

Computerized Neurocognitive Scanning: I. Methodology and Validation in Healthy People

Ruben C. Gur, Ph.D., J. Daniel Ragland, Ph.D., Paul J. Moberg, Ph.D., Travis H. Turner, B.A., Warren B. Bilker, Ph.D., Christian Kohler, M.D., Steven J. Siegel, M.D., Ph.D., and Raquel E. Gur, M.D., Ph.D.

Neuropsychological testing batteries are applied in neurobehavioral evaluations of brain disorders, including neuropsychiatric populations. They are lengthy, require expert administrators and professional scorers, and are prone to data handling errors. We describe a brief computerized neurocognitive "scan" that assesses similar domains with adequate reliability. The scan and a traditional battery were administered to a sample of 92 healthy individuals (44 men, 48 women) in a counterbalanced order. Both approaches showed a significant "sex-typical" gradient, with women outperforming men in verbal memory relative to spatial

tasks. Both methods also yielded similar profiles of sex differences, with the additional computerized measure of face memory showing better performance in women. Age effects were evident for both methods, but the computerized scan isolated the effects to speed rather than accuracy. Therefore, the computerized scan has favorable reliability and construct validity and can be applied efficiently to study healthy variability related to age and gender.

[Neuropsychopharmacology 25:766–776, 2001]

© 2001 American College of Neuropsychopharmacology.
Published by Elsevier Science Inc.

KEY WORDS: *Neuropsychological testing; Neurocognition; Computerized assessment; Sex differences; Aging; Psychometrics*

The relevance of neuropsychological testing to psychiatric diagnosis and treatment has been the focus of recent investigation (Sharma and Harvey 1999). Traditional neuropsychological test batteries have been developed and validated in neurological populations with focal and diffuse lesions (Benton 1994; Kaplan 1990). Assess-

ment of cognitive deficits has become integrated in neurological research and practice. For example, a consensus battery established for Alzheimer's research (Consortium to Establish a Registry for Alzheimer's Disease; CERAD) has benefited the field by facilitating collaborative studies across centers (Morris et al. 1989). Similarly, the value of neuropsychological testing in assessing effects of toxin exposure has been recognized, and standards have been established (Baker et al. 1985). Neuropsychological batteries usually include measures of executive functions—abstraction and mental flexibility, attention—as well as verbal and spatial memory, language, spatial processing, and sensorimotor function (Benton et al. 1994; Golden et al. 1991; Halstead 1947; Jarvis and Barth 1994; Reitan and Davison 1974; Saykin et al. 1991, 1995). These batteries have provided a method to link domains of cognitive performance with regional brain functioning, and algorithms have been developed to test models of such linkage formally (e.g., Gur et al. 1990). Application of neuropsychologi-

From the Brain-Behavior Laboratory, Neuropsychiatry Section, Department of Psychiatry (RGG, JDR, PJM, THT, CK, SJS, REG, WBB) and the Department of Biostatistics and Epidemiology (WBB), University of Pennsylvania, Philadelphia, Pennsylvania, USA.

Address correspondence to: Ruben C. Gur, Ph.D., Brain Behavior Laboratory, Department of Psychiatry, University of Pennsylvania, 10 Gates Building, 3400 Spruce St., Philadelphia, PA, 19104. Tel.: (215)662-2915; Fax: (215)662-7903. E-mail: gur@bbi.med.upenn.edu

Received January 11, 2000; revised April 17, 2001; accepted April 19, 2001.

Online publication: 4/26/01 at www.acnp.org/citations/Npp042601112.

cal batteries in psychiatric research and practice has been initially slow, but more recently, it has become consolidated into efforts to understand the involvement of neural systems in the pathophysiology of such major disorders as schizophrenia (Censits et al. 1997; Gold et al. 1992; Goldberg et al. 1987; Heinrichs and Zakzanis 1998; Mirsky and Duncan 1986; Saykin et al. 1991, 1994).

Although the traditional neuropsychological batteries have been useful in establishing links between behavioral deficits and brain dysfunction, they have several limitations when considered in the context of clinical neuroscience and treatment research. Weaknesses related to use in large-scale multicenter studies are length, complexity of administration and scoring, and vulnerability of data-handling methods. Traditional batteries take several hours to administer and require a trained technician supervised by a neuropsychologist. Scoring often involves expert judgment, and procedures for guarding against drifts are essential. Data entry is manual with inherent necessity for costly procedures to eliminate errors. An additional limitation of traditional batteries is that they are difficult to administer during functional imaging studies aimed at probing brain systems that regulate behavior (Gur et al. 1992). Subtests in these batteries are typically constructed to detect broadly defined deficits, rather than measure specific, unitary neurocognitive constructs. This feature hinders bridging, on the one hand, to basic human research in cognition and, on the other hand, to animal work. Finally, traditional tests necessarily confound speed and accuracy, two complementary features of performance that have been shown to relate differently to regional brain activation (Gur et al. 1988).

To address these limitations, we have developed a set of computerized neurobehavioral measures aimed specifically at integrating structural and functional neuroimaging studies. Our general approach to task development and validation was previously detailed (Gur et al. 1992). Here we present data on the implementation of the computerized tests in a normative sample that has also received the traditional battery. Because the computerized tests were designed to tap very specific neurocognitive systems and not to mirror the traditional battery, our goal was not to replace the traditional battery for all neurocognitive domains. However, the computerized tests retain reasonable psychometric properties, and our more limited goals were to present the psychometric properties of the computerized measures and to compare the sensitivity of the computerized and traditional measures to sex differences and age effects, which are major moderating variables in studies of brain and behavior (Kimura and Harshman 1984; Kramer et al. 1997; Saykin et al. 1995; Zec 1995). For a neurocognitive scan to be considered sensitive to normal individual differences in these measures, it will need to show that: (1) women perform better on mem-

ory tasks (McGivern et al. 1997; Ruff et al. 1989), relative to men who perform better on spatial processing and motor tasks (Caplan et al. 1997; Collins and Kimura 1997; Gur et al. 1999; Saykin et al. 1995; Silverman et al. 1996); (2) age is associated with decline in performance, which is more pronounced for frontotemporal functions (abstraction, attention, memory), and more pronounced for speed than for accuracy (Earles and Kersten 1999; Laursen 1997; Park et al. 1996; Sliwinski and Buschke 1999; Verhaeghen and Salthouse 1997).

In addition to comparing the computerized and traditional measures on sensitivity to sex differences and age effects, we sought to examine whether the entire array of computerized tests could be considered a "scan" of neurocognitive abilities that would yield comparable measures of major neurocognitive domains. The main difficulty in correlating the traditional with the computerized measures is that the former confound speed and accuracy, while the latter approach separates them. To make the scores comparable, we defined "efficiency" scores for the computerized battery, in which accuracy is divided by speed (performance accuracy per unit of time). Another difficulty in comparing the scores is the way in which the tasks were designed. Traditional tests are aimed at broadly defined domains of major clinical relevance, while the computerized tasks are narrowly defined to activate specific brain circuits. Although specific tests from the computerized scan have compared favorably with similar tests from the traditional battery (Kurtz et al. 2001; Ragland et al. 1995; Glahn et al. 1997; 2000), the traditional measures sample broader aspects of behavior and we had to "blur" the grouping of the computerized tasks to include measures of Executive, Memory, Intellectual, and Sensorimotor functions.

METHODS

Subjects

Participants were 92 healthy individuals (44 men, 48 women), recruited by newspaper advertisement. They underwent medical, neurological, and psychiatric (SCID-NP; Spitzer et al. 1996) evaluations (Shtasel et al. 1991) including laboratory tests. Subjects had no history of a disorder or event that might affect brain function including hypertension (blood pressure >140/90), cardiac disease, diabetes, endocrine disorders, renal disease, chronic obstructive pulmonary disease, cerebrovascular disease, head trauma with loss of consciousness, seizure disorder, migraines, or other neurological condition. They had no family history of schizophrenia or affective illness in first degree relatives. Men were 30.9 ± 10.7 years old (range 19.2–69.1) with 15.1 ± 1.8 years of education (range 12–18). The corresponding values for women were age: 27.2 ± 8.1 years (range

18.2–62.5); education: 15.8 ± 2.3 (range 12–22). Men and women did not differ on any of these measures.

Neuropsychological Assessment

Traditional Neuropsychological Battery. A comprehensive paper and pencil neuropsychological test battery (Saykin et al. 1994, 1995; Censits et al. 1997) was administered by trained examiners. A second examiner independently rescored test data to eliminate errors and permit assessment of interrater reliability. These scores are entered into a computerized database and checked for range, internal consistency, and secular drifts. The

test battery is listed in Table 1 including references for administration and scoring procedures used.

Test scores were standardized relative to all healthy people in our database (z-scores; mean=0, standard deviation = 1), and grouped into eight summary measures by averaging each subject's z-scores on tests assessing the same functional domain. Following Censits et al. (1997), summary measures were calculated for the following domains: abstraction (ABF), attention (ATT), verbal memory (VMEM), spatial memory (SMEM), language abilities (LAN), spatial abilities (SPA), sensory (SEN), and motor functions (MOT). The variables comprising these domains are presented in Table 1.

Table 1. Neuropsychological Performance of Healthy People on the Traditional Test Battery

Cognitive Domain and Component Test Variables	Time ^k (min)	Male		Female	
		Mean	SD	Mean	SD
Executive: Abstraction/flexibility (ABF)	15–20				
Percentage categories, WCST ^a		5.6	2.5	6.5	1.8
Percentage perseverative errors, WCST ^a		12.0	7.2	10.3	4.2
Attention (ATT)	30				
Seashore rhythm, total correct, HRB ^b		27.9	2.1	27.4	2.6
Trails A time (sec), HRB ^b		23.8	7.7	22.0	7.2
Trails B time (sec), HRB ^b		56.8	19.8	47.0	13.9
Digit span and digit symbol, raw score, WAIS-R ^c		16.8	4.1	16.9	3.8
Vigilance, total correct, CPT ^d		29.5	0.9	29.2	1.6
Memory: Verbal (VMEM)					
Logical memory immediate recall, WMS-R ^e	35	27.1	6.4	30.5	6.0
Logical memory delayed recall, WMS-R ^e		23.2	7.1	28.1	6.6
Learning trials 1 through 5, CVLT ^f		59.3	9.6	63.4	8.5
Spatial (SMEM)	10				
Design reproduction immediate recall, WMS-R ^e		36.4	3.5	36.4	3.6
Design reproduction delayed recall, WMS-R ^e		35.1	4.7	34.4	4.0
Intellectual: Language (LAN)					
Controlled oral word association, MAE ^g	30	44.8	9.3	48.6	11.3
Animal naming, BDAE ^h		24.7	4.7	25.5	4.7
Reading sentences and paragraphs, BDAE ^h		9.7	0.6	9.8	0.4
Token test, 12-item version, MAE ^g		23.2	1.4	23.1	1.7
Mattis aphasia exam visual naming		53.5	6.4	53.7	5.0
Spatial (SPA)	20				
Total correct, JOLO ⁱ		26.9	2.5	24.4	3.4
Block design, raw score, WAIS-R ^c		34.1	9.4	35.0	9.9
Sensorimotor: Sensory (SEN)					
Stereognosis time (sec.), right hand, LNNB ^j	5	8.1	3.5	8.0	2.6
Stereognosis time (sec.), left hand, LNNB ^j		7.4	4.5	6.9	2.6
Stereognosis errors, right hand, LNNB ^j		0	0	0.2	0.7
Stereognosis errors, left hand, LNNB ^j		0.1	0.6	0.2	0.7
Motor (MOT)	5				
Finger tapping, right hand, HRB ^b		49.9	7.1	45.2	6.0
Finger tapping, left hand, HRB ^b		45.1	8.8	42.1	5.4

^aWCST = Wisconsin Card Sorting Test (Heaton 1981).

^bHRB = Halstead Reitan Battery (Reitan and Woltson 1985).

^cWAIS-R = Wechsler Adult Intelligence Scale - Revised (Wechsler 1981).

^dCPT = Continuous Performance Test (Gordon 1986).

^eWMS-R = Wechsler Memory Scale - Revised (Wechsler 1987).

^fCVLT = California Verbal Learning Test (Delis et al. 1983).

^gMAE = Multilingual Aphasia Exam (Benton and Hamsher 1976).

^hBDAE = Boston Diagnostic Aphasia Exam (Goodglass and Kaplan 1983).

ⁱJOLO = Judgment of Line Orientation (Benton et al. 1975).

^jLNNB = Luria-Nebraska Neuropsychological Battery (Golden et al. 1991).

^kTotal time: 2 h, 35 min.

Computerized Neuropsychological Scan. The computerized neuropsychological tests were designed to yield quantitative measures of performance on behavioral domains that can be linked to regional brain function. All tasks were developed on Macintosh® computers using the PowerLaboratory® platform (Chute and Westall 1997). The test battery is listed in Table 2 including references for published tasks. Because not all tests in the computerized battery have been published, a brief review of the tasks is provided. The task development and validation process were detailed in Gur et al. (1992).

Briefly, the tasks are designed to be used in neuroimaging studies, yet provide sufficient psychometric sensitivity to enable their application in neuropsychological assessment. Tasks first underwent a process of construction and evaluation by a team of experimental and clinical neuropsychologists. This group took the tasks through conceptualization, initial item analysis, and construction of a preliminary version. The resulting version was submitted to psychometric study to assess reliability and construct validity and obtain normative data. Data were also obtained at that stage to document comparability of test versions and the effects of retesting (“practice effects”). The tests listed in Table 2 have

undergone this process and can be administered using desktop or portable computers. The tests are as follows:

Abstraction Inhibition and Working Memory Task (AIM). The AIM (Glahn et al. 2000) is designed as a measure of abstraction and concept formation with and without additional working memory loads. It presents subjects with five shapes: two in the upper right and two in the upper left corner of a computer screen, with a fifth target object appearing in the center of the screen, below the other stimuli (Figure 1a). The participant’s task is to pair the target object with the objects on either the left or right. On half the trials, an additional working memory maintenance requirement is superimposed on this basic module by adding a delay between the presentation of the target and other objects. Total number correct and median reaction time for correct responses were selected as performance measures.

Penn Inhibition Test (PIT). The PIT was based on the competing programs component of the Executive Control Battery (Goldberg et al. 1989). The task is divided into visual and auditory subtests. The visual subtest begins by presenting the subjects either an individual blue dot or two sequential blue dots, with the instruction to make a single key press for two dots and a double key

Table 2. Neuropsychological Performance of 92 Healthy Control Subjects on the Computerized Neurocognitive Scan

Cognitive Domain and Component Test Variables	REL ^j	Time ^k (min)	Male		Female	
			Mean	SD	Mean	SD
Executive: Abstraction/Flexibility (ABF)		19				
Abstraction inhibition and working memory, AIM ^a	0.78		48.4	5.1	48.1	5.9
Penn Inhibition Test, PIT	0.80		14.4	0.6	14.7	1.6
Computerized Ravens Progressive Matrices, CRPM ^b	0.88		47.0	6.6	48.3	6.7
Attention (ATT)		10				
Penn Continuous Performance Test, PCPT ^c	0.78		34.8	2.2	35.6	0.8
Computerized Stroop Test, CSTR ^d	0.90		34.7	4.6	34.8	2.4
Memory: Word (WMEM)						
Penn Word Memory Test, PWMT ^e	0.83	4	36.4	2.7	36.6	2.4
Facial (FMEM)						
Penn Face Memory Test, PFMT ^e	0.60	4	33.2	2.9	34.4	3.0
Spatial (SMEM)						
Visual Object Learning Test, VOLT ^f	0.87	9	69.0	6.7	69.9	5.0
Intellectual: Language (LAN)						
Penn Verbal Reasoning Test, PVRT ^g	0.76	5	79.2	12.3	75.7	12.4
Spatial (SPA)						
Computerized Benton Line Orientation Test, CJOLO ^h	0.87	6	25.5	4.2	22.9	5.4
Sensory-Motor ability (SM)						
Pursuit Rotor Task PRT ⁱ	0.79	3	-5.0	10.6	-4.0	8.2

^aAIM (Glahn et al. 2000).

^bCRPM (Raven 1960).

^cPCPT (Kurtz et al. in press).

^dCSTR (Stroop 1935).

^ePWMT and ^fPFMT (Gur et al. 1993).

^fVOLT (Glahn et al. 1997).

^gPVRT (Gur and Reivich 1980).

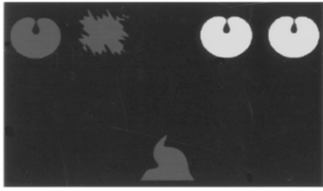
^hCJOLO (Benton et al. 1975).

ⁱPRT (Chute and Westall 1997).

^jREL = Cronbach’s Alpha.

^kTotal time: 1 h.

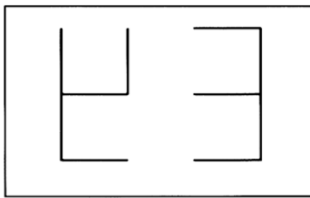
a. AIM



c. PFMT



b. PCPT



d. VOLT

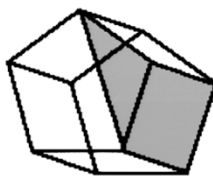


Figure 1. Examples of stimuli used in the neurocognitive scan.

press for a single dot. Individuals are required to perform to a criterion (10 sequential correct responses) on each stimulus type before advancing to the next stage. During the second stage, subjects are presented with 20 trials in which stimuli are randomly alternated. This is followed by a final stage in which each stimulus type is presented for nine trials followed by a tenth trial, during which the alternate stimulus is presented. The auditory subtest follows the same format, but instead of visual dots, presents auditory clicks. The individual is instructed to make a single key press when they hear two sequential clicks and a double key press when they hear a single click. Total number correct and median reaction time for correct responses were selected as performance measures.

Raven's Progressive Matrices. This is a computerized version of the standard paper and pencil task (Raven 1960). It is a multiple-choice task that requires subjects to conceptualize spatial, design, and numerical relationships that range in difficulty from very easy to increasingly complex. The computerized version was constructed by scanning and digitizing the stimulus cards from the original task. Instructions and scoring follow standard published procedures. Total number correct and median reaction time for correct responses were selected as performance measures.

Stroop. This is a computerized version of the standard paper and pencil task (Stroop 1935), designed to test an individual's ability to shift his or her perceptual set to conform to changing task requirements. The task requires subjects to name the color of the ink in which the words

are written quickly, ignoring the content of the word. On some trials, the stimuli are incongruent (e.g., the word "blue" is printed in yellow ink), and the subject must inhibit the prepotent response of reading the word, rather than the color in which it is printed, to respond correctly. Because of problems with voice recognition software, the computerized version of the task does not accept oral responses and, instead, requires subjects to make a button press on a computer game pad that is configured with red, blue, yellow, and green keys. Therefore, on incongruent trials on the computerized version, the subject must press the colored button that matches the color of the word and inhibit pressing the colored button that matches the content of the word. The task randomly presents 30 congruent and 30 incongruent trials. Total number correct and median reaction time for correct responses were selected as performance measures.

Penn Continuous Performance Test (PCPT; Kurtz et al. in press). During this task, the participant is asked to respond to a set of vertical and horizontal lines (a seven-segment display) whenever they form a digit (Figure 1b). Because each judgment is made on the basis of the present stimulus, working memory demands are minimized. Total number of true positive responses and median reaction time for true positive responses were selected as performance measures.

Penn Word Memory Test (PWMT; Gur et al. 1993). The PWMT is a forced-choice recognition task in which participants are shown 20 target words and asked to try to remember them. The encoding trial is followed by immediate and 20 min delayed recognition trials, dur-

ing which subjects are presented with the target words mixed with 20 distractors. New distractors are presented at each delay (for a total of 40) and are equated for frequency, length, concreteness, and imageability using Paivio's norms (Paivio et al. 1968). Total number of true-positive responses and median reaction time for true-positive responses were selected as performance measures.

Penn Face Memory Test (PFMT; Gur et al. 1993). The PFMT parallels the PWMT and consists of 20 target faces and 40 foils (20 for each test trial). Stimuli are black and white photographs of faces balanced for gender and age (Figure 1c). All faces are of neutral emotional expression, as determined by 12 raters. Procedures for administration and scoring of the task are identical to those for the Word Recognition Task. Total number of true-positive responses and median reaction time for true-positive responses were selected as performance measures.

Visual Object Learning Test (VOLT; Glahn et al. 1997). The VOLT was designed as a spatial analog of the California Verbal Learning Test (Delis et al. 1983). It uses 20 Euclidean shapes as learning stimuli (Figure 1d) that are presented over four learning trials, followed by short and long delay test recall. New distractor shapes are used in every test trial. Total number and median reaction time of true-positive responses across learning trials, short, and long delay were selected as performance measures.

Penn Verbal Reasoning Test (PVRT; Gur et al. 1987). The PVRT consists of age-appropriate verbal analogy problems. Individuals are presented with 30 analogies in a multiple-choice format. Total number correct and median reaction time for correct responses were selected as performance measures.

Computerized Judgment of Line Orientation (CJOLO). This is a computerized adaptation of the original paper and pencil task (Benton et al. 1975). Participants are shown two lines at an angle and are asked to indicate the corresponding lines on a simultaneously presented array. Task difficulty is defined by the length of the stimulus lines. Instructions and scoring follow standard published procedures. Total number correct and median reaction time for correct responses were selected as performance measures.

Pursuit Rotor Task (PRT). This is a standard rotor-pursuit paradigm incorporated into the PowerLaboratory® platform (Chute and Westall 1997). The task consists of five 60 s blocks during which the subject is required to trace a dot moving around the circumference of a circle at a rate of 100 mm/s. Blocks are separated by 30 s, and subjects use a light pen on a digitizing tablet to make the response. The difference in time on target (in

ms) during trial 5 minus time on target during trial 1 is calculated as a learning score and selected as a performance measure.

These individual scores were grouped to scan the following functional domains: abstraction and flexibility (ABF), attention (ATT), verbal memory (VMEM), face memory (FMEM), spatial memory (SMEM), language (LAN), spatial abilities (SPA), and sensorimotor (SM). The variables comprising these functions are presented in Table 2. Note that most assignments of tests to domains are straightforward and unambiguous. The exception is the assignment of the Stroop as an attentional test, when it could be argued that it measures mental flexibility and belongs to ABF. Changing the assignment of this measure did not affect the over-all results, and analyses comparing traditional with computerized measures combined ATT and ABF into a single "executive" domain.

Procedures

After establishing that participants were healthy, the purpose of the study was explained and informed consent was obtained. The traditional battery and the computerized scan were administered in a counterbalanced order within a week (mean \pm SD 2.4 \pm 1.7 days). All participants received both procedures. Tests were administered in a standard testing room, which was well lit and quiet. The traditional battery was administered by trained neuropsychology Fellows; whereas, the computerized scan was administered by Fellows or research assistants.

For both the traditional battery and the computerized scan, procedures were included to ensure that participants understood the instructions and could provide valid responses. For the traditional battery, this was done by interaction with the examiner. For the computerized scan, subjects first activated a program that enabled them to demonstrate facility in using the pointing device, and a brief training procedure preceding each test evaluated comprehension of test instructions and response requirements.

Data Analysis

The tests in both batteries were grouped into their functional domains using z-scores as described above. To test the sensitivity of each battery to sex differences in the neurocognitive profiles, a "female typical" gradient (Gur et al. 1999) was calculated as the average memory score (where women are expected to perform better than men) minus the spatial score (where men are expected to outperform women). For the traditional battery, the memory score included verbal and spatial measures; whereas, for the computerized scan, we added the face memory score. The hypothesis of sensi-

tivity to sex differences was tested by determining whether the female typical gradient differed from zero within sex, higher than 0 in women, and lower than 0 in men (paired *t*-tests, a stringent test of the hypothesis). We also compared this gradient between males and females to evaluate the hypothesis that the score is higher in women than in men (between-group *t*-test, a less stringent test).

Sensitivity to the effects of aging was evaluated by calculating the Pearson product moment correlations between age and performance on the functional domains. It was hypothesized that age effects are more pronounced for frontotemporal functions: abstraction, attention, and memory. The computerized scan also permitted testing the hypothesis that age associated decline is more pronounced for speed than for accuracy.

Direct comparison of the two approaches is hampered by differing goals and the confounding of speed and accuracy in the traditional battery. However, such comparison is desirable to establish test order effects and to examine their correlation. To overcome the greater targeted specificity of the computerized battery and consequent lack of overlap in specific neurocognitive domains, we have combined attention and abstraction/flexibility on both batteries into an "executive functioning" measure, and word, face, and spatial memory on the computerized battery to compose a "memory" measure that was compared with the average of oral and spatial memory on the traditional battery. Sensory was averaged with motor domains on the traditional battery profile to compare to the sensorimotor domain on the computerized profile. To overcome the difficulty of an inappropriate correlation of measures that separate accuracy from speed with those that do not, we have generated an "efficiency" index for the computerized battery. Efficiency was defined as the score for accuracy divided by the score for speed (performance accuracy per unit of time). Order effects were then examined by analysis of variance (ANOVA) on each domain measure, with order (dummy coded 0 = traditional first, 1 = computerized first) as a grouping factor and approach (traditional, computerized) as a repeated-measures (within-group) factor. The correlation matrix between the traditional and computerized measures was calculated to examine comparability of scores.

RESULTS

Table 1 provides the means of men and women on the tests comprising the traditional battery, and Table 2 provides these means for the computerized scan. As can be seen in Table 2, the tests comprising the computerized scan had moderate to very high reliability. For comparability across tests and domains, subsequent analyses were performed on the z-transformed data.

Sex Differences

The "female typical" gradient showed the hypothesized direction of being higher for women than for men in both the traditional battery and the computerized scan. For the traditional battery, the gradient was significantly negative for men: -0.37 ± 0.66 , *t* (paired) = 3.73, $p < .001$, but was not significantly positive for women: 0.11 ± 0.79 , *t* (paired) < 1, ns. However, the difference between men and women was significant, between-group $t = 3.16$, *df* = 90, $p < .001$. For the computerized scan, the gradient was significantly negative for men: -0.39 ± 1.05 , *t* (paired) = 1.99, $p < 0.025$, and significantly positive for women: 0.39 ± 0.85 , *t* (paired) = 2.66, $p < .01$. The between-group difference was also significant, $t = 3.24$, *df* = 90, $p < .001$.

The profiles of men and women (Figure 2) on the traditional battery (top bar) and computerized scan (bottom bars) indicated considerable uniformity of variance for both, with no significant differences between variances by Bartlett's test. The two batteries yielded similar profiles of sex differences with the exception that the traditional battery did not show a sex difference in attention; whereas, the computerized scan showed better performance for women. Within the memory measures the traditional battery showed a more pronounced superiority in females for oral memory; whereas, the measure of face memory, added to the computerized scan, also showed better performance in women. The accuracy and speed measures indicated that sex differences can be seen in specific aspects of performance. Although women were both more accurate and faster for the attention and face memory measures, men were more accurate but not faster for the spatial-processing measures.

Age Effects

For the traditional battery, ABF, and VMEM showed significant decline with age, $r = -0.31$, and -0.24 , respectively, *df* = 90, $p < .01$, one-tailed. An examination of the accuracy and speed measures provided by the computerized scan showed few correlations between age and accuracy, ABF, and SMEM, -0.49 and -0.31 , respectively. By contrast, correlations between age and speed were evident in most domains including: ABF -0.46 , ATT -0.25 , FMEM -0.35 , SMEM -0.23 , LAN -0.24 , SPA -0.26 . all $p < .01$, one-tailed.

Order Effects and Correlations Between Traditional and Computerized Measures

No main effects or interactions of order of administration have approached statistical significance in the ANOVA for any of the functional domain scores. The correlations between the domain scores from the traditional battery and comparable "efficiency" scores from the computer-

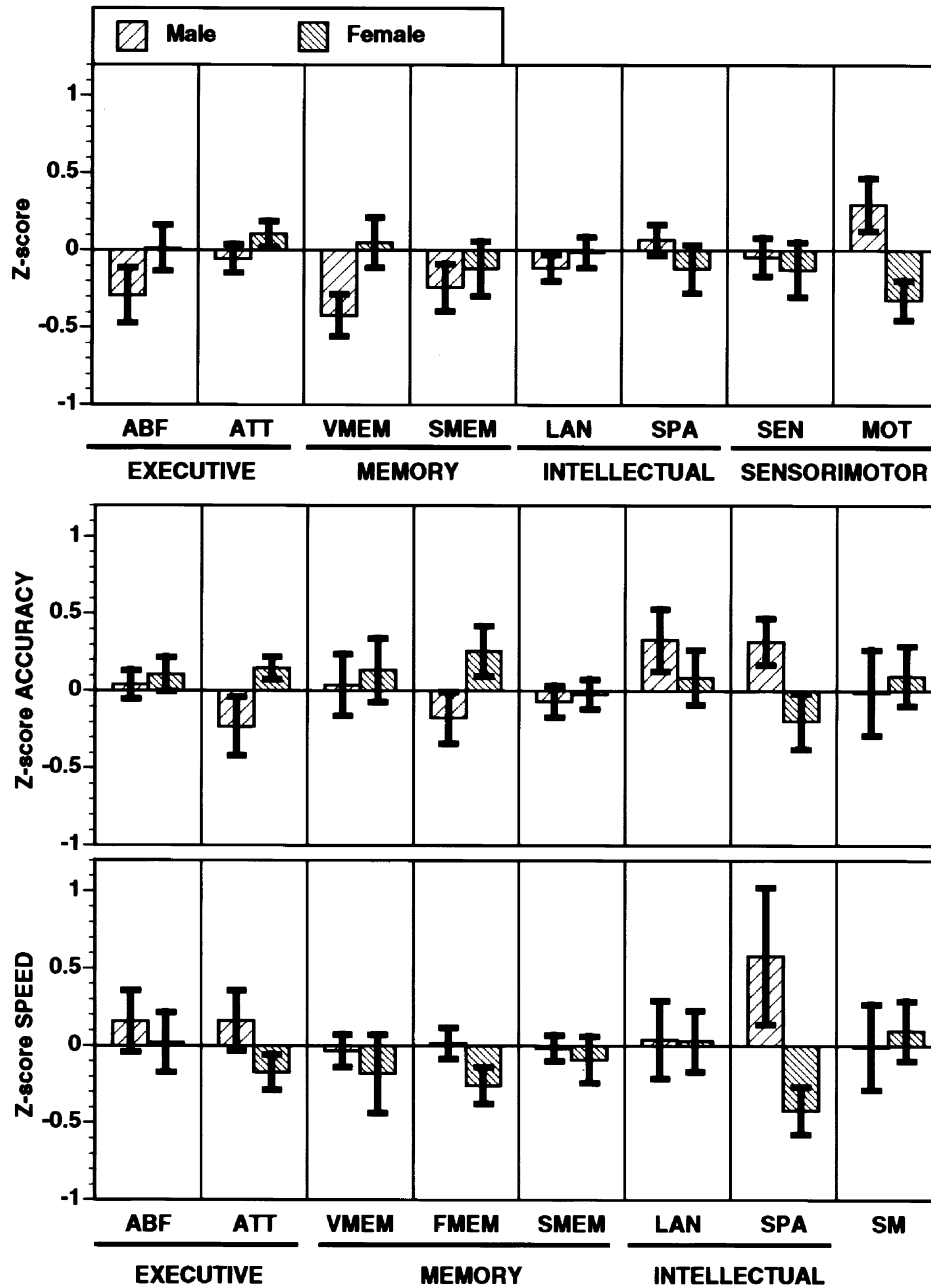


Figure 2. The neurocognitive profile of men and women (means \pm SEM) on the traditional battery (top bar) and the computerized scan accuracy (middle), and speed (bottom), eurocognitive domains as defined in Methods.

ized battery are presented in Table 3. As can be seen, the correlations between comparable domains (diagonal) are significant and higher than for correlations with other domains. Note that the correlations are moderate and for the “sensorimotor” domain the correlation is low.

DISCUSSION

The computerized scan approach seems to be feasible with healthy people, who responded favorably, were

able to follow instructions, and performed at over 70% accuracy levels. Participants who had no previous experience with computers could be trained using the training module. Furthermore, older adults had no difficulties with the display and procedures.

The computerized scan was at least as sensitive as the traditional battery to sex differences in cognitive performance. An index reflecting the hypothesized “female typical” gradient of superiority in memory relative to spatial processing showed a similar magnitude of sex differences for the computerized scan and the traditional

Table 3. Correlations Between the Traditional Measures of Major Cognitive Domains and the Computerized "Efficiency" Measures for the Comparable Domains

Traditional		Computerized			
		Executive Memory	Intellectual	Sensorimotor	
Executive	r	0.52	0.18	0.35	0.26
	p	0.0001	0.0982	0.0036	0.0221
Memory	r	0.26	0.53	0.32	0.17
	p	0.0211	0.0001	0.0079	0.1101
Intellectual	r	0.20	0.14	0.67	0.17
	p	0.0818	0.1949	0.0001	0.1185
Sensorimotor	r	-0.18	0.07	0.15	0.28
	p	0.1011	0.5571	0.1133	0.0189

battery. Both methods were able to provide the less stringent support for the hypothesis by showing a significant difference between men and women in the expected direction of women having more positive values than men. However, although the traditional battery provided the more stringent support of a significant within-sex deviation from zero only for men, the computerized scan met this criterion for both men and women.

A more detailed evaluation of the major neurocognitive domains indicated that, although verbal memory showed the largest difference favoring females on the traditional battery, the difference for word memory was marginal for the computerized scan. This probably reflects the inclusion in the traditional battery of the CVLT, which examines verbal learning. On the other hand, the face memory task from the computerized scan showed a robust difference favoring females. The computerized scan also revealed a sex difference in attention, with women showing better performance on this version of the CPT. Although poorer performance of men on attentional tasks seems consistent with the higher preponderance of attention deficit disorders in boys (Rhee et al. 1999), the issue of whether there are sex difference in attention, or whether they are stimulus-dependent, is still debatable (Dittmar et al. 1993; McGivern et al. 1997). Perhaps the addition of working memory demands in other versions of the CPT mask the better ability of women to attend vigilantly.

The advantage of the computerized scan in providing separate measures of speed and accuracy was manifested in the ability to characterize sex differences better in performance. Although women are both more accurate and faster than men for the face memory task, the better performance of women in attention is more pronounced for accuracy than for speed. In men, the better performance on spatial tasks was evident only for accuracy with speed being equal to that of women. These complementary aspects of performance may also be differentially linked to neural activation as can be assessed in functional imaging studies (Gur et al. 1988, 1997).

Both methods were about equally sensitive to the effects of healthy aging. Furthermore, as would be expected in this healthy population, the associations between cognitive performance and chronological age, although significant, were consistently small for both traditional and computerized measures. In both, aging showed an effect on measures related to frontotemporal functioning. Here again, however, the computerized scan had the advantage of enabling better specification of the facet of performance more vulnerable to aging. Consistently, the correlations with age were more pronounced for speed than for accuracy. This is compatible with other studies evaluating age effects on performance (Zec 1995). Thus, decline in performance that includes accuracy can be interpreted as reflecting processes distinguishable from speed-related decline associated with healthy aging.

Global measures of functional domains showed moderate correlations between the traditional and computerized measures, with the exception of the sensorimotor domain, where the measures were uncorrelated. Although the moderate correlations among comparable domains are encouraging, they are not sufficiently high to suggest that the approaches are interchangeable. Thus, correlations of 0.52 between the computerized and the traditional measures of executive functions and 0.53 between the respective measures of memory are insufficient to ensure an investigator who has been using the traditional battery that equivalent measures will be obtained with the computerized approach. However, these global measures tap areas that do not entirely overlap. The computerized tasks were designed to be used in functional imaging studies and were, therefore, targeted to very specific neurocognitive domains. The traditional battery, by contrast, consists of tests that were psychometrically designed to tap broader domains, emphasizing sensitivity to deficits associated with brain disorders. It is noteworthy that when individual computerized tasks, included in the present battery, have been directly compared with traditional measures, they yielded high correlations and reasonable construct validity (Glahn et al. 2000; Kurtz et al. 2001). Additional computerization of tasks not currently covered by the scan is needed to generate a set of measures more nearly equivalent to the traditional parameters.

The cognitive scan in its present form has several limitations. Most importantly, it does not incorporate the technology needed for assessing verbal learning as is traditionally done with the CVLT. We have evaluated necessary voice recognition technology and have not been able to configure the appropriate platform for such testing. The scan is also limited to measuring cognitive domains and does not include measures of emotion processing. However, computerized measures of facial affect discrimination have been developed (e.g., Kohler et al. 2000) and could be incorporated in future

studies. Although other tests could also be added, the purpose of this scan is to provide a time-efficient, reliable, and error-free estimate of the neurocognitive profile.

The computerized scan approach seems to be ready for clinical and research applications. Although it still lacks measures requiring voice recognition, it is time efficient, easy to administer, and provides data that assess both accuracy and speed. The computerized format also facilitates data acquisition, transfer, and analysis, features that may be essential for large-scale collaborative studies. Finally, the availability of the worldwide web makes it convenient to upgrade and update versions for distribution, assuring uniform data collection. Our finding that the scan is sensitive to sex differences and age effects in healthy people suggests that it would be sensitive to the less subtle effects of neuropsychiatric illness. The ability to administer the computerized scan to young and older adults is encouraging, and we are currently evaluating the feasibility of its extension to children.

ACKNOWLEDGMENTS

This research was supported by NIH Grants MH-43380, MH-42191, MH-01336, MH-53326, MO1RR0040, the EJLB Foundation and the Bosworth Family Fund. We thank Lyn Harper Mozley for invaluable help and advice, Brian C. Klimas for help in data collection and management, and the dedicated research assistants and Fellows who administered the tests.

REFERENCES

- Baker EL, Letz RE, Fidler AT, Shalat S, Plantamura D, Lyndon M (1985): A computer-based neurobehavioral evaluation system for occupational and environmental epidemiology: Methodology and validation studies. *Neurobehav Toxicol Teratol* 7:369–377
- Benton AL, Hamsher K (1976): *Multilingual Aphasia Examination*. Iowa City, IA, University of Iowa
- Benton AL (1994): Neuropsychological assessment. *Ann Rev Psychol* 45:1–23
- Benton AL, Hannay JJ, Varney NR (1975): Visual perception of line direction in patients with unilateral brain disease. *Neurology* 25:907–910
- Benton AL, Sivan AB, Des Hamsher K, Varney NR, Spreen O (1994): *Contributions to Neuropsychological Assessment: A Clinical Manual*. Oxford, UK, Oxford University Press
- Caplan PJ, Crawford M, Hyde JS, Richardson JTE (1997): *Gender differences in human cognition*. New York, Oxford University Press
- Censits DM, Ragland JD, Gur RC, Gur RE (1997): Neuropsychological evidence supporting a neurodevelopmental model of schizophrenia: A longitudinal study. *Schiz Res* 24:289–298
- Chute DL, Westall RF (1997): Power Laboratory. Devon PA, MacLaboratory Incorporated
- Collins DW, Kimura D (1997): A large sex difference on a two-dimensional mental rotation task. *Behav Neurosci* 111:845–849
- Delis DC, Kramer JH, Kaplan E, Ober BA (1983): *California Verbal Learning Test (CVLT) manual*. New York, The Psychological Corporation
- Dittmar ML, Warm JS, Dember WN, Ricks DF (1993): Sex differences in vigilance performance and perceived workload. *J Gen Psychol* 120:309–322
- Earles JL, Kersten AW (1999): Processing speed and adult age differences in activity memory. *Exper Aging Res* 25:243–253
- Glahn DC, Cannon TD, Gur RE, Ragland JD, Gur RC (2000): Working memory constrains abstraction in schizophrenia. *Biol Psychiat* 47:34–42
- Glahn DC, Gur RC, Ragland JD, Gur RE (1997): Reliability, performance characteristics, and construct validity and initial application of the visual object learning test (VOLT). *Neuropsychology* 11:602–612
- Gold JM, Randolph C, Carpenter CJ, Goldberg TE, Weinberger DR (1992): Forms of memory failure in schizophrenia. *J Abstr Psychol* 101:487–494
- Goldberg E, Bilder R, Jaeger J, Podell K (1989): *Examination of frontal system functions (The Executive Control Battery)*. unpublished manual
- Goldberg TE, Weinberger DR, Berman KF, Pliskin NH, Podd MH (1987): Further evidence for dementia of the prefrontal type in schizophrenia? *Arch Gen Psychiat* 44:1008–1014
- Golden CJ, Purisch AD, Hammeke TA (1991): *Luria-Nebraska Neuropsychological Battery: Forms I and II*. Los Angeles, Western Psychological Services
- Goodglass H, Kaplan E (1983): *The assessment of aphasia and related disorders*, 2nd ed. Philadelphia, Lea & Febiger
- Gordon M (1986): *Manual, Gordon Diagnostic System*. New York, Gordon Diagnostic Systems Inc.
- Gur RC, Erwin RJ, Gur RE (1992): Neurobehavioral probes for physiologic neuroimaging studies. *Arch Gen Psychiat* 49:409–414
- Gur RC, Jaggi JL, Ragland JD, Resnick SM, Shtasel D, Muenz L, Gur RE (1993): Effects of memory processing on regional brain activation: Cerebral blood flow in normal subjects. *Int J Neurosci* 72:31–44
- Gur RC, Gur RE, Obrist WD, Skolnick BE, Reivich M (1987): Age and regional cerebral blood flow at rest and during cognitive activity. *Arch Gen Psychiat* 44:617–621
- Gur RC, Gur RE, Skolnick BE, Resnick SM, Silver FL, Chawluk JB, Muenz L, Obrist WD, Reivich M (1988): Effects of task difficulty on regional cerebral blood flow: Relationships with anxiety and performance. *Psychophysiology* 25:392–399
- Gur RC, Ragland JD, Mozley LH, Mozley PD, Smith R, Alavi A, Bilker W, Gur RE (1997): Lateralized changes in regional cerebral blood flow during performance of verbal and facial recognition tasks: Correlations with performance and "Effort." *Brain Cogn* 33:388–414

- Gur RC, Reivich M (1980): Cognitive task effects on hemispheric blood flow in humans: Evidence for individual differences in hemispheric activation. *Brain Lang* 9:78–92
- Gur RC, Saykin AJ, Benton A, Kaplan E, Levin H, Kester DB, Gur RE (1990): Behavioral imaging: III. Inter-rater agreement and reliability of weightings. *Neuropsychiat Neuropsychol Behav Neurol* 3:113–124
- Gur RC, Turetsky BI, Matsui M, Yan M, Bilker W, Hughett P, Gur RE (1999): Sex differences in brain gray and white matter in healthy young adults. *J Neurosci* 19:4065–4072
- Halstead WC (1947): *Brain and Intelligence*. Chicago, University of Chicago Press
- Heaton RK (1981): *Wisconsin card sorting test manual*. Lutz, FL, Psychological Assessment Resources
- Heinrichs RW, Zakzanis KK (1998): Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology* 12:426–445
- Jarvis P, Barth J (1994): *The Halstead–Reitan Neuropsychological Battery*. Lutz, FL, Psychological Assessment Resources
- Kaplan E (1990): The process approach to neuropsychological assessment of psychiatric patients. *J Neuropsych Clin Neurosci* 2:72–87
- Kimura D, Harshman RA (1984): Sex differences in brain organization for verbal and nonverbal functions. *Prog Brain Res* 61:423–441
- Kohler CG, Bilker WB, Hagendoorn M, Gur RE, Gur RC (2000): Emotion recognition deficit in schizophrenia: Association with symptomatology and cognition. *Biol Psychiat* 48:127–136
- Kramer JH, Delis DC, Kaplan E, O'Donnell L, Prifitera A (1997): Developmental sex differences in verbal learning. *Neuropsychology* 11:577–584
- Kurtz MM, Ragland JD, Bilker WB, Gur RC, Gur RE (2001): Comparison of two forms of the continuous performance test with and without working memory demands in healthy controls and patients with schizophrenia. *Schiz Res* 48:307–316
- Laursen P (1997): The impact of aging on cognitive functions. An 11-year follow-up study of four age cohorts. *Acta Neurologica Scand: Suppl* 172:7–86
- McGovern RF, Huston JP, Byrd D, King T, Siegle GJ, Reilly J (1997): Sex differences in visual recognition memory: Support for a sex-related difference in attention in adults and children. *Brain Cogn* 34:323–336
- Mirsky AF, Duncan CC (1986): Etiology and expression of schizophrenia: Neurobiological and psychosocial factors. *Ann Rev Psychol* 37:291–319
- Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C (1989): The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 39:1159–1165
- Paivio A, Yuille JC, Madigan SA (1968): Concreteness, imagery, and meaningfulness values for 925 nouns. *J Exp Psychol* 76:1–25
- Park DC, Smith AD, Lautenschlager G, Earles JL, Frieske D, Zwahr M, Gaines CL (1996): Mediators of long-term memory performance across the life span. *Psychol Aging* 11:621–637
- Ragland JD, Gur RC, Deutsch GK, Gur RE (1995): Reliability and validity of the paired-associate recognition memory test: A test of declarative memory using Wisconsin card sorting stimuli. *Psychol Assess* 7:25–32
- Raven JC (1960): *Guide to the standard progressive matrices*. London, H.K. Lewis
- Reitan RM, Davison LA (1974): *Clinical neuropsychology: Current status and applications*. New York, Hemisphere
- Reitan RM, Woltson D (1985): *The Halstead–Reitan Neuropsychological Test Battery, Theory, and Clinical Interpretation*. Tucson, AZ, Neuropsychology Press
- Rhee SH, Waldman ID, Hay DA, Levy F (1999): Sex differences in genetic and environmental influences on DSM-III-R attention deficit/hyperactivity disorder. *J Abnorm Psychol* 108:24–41
- Ruff RM, Light RH, Quayhagen M (1989): Selective reminding tests: A normative study of verbal learning in adults. *J Clin Exp Neuropsychol* 11:539–550
- Saykin AJ, Gur RC, Gur RE, Mozley D, Mozley LH, Resnick SM, Kester DB, Stafiniak P (1991): Neuropsychological function in schizophrenia: Selective impairment in memory and learning. *Arch Gen Psychiat* 48:618–624
- Saykin AJ, Gur RC, Gur RE, Shtasel DL, Flannery KA, Mozley LH, Malamut BL, Watson B, Mozley PD (1995): Normative neuropsychological test performance: Effects of age, education, gender, and ethnicity. *Appl Neuropsychol* 2:79–88
- Saykin AJ, Shtasel DL, Gur RE, Kester DB, Mozley LH, Stafiniak P, Gur RC (1994): Neuropsychological deficits in neuroleptic naive, first episode schizophrenic patients. *Arch Gen Psychiat* 51:124–131
- Sharma T, Harvey P (eds) (1999): *Cognition in Schizophrenia*. Oxford, Oxford University Press
- Shtasel DL, Gur RE, Mozley PD, Richards J, Taleff MM, Heimberg C, Gallacher F, Gur RC (1991): Volunteers for biomedical research: Recruitment and screening of normal controls. *Arch Gen Psychiat* 48:1022–1025
- Silverman I, Phillips K, Silverman LK (1996): Homogeneity of effect sizes for sex across spatial tests and cultures: Implications for hormonal theories. *Brain Cognit* 31:90–94
- Sliwinski M, Buschke H (1999): Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychol Aging* 14:18–33
- Spitzer RL, Williams JB, Gibbon M (1996): *Structured clinical interview for DSM-IV: Nonpatient version (SCID-NP)*. New York, New York State Psychiatric Institute
- Stroop JR (1935): Studies of interference in serial verbal reactions. *J Exp Psychol* 18:643–662
- Verhaeghen P, Salthouse TA (1997): Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychol Bull* 122:231–249
- Wechsler DA (1981): *Wechsler Adult Intelligence Scale Revised (WAIS-R)*, Manual. Cleveland, OH, The Psychological Corp
- Wechsler DA (1987): *Wechsler Memory Scale, Revised (WMS-R)*. Cleveland, OH, The Psychological Corp
- Zec RF (1995): The neuropsychology of aging. *Exp Gerontol* 30:431–442