

Journal of Clinical and Experimental Neuropsychology, in press

An interactive test of serial behavior: Age and practice alter executive function

Andrew Kuhlman¹, Deborah Little², & Robert Sekuler¹

¹Volen Center, Brandeis University; Waltham, MA

²Center for Stroke Research, University of Illinois, Chicago, IL

Address correspondence:

Robert Sekuler
Volen Center for Complex Systems
Brandeis University, MS 013
Waltham, MA 02454
Phone: 781-736-3277
Fax: 781-736-2398
e-mail: vision@brandeis.edu

ABSTRACT

We describe an interactive, computer-based test inspired by the Trail Making Test (TMT). In this new test, young (mean age=20.4 years) and older participants (mean age=74.9 years) used natural, pointing responses to order series of numbers, letters, or intermixed letters and numbers. This interactive test, which avoids several deficiencies of TMT, assesses participants' baseline speed for detecting and responding to individual items, captures the time for each response, and segregates erroneous responses from correct ones. The inter-response times with intermixed letters and numbers showed that participants did not always switch between the two different ordering-tasks, but instead often recall letter-number pairs as single units. An additive factors analysis decomposed test times into functional components, including executive function, which took longer in older participants. With modest practice, both young and older participants sped up their ordering of intermixed numbers and letters, probably reflecting increased automaticity and reduced dependence upon executive function.

An interactive test of serial behavior: Age and practice alter executive function

The Trail Making Test (TMT), a mainstay of neuropsychological evaluation (Reitan, 1958), requires participants to draw a single, continuous line (a trail) through randomly-located items on a sheet of paper. In TMT's Form A, the trail is drawn through numbers, which are to be connected in increasing numerical order, 1, 2, 3, . . . 25. In form B, the trail is drawn through spatially intermingled numbers and letters, alternating between the two kinds of items, 1, A, 2, B, 3, C, . . . 12, L, 13. The total time to complete each form is measured, and performance summarized by scoring each test separately, computing the difference between the two total times, or by their ratio (Arbuthnott & Frank, 2000; Drane, Yuspeh, Huthwaite, & Klingler, 2002).

In part, TMT's wide use reflects its presumed sensitivity to executive function. More specifically, differential performance on TMT's two forms has been linked to the ability to complete complex executive functions such as planning of actions and switching between tasks. Because normal aging slows performance on TMT's Form B relative to Form A (Davies, Spelman, & Davies, 1981; Hays, 1995), TMT might be a useful indicator for early age-related changes in the frontal lobe, particularly the dorsolateral prefrontal cortex, and other regions of the brain crucial for executive functions (Cowell et al., 1994; Gunning-Dixon & Raz, 2000). In fact, major accounts of cognitive aging emphasize behavioral changes thought to be associated with the frontal lobe, including reductions in working memory (Wingfield, Lindfield, & Kahana, 1998) and widespread weakening of inhibitory processes (Hasher & Zacks, 1988). Moreover, there is growing consensus that various aspects of executive function depend crucially upon prefrontal cortex (Shallice, 2002).

In order to dissect age-related changes into functional components, Salthouse and colleagues implemented two different variants of TMT. Both afforded the opportunity to analyze the times associated with each individual response, which is not possible in the original test. With one test variant, participants used a keyboard to navigate through a matrix of targets on a computer display (Salthouse & Fristoe, 1995). In the other, participants drew a trail through a simplified, paper and pencil version of TMT (Salthouse et al., 2000). Both studies showed age-related slowing on the mixed letter-number versions of TMT. The application of a structural equation model suggested that the age effects might not come from a factor unique to Form B, but rather might have come from an exaggeration of some factor that also influenced Form A. Specifically, Salthouse et al. (2000) proposed that age-related slowing on TMT was attributable to slowed perceptual processing, that is, in distinguishing among visually similar items. They argued that there was little sign of age-related effects on executive function.

The apparent contradictions in the literature stimulated us to take a more direct approach to understanding the origin of age-related slowing on TMT, including possible contributions from age-related changes in visual perception (Salthouse et al., 2000). In the course of our work, we devised computer-based interactive tests, which in the aggregate we call the Massachusetts Ordered Behavior test (MOB_t). We then used MOB_t to identify the basic components of task switching operations in TMT, and examined the roots of age-related changes in task switching. To examine the impact of age-related perceptual changes, we directly manipulated the visibility of target items, and measured how this manipulation affected different ordering behaviors.

We were particularly interested in isolating the contribution of executive function to MOB_t . In the literature, the term "executive function" has been used in various ways. Here, however, it refers to operations that are required to alternate between different tasks (Meyer & Kieras, 1997; Duncan, 1986; Norman & Shallice, 1986), e.g., searching for an item in an ordered series of numbers versus searching for an item in an ordered series of letters. Task

switching has generally been investigated by having participants switch between two or more well-learned and practiced tasks. Efficiency of task switching is generally reported as the cost (increased response time and/or increased errors) associated with having to switch between the two tasks (Monsell, 2003). Although these costs are reduced when the switch is predictable, residual negative impacts on behavior remain (Monsell, Sumner, & Waters, 2003), and these impacts seem to correlate with performance on TMT (Arbuthnott & Frank, 2000). The underlying assumption in TMT is that participants must maintain both strings of well learned information (letters and numbers) and switch between the two to identify the next target required. It is believed that this switch between tasks lengthens response times in B relative to Form A. MOB_t has a comparative advantage over TMT for determining executive function's contribution to age-related changes in performance because it allows performance to be evaluated from both the aggregate time to complete an entire set of items, as in TMT, and from the individual times collected for each target. Examining performance on individual trials affords greater analytical and statistical power as compared to TMT. In addition, the collection of these individual trial times allow for the directed evaluation of executive functions, and changes in executive function over time, that occur between items, such as the case in switching between letters and numbers.

In designing MOB_t we sought to rectify various deficiencies in TMT (Spren & Strauss, 1998), including several that seem not to have been discussed previously. The following paragraphs briefly describe the flaws we discerned in TMT that might exaggerate age-related effects on downstream measures of task-switching, and the steps MOB_t took to rectify these flaws. The goal was an uncontaminated estimate of the overhead associated with executive function.

Two classes of deficiencies were of special concern. The first class included administrative and methodological details; the second class included higher-order cognitive functions, aside from task-switching, which can affect performance on TMT. These cognitive functions include visual search, spatial memory, and learning.

Methodological Factors

Our primary concern with the interpretation of data collected from the TMT is that erroneous responses can not be examined. In TMT, errors are made when the trail is drawn to an incorrect item. Participants are instructed to detect and then correct their own errors, or, failing that, the test administrator is supposed to guide necessary corrections. Although errors affect the time for completing a trail, frequency of errors is not included in the test score. By recording each individual response, MOB_t eliminates this problem. In addition to an automated error correction scheme, MOB_t allows any error's effects to be investigated, characterized, and expunged as appropriate.

However advantageous computerized administration may be, there is concern that this approach might negatively affect particular populations, such as older adults, who may have limited access and experience with computers. To counteract any disadvantage that older participants might experience with a test's automated computer administration, responses were made using a touch-sensitive screen fitted over the computer display. Participants responded by touching the screen with the index finger of their preferred hand, executing a simple, natural, and highly-practiced response. Moreover, before testing began each participant had a comprehensive, automated practice tutorial on each of MOB_t's sub-tests. An additional brief practice session was offered before every test.

Finally, TMT's original format does not allow for multiple administrations within a short period of time. Successive administrations of TMT could be contaminated by implicit learning of target locations (Chun & Jiang, 1998; Musen & Treisman, 1990), and these effects of learning might be different across individuals or groups. MOB_t nullified the value

of learning particular sequences of locations: item locations were re-randomized after every response. Repeated testing also allowed us to assess how practice might influence the various cognitive components required by MOB_i.

Cognitive Factors

Of particular concern were aspects of TMT that might differentially affect young and older participants in a way that would be mistaken for an age-related change in executive function. For example, it is known that the efficiency of visual search declines with increasing age (Plude & Hoyer, 1981, 1996; Hasher, Stoltzfus, Zacks, & Rypma, 1991; Salthouse, 1993). Such age-related declines in search speed are magnified when potential targets cover relatively large areas (Burton-Danner, Owsley, & Jackson, 2001). At normal viewing distances, TMT requires visual search of an area of about 1200 square degrees of visual angle, and a recently introduced version of TMT-B requires search of an even larger area (Delis, Kaplan, & Kramer, 2001). This search requires a series of saccadic eye movements and fixations among items separated from one another by variable distances and directions (Araujo, Kowler, & Pavel, 2001). As a result of its format, TMT's results could be distorted by individual differences in efficiency and speed of search (Crowe, 1998; Ehrenstein, Heister, & Cohen, 1982; Gaudino, Geisler, & Squires, 1995). In contrast, MOB_i's display spanned less than one-fourth the area covered by items in TMT, and MOB_i's targets occupied equidistant, consistent and well-separated locations. MOB_i's layout of items could also reduce age-related changes from visual clutter which older people find especially troublesome (Kline et al., 1992; Kosnik, Winslow, Kline, Rasinski, & Sekuler, 1988; Kosnik, Sekuler, & Kline, 1990; Sekuler & Ball, 1986).

It is also known that memory for spatial location declines with age (Newman & Kaszniak, 2000; Burton-Danner et al., 2001), and that encoding of location is obligatory, occurring even when such information is irrelevant to the goals of the current task (Hasher & Zacks, 1979). Because spatial memory of previously-seen item locations could direct and speed search in TMT, individual differences in registering and retaining spatial information could produce individual differences in the speed and accuracy of TMT performance. MOB_i was designed to eliminate this confound. Two successive targets were never presented in the same location. This greatly reduced the possibility that inhibition of return would affect the outcome latencies (Bennett & Pratt, 2001). (Inhibition of return refers to the finding that response times are typically slower for targets that appear at previously attended locations than for targets at novel locations.) Further, this control removed any effect of memory for spatial location with repeated presentations. However, this control also may have increased the working memory load in MOB. In TMT, while attempting to select the next item in a sequence, a subject could see the last item chosen. By eliminating this potential external reference, MOB forces subjects to retain the last item chosen in working memory, increasing the load on working memory.

We also wanted to determine whether age-related differences in TMT performance arose from age-related changes in visual perception (Salthouse & Fristoe, 1995; Salthouse et al., 2000). Individual differences in one particular visual ability, contrast sensitivity, broadly affect visual perception (Sekuler & Sekuler, 2000), including the speed with which alphabetic characters are encoded (Groth, Gilmore, & Thomas, 2003). Because of well-documented age-related changes in contrast sensitivity (Higgins, Jaffe, Caruso, & deMonasterio, 1988; Owsley, Sekuler, & Siemsen, 1983), it seemed reasonable to hypothesize that some age-related changes in TMT might arise from age-related decreases in contrast sensitivity.

MOB_i's components allow for direct investigation of factors that might contribute to age-related changes in performance on the test. These putative factors include baseline response speed, executive function, contrast sensitivity, and familiarity with the task. MOB_i makes it possible to evaluate these factors while reducing the impact of other higher-order

cognitive functions, such as visual search or spatial memory.

METHODS

Participants

Fifteen young (11 female, 4 male, mean age = 20.4, SD = 0.99, Range = 18-22 years) and 15 older participants (5 female, 10 male, mean age = 74.9 SD = 6.46, Range = 68-92 years) served for pay. All were right handed, and none reported cognitive, visual, or motor deficiencies other than the need for corrective lenses. Each participant's contrast sensitivity was measured with the Pelli-Robson charts (Pelli, Robson, & Wilkins, 1988). Additionally, each participant's Snellen distance acuity was measured with an Optec 2000 vision tester. The younger participants demonstrated better visual acuity (Range = 20/18 - 20/100, M = 20/17, SD = 20/5.05) than the older participants (Range = 20/17 - 20/50, M = 20/28, SD = 20/13.24). Contrast sensitivity was also better in the younger (Range = 1.80-1.95, M = 1.93, SD = 0.053) as compared to the older adults (Range = 1.50-1.95, M = 1.75, SD = 0.15). (Note that the mean and standard deviation for the younger subjects omits the acuity measure for one younger subject whose acuity, 20/100, appears to be an outlier.)

Procedure

Participants were tested on each of MOB_i's four component tests, which were presented in eight, randomized blocks. Testing was preceded by a computer-based tutorial and practice opportunities.

To evaluate contrast's impact on performance, each of MOB_i's components was run with all test items set to one of five different, logarithmically-spaced contrast levels, 0.14, 0.23, 0.35, 0.59, and 0.96. Note that while the contrast of test items varied, the contrast of the eight squares was fixed at 0.96. This meant that although low contrasts might make targets and distracter items more difficult to recognize, the surrounding, highly-visible squares made it easy to localize items. Over eight blocks of tests, all of MOB_i's components were run four times at the highest contrast, and once at each of the other four contrast levels.

For each participant, order of testing was randomized independently for various contrast levels and for various sub-tests. The extra replications at the highest contrasts enabled more detailed comparisons between the younger and older adults. Participants were told to respond to each target as quickly as possible, bearing in mind that while accuracy was important, speed was more important.

Test description

MOB_i was comprised of four sub-tests: Speed control, Numbers only, Letters only, and Numbers and Letters. Nineteen distinct display frames made up each sub-test. Each sub-test was presented a total of 8 times. Four of these repetitions were conducted at high contrast between the target and display. One additional presentation of each sub-test was also carried out at each of 4 additional, lower levels, of contrast. Thus the protocol consisted of 4 sub-tests, each presented at 5 different levels of contrast. These manipulations produced a total of 36 tests. To reduce fatigue, the tests were grouped into 8 blocks.

MOB_i was implemented in Matlab (The Mathworks, Natick Massachusetts) on an Apple iMac computer whose display had been fitted with a touch screen (Mass MultiMedia, Colorado Springs, Colorado). During testing, participants responded by touching a region on the touch screen with the index finger of their preferred hand.

All of MOB_i's sub-tests shared the same visual format, which facilitated comparisons among the tests. This format comprised eight squares, spaced evenly around the circumference of an imaginary circle. At the average viewing distance, 42 cm, the circle's diameter was 18.9 degrees visual angle. Eight squares, each 3.1 degrees, defined the locations within which test items were presented. The centers of neighboring squares were separated by

7.4 degrees. The display background was white (56 cd/m^2), and unless otherwise indicated, squares, target and distracter items were high contrast, black ($\sim 0.3 \text{ cd/m}^2$). The squares' consistent position and their high contrast were meant to aid visual search.

Nineteen successive display frames made up each of MOB_t 's four sub-tests. A display frame remained visible until the participant made the correct response on the touch screen. Then, with an imperceptible delay ($< 2 \text{ msec}$), a new display frame replaced its predecessor. Each sub-test took approximately 30 seconds to complete. To minimize potentially disruptive transients (Yantis & Jonides, 1984), the monitor's background luminance remained constant as one frame succeeded another.

In each of MOB 's component sub-tests, when a subject touched a square that did not contain the target item, or touched some spot on the screen that was not in any of the squares, the computer immediately emitted a distinctive beep and highlighted the correct square, that is the square that should have been touched. To highlight the correct item, the outline square around that correct item instantly tripled in thickness, a change that was highly visible. After the correct item had been highlighted, the computer's error correction routine prevented the presentation of a new frame until the subject touched the correct square.

The following paragraphs give brief descriptions of MOB_t 's four sub-tests. Figure 1 shows sample displays for the first four frames for each sub-test. When looking over the samples, it is important to keep in mind that every time a test was administered, the distribution of items over the eight squares was randomized anew. As a result, the displays actually presented were not identical to the samples in Figure 1.

<FIGURE 1>

Speed Control test (SC_t)

This sub-test assessed participants' baseline response speed. As illustrated in Figure 1A, a letter X was centered in one of the eight squares, chosen at random for each frame. The participant had to touch anywhere in the square, or within 0.5 cm around the square. Because all other squares were empty, the lone target could be processed pre-attentively, that is, the target popped-out perceptually, eliminating the need for search or scrutiny (Treisman, 1988).

Numbers test (N_t)

This sub-test, which is illustrated in Figure 1B, is analogous to Form A of TMT. On display frame n , the numbers n to $n+7$ were presented, one inside each of the eight squares. On display frame n , the number n was the target, with the remaining seven numbers serving as distracters. The participant's task on the n^{th} frame was to touch the square that contained the number n . In other words, on frame 1, the participant touched the square that contained 1, on frame 2, the participant touched the square that contained 2, etc. The distribution of numbers over the eight squares was randomized anew for each frame.

Letters test (L_t)

This sub-test was like N_t , but substituted the letters A through S for the numbers 1 through 19 (see Figure 1C). L_t has no analogue in TMT; it provides a measure of the speed with which letters could be ordered. As in N_t , the distribution of items over the eight squares was randomized anew for each frame.

Numbers and Letters test (NL_t)

This sub-test was analogous to TMT's Form B. As illustrated in Figure 1D, each display frame comprised four letters and four numbers. From one display frame to the next, the target alternated between numbers and letters, that is 1, A, 2, B, 3, C, \dots I, 10. Accurate performance is thought to require participants to switch between letters and numbers, and also, as in N_t and L_t , to remember which letter or number came next in the target sequence.

General Characteristics of MOB_t

In N_t, L_t, and NL_t, once an item had been touched, that item and all accompanying items disappeared, and were replaced by the items for the next display frame. As a result, successive responses in all three of these tests required that participants remember which item had been most recently selected.

Within any display frame of N_t, L_t, or NL_t, the non-target (distracter) items comprised the seven items that immediately followed the target in the appropriate sequence. For example, if 5 were the target on the fifth frame of an N_t sequence, the accompanying distracters would be 6 through 12. Similarly, on the third frame of a L_t series, C would be the target, and the accompanying distracters would be the letters D through J. Finally, in an NL_t series, when the target was C, it was accompanied by 4, D, 5, E, 6, F, and 7 as distracters. This was done to maintain comparability with TMT, where participants mark targets as they go, and therefore no longer need to take account of such targets. However, unlike in TMT where physically marking “used” targets could reduce task difficulty as the test progressed by narrowing the search possibilities, MOB_t’s design makes for consistent levels of difficulty throughout a series.

The purpose of the current investigation was to address five main questions that arise from TMT. Our design of MOB_t was meant to address these ongoing issues. First, although contrast sensitivity declines with advancing age we do not expect to find age-related switching costs as a result of contrast sensitivity. Instead, we predict that contrast would have to be greatly reduced before it would affect outcome performance. Second, based on the TMT literature with older adults we predict that older adults will show a greater cost associated with switching between letters and numbers. As part of this increased cost, older adults will show an increased tendency to perseverate, failing to switch between letter and number series. Third, rather than switching between letters and numbers during the NL_t we would predict that, at least early in the alphabet (with over-learned item pairs) subjects will recall a number and its paired letter (e.g., 1-A, 2-B) as a single unit. This will result in a chunking of items such that the latency to respond to the first item in the pair is greater than the latency to respond to the second item in the pair. Fourth, the difficulty of switching between letters and numbers as measured by latency in MOB should vary over the number-letter series (differential familiarity and practice with order information early as compared to late in the alphabet). Finally, we expect that practice will differentially affect various test subcomponents. In particular, as task difficulty increases (SC_t-N_t-L_t-NL_t), we expect that the effect of practice would increase. In the following section, the organization of results follows the order of hypotheses presented here.

RESULTS & DISCUSSION

Relatively few responses were errors. For young participants, 2.1% of total responses were errors and for the older participants, 4.7% of their responses were errors. A response that immediately followed an error was almost always correct, but tended to have an abnormally long latency. As a result, we eliminated from our analysis all response times associated with errors, as well as response times immediately following an error. Our main measure of each participant’s performance was the mean response time per display frame, excluding (i) responses on which the participant made an error, and (ii) responses to the next target in the series. These latencies are presented both as times averaged across each sub-test (average RT across all display frames) and as the averaged times for each serial position (averaged across repetition of each sub-test).

Overall effects

Figure 2 shows results from tests with the highest contrast targets and distracters. The overall difference in times for older and young participants, which averaged more than half a second per response, was significant $F(1, 28) = 90.37, p < .01$; on every sub-test, younger participants' responses were faster than the responses of their older counterparts. Second, there was an overall interaction between age and type of sub-test, with older adults being especially slowed on NL_t , $F(3, 84) = 23.85, p < .01$. In addition, for both age groups, performance on SC_t was significantly faster than for any other test ($p < .01$); performance on NL_t was significantly slower than for any other test ($p < .01$). Performance on N_t and on L_t did not differ significantly from each other ($p < .05$).

<FIGURE 2>

Contrast sensitivity

The effect of contrast variation does not notably differ among MOB_t 's sub-tests, for either young or older adults, $F(12, 336) = 1.318, p > 0.20$. As Figure 3 shows, contrast variation has little effect until extremely low contrast values are reached. Young and older participants' mean response times as a function of the test items' contrast, averaged over MOB_t 's four sub-tests are presented in Figure 3. The results show that the overall mean response times increase with decreasing contrast, $F(4, 112) = 33.98, p < .01$. Moreover, contrast's effect on older participants' performance was considerably more powerful than its effect on young participants' performance, $F(4, 112) = 12.14, p < .01$.

<FIGURE 3>

As a first step to evaluating age-related consequences of contrast variation, we used a least squares criterion to find the line that best fit the young participants' results ($r^2 = 0.93$). This line is drawn through the young participants' data in Figure 3 for NL_t only. That same best-fitting line has been drawn also through the older participants' data, but has been shifted vertically by 0.63 seconds, which is the difference between the two groups' mean times at the highest contrast. If the difference between groups were entirely explained by their relative response times at the highest contrast, all of the older participants' mean response times would be well represented by this vertically shifted best-fitting line. Although the shifted best-fitting line adequately captures older participants' data at the three highest contrasts, as contrast decreases further, the data increasingly deviate from the line. This divergence signifies that older participants' response times are slowed more by very low contrast than are the response times of their young counterparts.

The divergence just described suggests also that age-related differences on MOB_t cannot be explained by the modest difference between the two groups' contrast sensitivity (see Participants). As expected, young and older participants differed in contrast sensitivity (Higgins et al., 1988; Owsley et al., 1983). Young participants' mean contrast sensitivity was 1.93; older participants' mean contrast sensitivity was only 1.76. This difference of 0.17 \log_{10} units means, on average, the older participants needed 48% more contrast to detect visual targets (e.g., $\log_{10} 1.48 = 0.17$). As a result, age-related differences in contrast sensitivity could be compensated for if older participants were given test items whose contrast was 48% higher than items presented to young participants. Figure 3 shows that this compensation would fall far short of equating older and young participants' performance on MOB_t . In fact, a single step along the horizontal axis, from a contrast 0.96 to 0.59, represents a change in contrast (63%) larger than the age-related difference in contrast sensitivity. This supports the proposition that reduced contrast threshold cannot account for older participants' slowed performance on MOB_t .

Analysis of errors

All errors can be put into one of two categories. For all groups of participants and all

sub-tests (except the NL_t), the most common errors were motor errors, in which the participant either touched the screen outside of any target location, or touched the same box twice in rapid succession (typically between 10 and 90 msec), suggesting that the finger made, broke, and then quickly re-made contact with the screen. These errors have little to do with cognitive function. Non-motor errors occurred when participants lost their way, either within a series of like items, for example, touching the number 7 when they should have touched the number 5, or between series of different types of items, for example, touching a letter when they should have touched a number.

In Figure 4, motor and non-motor errors are displayed separately for each of the four sub-tests. Note that the only types of errors possible in the SC_t were motor errors. Also, the non-motor errors (other errors) in Figure 4 for both the N_t and L_t only represent within-series errors, whereas the non-motor errors for NL_t are a composite of both within and between series errors. On average, older participants made more errors than did young participants, and non-motor errors were relatively more frequent in NL_t than in other sub-tests.

<FIGURE 4>

As indicated above, two distinct types of mistakes comprised non-motor errors. The first type involves *within-category errors* (even numbered display frames), which occur when participants choose an incorrect target that is in the same category as the correct target (for example, choosing the wrong number when the target was a number, or choosing the wrong letter when the target was a letter). The second type of non-motor error, *Between-category errors* (odd numbered display frames), occurred with the choice of a target that is in a different category from the correct target (for example, choosing a letter when the target was a number, or vice versa).

<FIGURE 5>

Figure 5 shows the relative distributions of non-motor errors over the seven non-targets (distracters) that shared the display frame with the target item for NL_t . The distributions of these errors have been normalized for both younger and older adults. Note that one particular ordinal position attracted more wrong choices than other positions did. The consistent predominance of errors at $n+2$ is evidence of perseveration. Such errors arose when participants neglected to switch tasks when choosing a target, mistakenly selecting the next letter (when the previous target had been a letter), or the next number (when the previous target had been a number). Interestingly, younger participants made only about half as many errors overall as older participants, including errors of perseveration, failures to switch tasks. Despite this substantial age-related difference in overall perseveration rates, Figure 5 shows that both groups showed the same relative propensity to perseverate.

Inter-response times (IRTs) and response bursting

When participants attempt to recall a series of items, inter-response times (IRTs) provide valuable clues to the way in which recalled items were stored in memory. For example, the pattern of IRTs may reveal clustering or chunking of items in memory (Kahana & Loftus, 1999). Successive items from a shared category tend to be recalled with shorter IRTs than items from different categories (Wingfield et al., 1998). This phenomenon, sometimes called response bursting, is not limited to free recall with verbal material; similar bursting, for example, is seen in the timing of successive moves through a previously learned, complex virtual environment (Kirschen, Kahana, Sekuler, & Burack, 2000) and in other settings as well.

To examine the pattern of IRTs in MOB_t , data from NL_t were sorted according to the target's serial position within the 19-item sequence of numbers and letters, where "1" occupied the first position, "A" the second, "2" the third, "B" the fourth, and so on. Figure 6 shows participants' mean response time for each of the 19 target items in NL_t , from 1, A, 2, .

·10. To facilitate comparisons between young and older participants, response times were normalized by setting each group's mean time to unity. Thus, in Figure 6, the value 1.2 signifies a response time 20% above the series' mean time. This analysis used data only with targets of the highest contrast.

<FIGURE 6>

The curves in Figure 6 exhibit several interesting features. First, response times tended to be relatively short at the sequence's start, and lengthen progressively throughout the series. On average, relative response times change by about 50% over the course of the series. This slowing of response cannot be attributed to simple fatigue because no such slowing was seen with sub-tests other than NL_t . Instead, we believe that it reflects the additional cognitive load imposed by switching operations unique to this sub-test.

Prior to testing, participants were already well familiar with letter series and number series, but when they had to shift back and forth between the two series a relatively unfamiliar skill was required: participants had to track simultaneously targets drawn from two separate series. This additional challenge may be responsible for a second noteworthy feature in Figure 6, the distinct up-and-down, scalloping pattern in the middle portion of each curve. Note that responses for most of the letter target items are faster than they are for the immediately preceding or immediately following number targets. These systematic swings in response time comprise about 10-20% of the total scaled times. We believe that this scalloping shows that some pairs in NL_t were encoded and retrieved as single combined items, for example, "3-C," "4-D," and so on as opposed to switching between tasks, although the task was presented in a serial, rather than parallel fashion. Note that this strategy, implied by the pattern of inter-response times, is quite different from the usual depiction of the task on Form B of TMT which is one of holding two streams of information in parallel and switching between the two streams to complete the task. This result may require some re-thinking of how TMT-B's task and results are conceptualized.

Practice

The effects of multiple administrations were examined for both age groups. Each subtest was presented 4 times across the experimental session. The presentation order for the subtests was randomized so that the actual time between repetitions of any one subtest varied. Practice had no overall effect for either the young or the older adults for N_t and L_t . And the overall effects of age were not significantly affected by multiple administrations, $F(3,84)=2.429, p=.07$. However, practice produced a significant, but small reduction in response times for the motor baseline ($SC_t; F(3,84)=3.159, p=.029$), and exerted a more substantial effect on NL_t ($F(3,84)=11.517, p<.01$). In fact, as Figure 7 suggests, by the fourth repetition of that condition, performance on NL_t has become very nearly as fast as performance on either N_t or L_t for the younger adults and is greatly reduced for the older adults. Although additional data would be required to establish this point with certainty, we hypothesize that practice with NL_t allows participants to chunk additional pairs of responses (see Figure 6). This reduces the need to switch, making NL_t 's series of targets more similar to the well-learned series in N_t and L_t .

<FIGURE 7>

Additive factors analysis

We took a simple additive factors approach to characterize executive function's contribution to performance with NL_t , (Sternberg, 1999). This approach seeks to describe some overall effect as a sum of the effects associated with several separable factors. Although individual factor effects can be linear, quadratic, or higher order, cross product terms (interactions) involving two or more factors are assumed to be zero. We took this simple approach because it promised to account for the results with few free parameters. In addition,

this approach makes it easy to detect errors in predictions, which would be diagnostic of non-zero interactions that had been overlooked. The analysis rests on two assumptions.

1. The visual and motor operations demanded by SC_t contribute equally to each of MOB_t 's three other tests. The time required by these operations can therefore be subtracted from the performance times of each of the other tests, leaving higher order cognitive processes as a residual.

2. NL_t includes the ordering or sequencing operations required by N_t and L_t but also recruits executive processes, which shift from one sequence to the other, and back.

These two assumptions lead to a simple quantitative account of the components of response times in NL_t . The average response time in NL_t , $\overline{NL_t}$, can be described as

$$\overline{NL_t} = SC_t + 0.5 \cdot [(L_t - SC_t) + (N_t - SC_t)] + Exec \quad [\text{Eq.1}]$$

Equation 1 assumes that each response time in NL_t comprises, in addition to the unique cognitive demands posed by NL_t , whatever time is required by SC_t , 1/2 of the mean response time in L_t (because NL_t presents only the letters A-I, but L_t uses twice as many), and 1/2 of the mean response time in N_t (because NL_t presents only the numbers 1-10, but N_t uses twice as many, 1-19). The time required by SC_t was subtracted from each of the two other tests because that time had already been represented, by the first term in Equation 1. The remaining term in Equation 1 represents NL_t 's unique executive component.

Because of the near equality of times associated with L_t and N_t , and because Nt includes all times associated with SC_t , the terms in Equation 1 can be rearranged and simplified to

$$Exec = NL_t - N_t \quad [\text{Eq. 2}]$$

To facilitate comparisons among groups, we normalized Equation 1's components within each age group, expressing each as a percent of the total time in NL_t . Table 1 shows that the largest fraction of response times in NL_t , from 39 to 44%, is consumed by simple visual detection and execution of motor responses. Number and letter ordering and recall (the N_t and L_t components) take up nearly equal fractions of response times for both groups. Most interesting is the suggestion that executive operations account for a small, consistent fraction of the response time in young participants (mean = 17%, standard error = 1.9%), but a considerably larger fraction, (mean = 29%, standard error = 2.4%) with older participants. (The results in Figure 7 suggest that executive function's contribution to NL_t may diminish with practice for both young and older adults. For the sake of simplicity, our model does not include such practice effects.)

<TABLE 1>

Considering the relative age-related increases in reaction times, NL_t 's unique component increases far more than the visuomotor component or the components required to order numbers or letters. The additive factors analysis confirms the idea that age brings particular changes in cognitive processes unique to NL_t .

GENERAL DISCUSSION

In addition to remediating various deficiencies in TMT, MOB_t was meant to separate component processes that likely contribute to performance on TMT. Although older participants were slowed by reduced item contrast, contrast sensitivity differences cannot account for the significant age-related differences in performance on MOB_t , or presumably, in TMT.

MOB_t successfully isolated a distinct executive function component of reaction time, and implicated this component in the age-related differences in performance on MOB_t (and

presumably, on TMT). This result appears to be inconsistent with the idea that age-related changes in TMT and variants of that test arise from perceptual slowing, not from an alteration in efficiency of executive function. The results also showed that this executive component was reduced with practice. We hypothesize that this reduction resulted from the replacement of switching operations with learned encoding and retrieval of chunked letter-number pairs, as reflected in Figure 6.

Although our study used variants of TMT rather than TMT itself, our results have important implications for that widely-used test. MOB_t's incorporation of an automated error detection and correction scheme offers one example of such implications. Although errors on MOB_t were not common (<5%), we believe it is valuable to have a consistent method that immediately detects any error, guides the participant to the correct response, and then makes appropriate adjustments to the resulting data. These important functions are missing from TMT, as well as a recent update of that test (Delis, Kaplan, & Kramer, 2001). In MOB_t participants have unusually long latencies not only when they make an error, but on the next correct response as well. If times associated with errors and times associated with immediately succeeding, correct response were included in participants' scores, those scores would spuriously exaggerate the difference between performance on TMT-A and TMT-B. Moreover, failure to censor data appropriately would generally produce an overestimate of the real magnitude of age-related differences in executive function. Because we have not tested other groups of participants that typically show abnormally large differences between TMT-A and TMT-B, we cannot say whether the same caveat would apply to them. However, this is a concern that should be borne in mind whenever TMT or related tests are administered.

Although failure to identify and censor error-related responses may exaggerate the real size of differences attributed to executive function, another factor may work in the opposite direction, causing such differences to be underestimated. Because MOB_t captures and analyzes individual responses, we could examine changes in performance on NL_t across its entire series of test items. As Figure 6 shows, response times with NL_t increase systematically from the test series' beginning (e.g., 1-A, 2-B) to its end (e.g., 9-I, 10-J). Because performance with N_t or L_t showed no such variation, differences between NL_t, on one hand, and N_t and L_t, on the other hand, would be smallest early in the NL_t series, and increase thereafter. Because well-learned items, early in the series, extract only a modest cost in switching between numbers and letters, TMT-B's single aggregate score will include component responses that are relatively quick (near the series' start) and responses that are relatively slow (from the series' end).

Throughout, we have adhered to a narrow definition of executive function, restricting the term to signify planning of actions and switching between tasks. Current models, however, tend to fractionate executive control into additional independent subprocesses (e. g., Shallice, 2002), including subprocesses responsible for the ability detect and correct one's errors. MOB_t identifies and separates erroneous responses, but at the same time, MOB_t's automated error handling routine eases the burden on participants to spot and then correct their own errors. As a result, the test does not allow full expression of individual differences in executive processes that are specifically associated with error detection and correction. We believe that this potential drawback is more than compensated by the methodological and analytical benefits from MOB_t's standardized error-handling procedure.

As noted earlier, one reason for TMT's popularity is its ability to reflect changes in the efficiency of executive function. There is growing consensus that several aspects of executive function depend crucially upon prefrontal cortex (Shallice, 2002). Consistent with this idea, Stuss and colleagues (2001) found that focal damage to the frontal lobe, particularly

damage to dorsolateral frontal areas, degrades performance on TMT Form B. (However, see Anderson, Bigler & Blatter, 1995). Recent magnetic resonance studies have shown that performance on TMT's form B, relative to form A, is strongly correlated with incidence of white matter hyperintensities (Takahashi, Murata, Omori, Kosaka, Takahashi, Yonekura, & Wada, 2004), and more generally with decreases in cerebral volume (Coffey, Ratcliff, Saxton, Bryan, Fried, Lucke, 2001), which commonly accompany aging. We believe that future research on structural correlates could benefit by using MOB as an index of executive function. The present study suggests that MOB's item by item analysis would not only enable a more fine-grained picture of overall executive function, but also that MOB's subtests would enable the isolation and analysis of several subcomponents of that overall function.

Earlier, we commented on some attempts to update and improve TMT (Salthouse & Fristoe, 1995; Salthouse et al., 2000). Here, it is worth noting one other effort, which has produced a modified paper-and-pencil version of TMT. A new test battery by Delis, Kaplan and Kramer (2001), which is meant to assess executive function, includes an expanded version of TMT. Their version provides separate tests of baseline motor speed and visual scanning. However, the paper-and-pencil format forecloses any chance of recording and analyzing individual responses, whose importance is amply demonstrated by our results.

Finally, we look on MOB_t as a promising, advanced prototype, rather than a finished product. However promising the present results may be, MOB_t must be validated with larger samples of other participants including participants of all ages, socioeconomic status, and educational background must be examined. Additionally, the efficacy of using such a measure in a clinical setting, for example, ease of scoring and interpretation, would have to be examined directly. Whatever final form the test may take, we believe that the present results clearly demonstrate the importance of collecting and then analyzing individual response times and errors.

ACKNOWLEDGEMENTS

We thank Kelly Addis and Arye Elfenboim for contributions at the start of this project, and Jon Peelle and Alice Cronin-Golomb for helpful comments on an early version of this report. Partially supported by NIH grant AG-04517. Address correspondence to R. Sekuler, Volen Center, Brandeis University, MS 013, Waltham MA 02454 (e-mail: vision@brandeis.edu).

REFERENCES

- Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: the costs of choosing the optimal path. *Vision Research, 41* (25-26), 3613-3625.
- Arbuthnott, K. & Frank, J. (2000) Trail making test, Part B as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology, 22*, 518-528.
- Anderson, C.V., Bigler, E.D., & Blatter, D.D. (1995). Frontal lobe lesions, diffuse damage, and neuropsychological functioning in traumatic brain-injured patients. *Journal of Clinical and Experimental Neuropsychology, 17*(6), 900-908.
- Bennett, P. J., & Pratt, J. (2001). The spatial distribution of inhibition of return. *Psychological Science, 12* (1), 76-80.
- Burton-Danner, K., Owsley, C., & Jackson, G. (2001). Aging and feature search: the effect of search area. *Experimental Aging Research, 27* (1), 1-18.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36* (1), 28-71.
- Coffey, C.E., Ratcliff, G., Saxton, J.A., Bryan R.N., Fried, L.P., & Lucke, J.F. (2001). Cognitive correlates of human brain aging: a quantitative magnetic resonance imaging investigation. *Journal of Neuropsychiatry and Clinical Neuroscience, 13*(4), 471-485.
- Cowell, P. E., Turetsky, B. I., Gur, R. C., Grossman, R. I., Shtasel, D. L., & Gur, R. E. (1994). Sex differences in aging of the human frontal and temporal lobes. *Journal of Neuroscience, 14* (8), 4748-4755.
- Crowe, S. F. (1998). The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts a and b of the trail making test. *Journal of Clinical Psychology, 54* (5), 585-591.
- Davies, D. M., Spelman, M. S., & Davies, M. G. (1981). Combining psychometric data on brain damage and the influence of aging. *Perceptual & Motor Skills, 52* (2), 583-592.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). D-KEFS: Delis Kaplan Executive Function System. San Antonio: The Psychological Corporation.
- Drane, D. L., Yuspeh, R. L., Huthwaite, J. S., & Klingler, L. K. (2002). Demographic characteristics and normative observations for derived trail making test indices. *Neuropsychiatry, Neuropsychology & Behavioral Neurology, 15* (1), 39-43.
- Duncan, J. (1986). Disorganization of behavior after frontal-lobe damage. *Cognitive Neuropsychology, 3*, 271-290.
- Ehrenstein, W. H., Heister, G., & Cohen, R. (1982). Trail making test and visual search. *Archiv fur Psychiatrie und Nervenkrankheit, 231* (4), 333-8.
- Filley, C.M. (2000). Clinical neurology and executive dysfunction. *Seminars in Speech and Language, 21*(2), 95-108.
- 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* (4), 529-35.
- Groth, K., Gilmore, G. C., & Thomas, C. (2003). Impact of stimulus integrity on age differences in letter matching. *Experimental Aging Research, 29* (2), 155-172.
- Gunning-Dixon, F. M., & Raz, N. (2000). The cognitive correlates of white matter abnormalities in normal aging: a quantitative review. *Neuropsychology, 14* (2), 224-232.
- Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 17* (1), 163-169.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356-358.

- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, p. 193-225). New York: Academic Press.
- Hays, J. R. (1995). Trail making test norms for psychiatric patients. *Perceptual & Motor Skills, 80* (1), 187-94.
- Higgins, K. E., Jaffe, M. J., Caruso, R. C., & deMonasterio, F. M. (1988). Spatial contrast sensitivity: effects of age, test-retest, and psychophysical method. *Journal of the Optical Society of America, A, 5* (12), 2173-80.
- Kahana, M., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The nature of cognition* (p. 322-384). Cambridge: MIT Press.
- Kirschen, M. P., Kahana, M. J., Sekuler, R., & Burack, B. (2000). Optic flow helps humans learn to navigate through synthetic environments. *Perception, 29* (7), 801-818.
- Kline, D. W., Kline, T. J., Fozard, J. L., Kosnik, W., Schieber, F., & Sekuler, R. (1992). Vision, aging, and driving: the problems of older drivers. *Journal of Gerontology, 47* (1), P27-P34.
- Kosnik, W., Winslow, L., Kline, D., Rasinski, K., & Sekuler, R. (1988). Visual changes in daily life throughout adulthood. *Journal of Gerontology, 43* (3), P63-P70.
- Kosnik, W. D., Sekuler, R., & Kline, D. W. (1990). Self-reported visual problems of older drivers. *Human Factors, 32* (5), 597-608.
- Kramer, A. F., & Atchley, P. (2000). Age-related effects in the marking of old objects in visual search. *Psychology & Aging, 15* (June), 286-296.
- Meyer, D. E., & Kieras, D. E. (1997). Epic: A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3-65.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140.
- Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31* (3), 327-342.
- Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory & Cognition, 16* (1), 127-137.
- Newman, M. C., & Kaszniak, A. W. (2000). Spatial memory and aging: Performance on a human analog of the Morris water maze. *Aging, Neuropsychology, & Cognition, 7* (2), 86-93.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (p. 1-18). Plenum Press.
- Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research, 23* (7), 689-99.
- Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences, 2*, 187-199.
- Plude, D. J., & Hoyer, W. J. (1981). Adult age differences in visual search as a function of stimulus mapping and processing load. *Journal of Gerontology, 36* (5), 598-604.
- Plude, D. J., & Hoyer, W. J. (1986). Age and the selectivity of visual information processing. *Psychology & Aging, 1* (1), 4-10.
- Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual & Motor Skills, 8*, 271-276.
- Richardson, D. C., & Spivey, J. M. (2000). Representation, space and Hollywood squares: Looking at things that aren't there anymore. *Cognition, 76* (3), 269-295.
- Salthouse, T. A. (1993). Attentional blocks are not responsible for age-related slowing. *Journal of Gerontology, 48* (6), P263-P270.

- Salthouse, T. A., & Fristoe, N. M. (1995). Process analysis of adult age effects on a computer administered trail making test. *Neuropsychology*, 9 (4), 518-528.
- Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. J. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology*, 14 (1), 102-111.
- Sekuler, R., & Ball, K. (1986). Visual localization: Age and practice. *Journal of the Optical Society of America, A*, 3 (6), 864-867.
- Sekuler, R., & Sekuler, A. B. (2000). Visual perception and cognition. In J. G. Evans, T. F. Williams, J.-P. Beattie, B Land Michel, & G. K. Wilcock (Eds.), *Oxford textbook of geriatric medicine* (2 ed., p. 874-880). New York: Oxford.
- Shallice, T. (1994) Multiple levels of control processes. In: C. Umilta & M. Moscovitch (Eds.) *Attention and performance 15: Conscious and nonconscious information processing* (p. 395-420), Cambridge, MA: The MIT Press.
- Shallice, T. (2002) Fractionation of the supervisory system. In D. T. Stuss, & Knight, R. T. (Eds.), *Principles of frontal lobe function* (p.261-277), London : Oxford University Press.
- Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: administration, norms, and commentary*. New York: Oxford University Press.
- Sternberg, S. (1999). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science* (p. 703-786). New York: Oxford University Press.
- Stuss, D.T., Bisschop, S. M., Alexander, M.P., Levine, B., Katz, D., & Izukawa, D. (2001). The Trail Making Test: a study in focal lesion patients. *Psychological Assessment*, 13(2), 230-239.
- Takahashi, T., Murata, T., Omori, M., Kosaka, H., Takahashi, K., Yonekura, Y., & Wada, Y. (2004). Quantitative evaluation of age-related white matter microstructural changes on MRI by multifractal analysis. *Journal of Neurological Science*, 225(1-2),33-37.
- Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 40A(2), 201-237.
- Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology & Aging*, 13 (2), 256-266.
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 10 (5), 601-621.

TABLE & FIGURE CAPTIONS

Table 1. Additive factors estimated for both younger and older adults.

Figure 1. Samples of four display frames for each of MOB_t's four sub-tests.

Figure 2. Mean response times for various sub-tests in MOB_t for both younger and older adults are presented for tests run at the highest contrast.

Figure 3. Mean response times as a function of stimulus contrast for young (lower symbols and line) and older (upper symbols and line) participants. Results have been averaged over all of MOB_t's sub-tests. The least-squares fit for young participants is drawn through that group's data points; a second line for the older adults, with the same slope, but shifted upwards by 0.63 seconds is also shown.

Figure 4. Percent of trials on which an error was made. Results are shown separately for each of MOB_t's sub-tests.

Figure 5. Proportion of erroneous choices of various items in an NL_t display frame. The horizontal line at a proportion of 0.14 indicates the values expected if errors were uniformly distributed across a display frame's seven items.

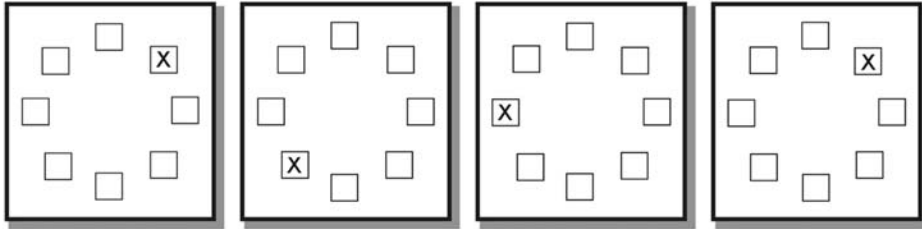
Figure 6. Normalized response time as a function of target position in NL_t series. Separate curves are shown for young participants and for older participants.

Figure 7. Mean response times as a function of practice for various sub-tests in MOB_t. Results are shown for each of the 4 repetitions at the highest contrast. Panel A: results from young participants; Panel B: results from older participants.

Factor as a Percent of Total Time in NLt	Subtest Unique			
	SC _t	N _t	L _t	NL _t
Young Adults	43	21	19	17
Older Adults	39	16	16	29

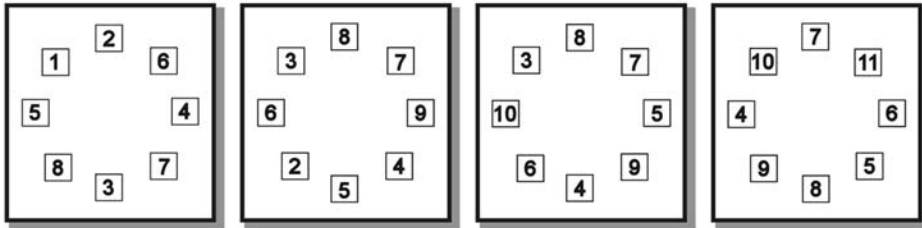
Speed Control test (SC_t)

A



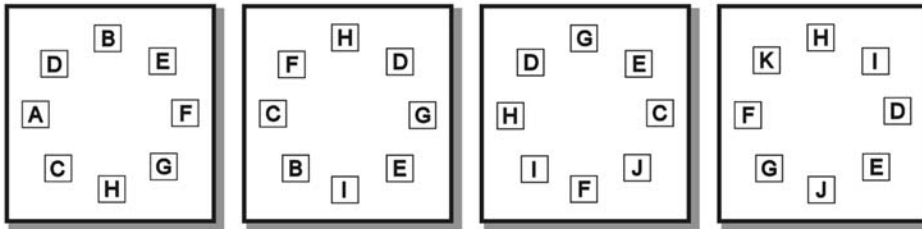
Numbers test (N_t)

B



Letters test (L_t)

C



Numbers Letters test (NL_t)

D

