 448.8	182.2 (447.0) 617.4	1.00 (1.68) 2.61
7	27	126.5 (166.4) 240.1	203.6 (284.5) 443.1	1.21 (1.64) 2.50
8	17	102.1 (138.5) 294.8	141.9 (237.2) 394.2	1.17 (1.47) 1.93
9	27	90.5 (116.5) 174.8	129.6 (183.3) 306.9	1.24 (1.50) 2.34
10	29	83.4 (113.9) 150.6	121.7 (167.8) 249.3	1.18 (1.51) 1.86
11	26	76.2 (100.0) 127.4	103.1(141.5) 200.4	1.05 (1.44) 1.93
12	20	63.0 (92.4) 129.3	78.3 (122.5) 167.8	1.08 (1.32) 1.65
13–14	38	66.7 (83.4) 107.9	93.1 (116.9) 169.5	1.22 (1.44) 1.69
15–19	46	58.8 (77.5) 99.4	69.9 (100.3) 147.2	1.09 (1.29) 1.61
20–29	22	63.5 (73.9) 112.0	79.2 (97.3) 162.5	1.20 (1.33) 1.69
30–39	30	68.5 (83.4) 104.4	81.9 (109.2) 168.1	1.14 (1.34) 1.69
40–49	43	68.2 (84.1) 114.6	87.9 (115.3) 168.3	1.15 (1.45) 1.63
50–59	13	82.4 (93.8) 119.5	99.4 (128.6) 161.0	1.14 (1.29) 1.68
60–76	26	91.0 (114.2) 180.6	116.8 (172.7) 239.6	1.02 (1.38) 1.91

**Table 4** Percentiles (10, 50, and 90) by age for the Psychology Experiment Building Language Wisconsin (Berg) Card Sorting Test for the trials to complete the first set, percentages of errors, and the percentages of perseverative errors

Age	N	Trials to Complete 1st Set	% Errors	Perseverative Errors
		10 (50) 90	10 (50) 90	10 (50) 90
7–9	28	10.0 (21.0) 70.1	20.3 (28.9) 50.8	10.9 (16.4) 25.2
10–12	43	11.4 (19.0) 55.6	17.5 (25.8) 49.7	8.1 (13.3) 28.3
13–14	34	11.0 (18.5) 74.0	13.9 (22.7) 53.9	6.6 (12.1) 23.0
15–19	29	11.0 (15.0) 53.0	14.8 (26.6) 47.7	10.7 (15.6) 28.1
20–29	29	10.0 (12.0) 30.0	13.0 (20.3) 38.3	8.6 (13.3) 28.1
30–39	24	10.0 (12.0) 103.0	11.1 (19.5) 55.1	5.2 (10.6) 30.9
40–49	29	10.0 (17.0) 57.0	14.4 (26.6) 59.4	8.1 (13.3) 28.1
50–89	30	10.0 (15.5) 39.7	16.6 (34.8) 57.7	9.4 (19.5) 28.8

**Table 5** Percentiles (10, 50, and 90) by age for the Psychology Experiment Building Language Tower of London test for the number of extra moves and time (seconds)

Age	N	Moves			Time		
		10	50	90	10	50	90
6–8	44	7.5	27.0	57.0	150.6	265.0	503.2
9–10	42	6.0	15.5	33.0	113.1	161.6	272.1
11–12	32	5.0	13.0	37.7	98.4	135.6	224.4
13–15	30	5.1	15.0	31.9	89.6	116.9	203.9
16–19	22	4.3	8.5	22.0	73.0	120.4	245.9
20–29	27	3.8	10.0	20.4	83.8	127.1	227.4
30–39	40	2.1	8.0	23.0	95.8	150.7	248.8
40–49	32	3.0	10.0	20.1	110.6	165.2	260.1
50–59	20	2.4	11.5	37.9	106.7	159.5	293.0
60–82	36	4.7	13.0	25.9	138.7	216.4	443.6

**Table 6** Percentiles (10, 50, and 90) by age and sex for the Psychology Experiment Building Language Time-Wall for inaccuracy ( $\times 100$ ) and percentages of late responses (LRs)

Age	N	Males		N	Females	
		Inaccuracy	% LR		Inaccuracy	% LR
5–9	28	6.4 (13.6) 21.1	5.6 (16.7) 44.4	29	7.5 (15.5) 49.2	5.6 (22.2) 55.6
10–13	36	6.0 (9.0) 49.4	5.6 (19.4) 57.2	27	6.5 (9.4) 57.8	5.6 (16.7) 46.7
14–19	36	4.4 (6.7) 14.3	11.1 (36.1) 64.4	21	5.0 (6.3) 15.9	5.6 (22.2) 38.9
20–49	30	3.5 (6.0) 26.0	5.6 (30.6) 65.6	29	4.7 (7.7) 11.5	5.6 (16.7) 38.9
50–79	10	3.2 (8.0) 51.7	6.1 (33.3) 71.1	22	3.9 (6.4) 14.1	8.9 (25.0) 59.4

## References

- Army Individual Test Battery. (1944). *Manual of direction and scoring*. Washington, DC: War Department Adjutant General's Office.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15–22.
- Berteau-Pavy, D., & Piper, B. J. (2011). *Contributions of age, but not sex, to mental rotation performance in a community sample* (PEBL Tech. Rep. 2011–02), Retrieved from <http://sites.google.com/site/pebltechnicalreports/home/pebl-technical-report-2011-02>
- Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *8*, 219–228.
- Conners, C. K. (1985). The computerized continuous performance test. *Psychopharmacology Bulletin*, *21*, 891–892.
- Davidson, P., Gao, F., Mason, W., Winocur, G., & Anderson, N. D. (2008). Verbal fluency, Trail Making, and Wisconsin Card Sorting Test performance following right frontal lobe tumor resection. *Journal of Clinical and Experimental Neuropsychology*, *30*, 18–32.
- Davis, H. P., & Keller, F. R. (1998). *Colorado assessment test manual*. Colorado Springs: Colorado Assessment Tests.
- DeLuca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., et al. (2003). Normative data from the CANTAB: I. Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology*, *25*, 242–254.
- Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, *26*, 874–890.
- Fray, P. J., & Robbins, T. W. (1996). CANTAB battery: Proposed utility in neurotoxicology. *Neurotoxicology and Teratology*, *18*, 499–504.
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60.
- Gaudino, E. A., Geisler, M. W., Squires, N. K. (1995). Construct validity in the Trail Making Test: What makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, *17*, 529–535.
- Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*, 404–411.
- Greenberg, L. M., & Waldman, I. D. (1993). Developmental normative data on the test of variables of attention (T.O.V.A.™). *Journal of Child Psychology and Psychiatry*, *34*, 1019–1030.

- Guimarães, C. A., Li, L. M., Rzezak, P., Fuentes, D., Franzon, R. C., Montenegro, M. A., et al. (2007). Temporal lobe epilepsy in childhood: Comprehensive neuropsychological assessment. *Journal of Child Neurology*, *22*, 836–840.
- Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., et al. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, *187*, 254–262.
- Haaland, K. Y., Vranes, L. F., Goodwin, J. S., & Garry, P. J. (1987). Wisconsin Card Sort Test performance in a healthy elderly population. *Journal of Gerontology*, *42*, 345–346.
- Hänninen, T., Hallikainen, M., Koivisto, K., Partanen, K., Laakso, M. P., Riekkinen, P. J., et al. (1997). Decline of frontal lobe functions in subjects with age-associated memory impairment. *Neurology*, *48*, 148–153.
- Hartman, M., Bolton, E., & Fehnel, S. E. (2001). Accounting for age differences on the Wisconsin Card Sorting Test: Decreased working memory, not inflexibility. *Psychology of Aging*, *16*, 385–399.
- Huizinga, M., Dolan, C. V., & Van der Molan, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017–2036.
- Jerison, H. H., & Arginteanu, J. (1958). *Time judgements, acoustic noise, and judgement drift* (WADC-TR-57-454). Wright-Patterson Air Force Base, OH: Wright Air Development Center (AD-12093).
- Jerison, H. H., Crannell, C. W., & Pownall, D. (1957). *Acoustic noise and repeated time judgements in a visual movement protection task* (Tech. Rep. WADC-TR-57-54). Wright-Patterson Air Force Base, OH: Wright Air Development Center.
- Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199–225.
- Korkman, M., Kemp, S. L., & Kirk, U. (2001). Effects of age on neurocognitive measures of children ages 5 to 12: A cross-sectional study on 800 children from the United States. *Developmental Neuropsychology*, *20*, 331–354.
- Kubo, M., Shoshi, C., Kitawaki, T., Takemoto, R., Kinugasa, K., Yoshida, H. C., et al. (2009). Increase in prefrontal cortex blood flow during the computer version Trail Making Test. *Neuropsychobiology*, *58*, 200–210.
- Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59–90.
- Luciana, M., Collins, P. F., Olson, E. A., & Schissel, A. M. (2009). Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity. *Developmental Neuropsychology*, *34*, 461–475.
- Lyvers, M., & Tobias-Webb, J. (2010). Effects of acute alcohol consumption on executive cognitive functioning in naturalistic settings. *Addictive Behaviors*, *35*, 1021–1028.
- Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Brain plasticity and functional losses in the aged: Scientific bases for a novel intervention. *Progress in Brain Research*, *157*, 81–109.
- Malloy-Diniz, L. F., Cardoso-Martins, C., Nassif, E. P., Levy, A. M., Leite, W. B., & Fuentes, D. (2008). Planning abilities of children aged 4 years and 9 months to 8½ years. *Dementia & Neuropsychologia*, *2*, 26–30.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Mueller, S. T. (2010a). A partial implementation of the BICA cognitive decathlon using the Psychology Experiment Building Language (PEBL). *International Journal of Machine Consciousness*, *2*, 273–288.
- Mueller, S. T. (2010b). *The PEBL manual: Programming and usage guide for the Psychology Experiment Building Language PEBL, Version 0.11*. Raleigh, NC: Lulu Press. ISBN 978-0-557-65817-6.
- Mueller, S. T. (2010c). PEBL: The psychology experiment building language, Version 0.10. Retrieved from PEBL Website: <http://pebl.sourceforge.net>
- O’Leary, M. R., Radford, L. M., Chaney, E. F., & Schau, E. J. (1977). Assessment of cognitive recovery in alcoholics by use of the trail-making test. *Journal of Clinical Psychology*, *33*, 579–582.
- Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. *Cognitive Aging: A Primer*, *11*, 3–19.
- Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and methodology. DTIC Document #ADA181697. Retrieved from <http://stinet.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA181697>
- Periáñez, J. A., Rios-Lago, M., Rodríguez-Sánchez, J. M., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., et al. (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology*, *22*, 433–447.
- Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*, 209–231.
- Piper, B. J. (2011). Age, handedness, and sex contribute to fine motor behavior in children. *Journal of Neuroscience Methods*, *195*, 88–91.
- Piper, B. J., Acevedo, S. F., Craytor, M. J., Murray, P. W., & Raber, J. (2010). The use and validation of the spatial navigation Memory Island test in primary school children. *Behavioural Brain Research*, *210*, 257–262.
- Piper, B. J., & Miller, C. (2011). *Preliminary behavioral observations with the PEBL “Hunt” test across the lifespan* (PEBL Tech. Rep. 2011–03). Retrieved from <http://sites.google.com/site/pebltechnicalreports/home/pebl-technical-report-2011-03>
- Psychological Assessment Resources. (2003). *Computerized Wisconsin Card Sort Task Version 4 (WCST)*. Lutz, FL: PAR.
- Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*, *19*, 393–394.
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.
- Schall, U., Johnston, P., Lagopoulos, J., Jüptner, M., Jentzen, W., Thienel, R., et al. (2003). Functional brain maps of Tower of London performance: A positron emission tomography and functional magnetic resonance imaging study. *NeuroImage*, *20*, 1154–1161.
- Sekuler, R., Watamaniuk, S. N. J., & Blake, R. (2002). Motion perception. In *Stevens's handbook of experimental psychology* (3rd ed.). New York: Wiley.
- Shad, M. U., Tamminga, C. A., Cullum, M., Haas, G. L., & Keshavan, M. S. (2006). Insight and frontal cortical function in schizophrenia: A review. *Schizophrenia Research*, *86*, 54–70.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B*, *298*, 199–209.

- Snyder, H. L., & Rice, V.J.B. (1990). *Complex cognitive performance and antihistamine use* Technical report for U.S. Army Medical Research and Development Command, Ft. Detrick, MD, Contract #DAMDI7-89-C-9073. DTIC Document #AD-A244 602.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological test administration, norms, and commentary* (3rd ed.). New York: Oxford.
- Stuss, D., Bisschop, S., Alexander, M., Levine, B., Katz, D., & Izukawa, D. (2001). The Trail Making Test: A study in focal lesion patients. *Psychological Assessment, 13*, 230–239.
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*, 203–214.
- Tsuchiya, E., Oki, J., Yahara, N., & Fujieda, K. (2005). Computerized version of the Wisconsin card sorting test in children with high-functioning autistic disorder or attention-deficit/hyperactivity disorder. *Brain and Development, 27*, 233–236.
- Ward, G., & Allport, A. (1997). Planning and problem solving using the five-disc Tower of London task. *Quarterly Journal of Experimental Psychology, 50A*, 49–78.
- Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's and Dementia, 4*, 428–437.
- Yeudall, L. T., Reddon, J. R., Gill, D. M., & Stefanyk, W. O. (1987). Normative data for the Halstead–Reitan neuropsychological tests stratified by age and sex. *Journal of Clinical Psychology, 43*, 346–367.
- Zakzanis, K. K., Mraz, R., & Graham, S. J. (2005). An fMRI study of the Trail Making Test. *Neuropsychologia, 43*, 1878–1886.