ENVIRONMENTAL RESEARCH 62, 125–147 (1993)

Comparison of Performance from Three Continents on the WHO-Recommended Neurobehavioral Core Test Battery[1]

W. KENT ANGER,* MARIA G. CASSITTO,† YOU-XIN LIANG,‡ RAFAEL AMADOR,§ JACOB HOOISMA,|| DAVID W. CHRISLIP,¶ DONNA MERGLER,** MATT KEIFER,†† JOSEPH HÖRTNAGL,‡‡ LIONEL FOURNIER,§§ BOHDAN DUDEK,|||| AND E. ZSÖGÖN¶¶

*The Oregon Health Sciences University, Portland, Oregon 97201; †University of Milan, IOH, Milan, Italy; ‡Shanghai Medical University, Shanghai, People's Republic of China; §National Autonomous University of Nicaragua, Leon, Nicaragua; ||Medical Biological Laboratory TNO, Rijswijk, The Netherlands; ¶Division of Biomedical and Behavioral Sciences, National Institute for Occupational Safety and Health, Cincinnati, Ohio 45226; **University of Québec in Montréal, Montreal, Canada; ††CARE, Leon, Nicaragua; ††University of Washington, Seattle, Washington 98195; ‡‡University of Innsbruck, Innsbruck, Austria; §§Hôpital F. Widal, Paris, France; ||||Nofer's Institute of Occupational Medicine, Lodz, Poland; and ¶¶National Institute of Occupational Health, Budapest, Hungary

Received July 8, 1992

To address the need for standardized test batteries, an expert group convened by the World Health Organization (WHO) and the U.S. National Institute for Occupational Safety and Health during 1983 proposed the Neurobehavioral Core Test Battery (NCTB) to identify nervous system effects of chemical exposures in human populations worldwide. To determine the feasibility of using the NCTB in varied cultures, a cross-cultural assessment was conducted under WHO auspices. Data were collected in 10 countries of Europe, North and Central America, and Asia from over 2300 males and females who were not exposed to chemicals at work, within five age ranges between 16 and 65. Results suggest that performance on two NCTB tests (Simple Reaction Time, Benton Visual Retention) is very similar in a broad range of countries and that performance on four other NCTB tests (Santa Ana, Digit Symbol, Digit Span, Aiming) is relatively more variable from country to country, in both males and females. However, data collected from very poorly educated males in one country revealed very low performance levels suggesting that the NCTB may not provide an adequate reference group for identifying (behavioral) neurotoxic effects in such populations. More research is thus needed on evaluating neurotoxicity in poorly educated subjects.

INTRODUCTION

Behavioral testing of occupational populations exposed to chemicals began in the 1960s (Hänninen, 1966) using the methods of neuropsychology and experimental psychology. Hänninen in Finland developed the first such behavioral test battery, refining it in a series of studies in the 1970s (Hänninen and Lindström, 1979). Research in this field has generally employed behavioral test batteries to assess the effects of neurotoxic or unknown chemicals, but the number of unique tests used in this field has expanded continuously. Johnson and Anger (1983) identified 60 unique tests that had been used in this research at the outset of the 1980s, and 250 unique tests had been employed in worksite behavioral research by the end of that decade (Anger, 1990).

Calls for standardized tests (e.g., Dews, 1975) were answered by a group con-

[1] Presented at the Fourth International Symposium on Neurobehavioral Methods and Effects in Occupational and Environmental Health, July 8–11, 1991, Tokyo, Japan.

0013-9351/93 $5.00

vened in Cincinnati, Ohio, during 1983 by the World Health Organization (WHO) and the U.S. National Institute for Occupational Safety and Health (NIOSH). That group proposed a battery which could identify or screen for nervous system effects of chemical exposures in human populations worldwide. This battery, the Neurobehavioral Core Test Battery (NCTB), is composed of seven behavioral tests: Digit Span, Digit Symbol, Benton Visual Retention, Simple Reaction Time, Santa Ana dexterity, Pursuit Aiming II, and Profile of Mood States (Johnson *et al.*, 1987).

Contemporaneously, Baker *et al.* (1985) developed a battery of 19 tests, the Neurobehavioral Evaluation System (NES), including variants of five of the NCTB tests, using computer implementation. The NES has become the most widely used behavioral test battery to assess the neurotoxicity of chemicals by computer (Letz, 1990; Anger, 1990), and the NCTB is the most widely used battery administered by a human tester (Cassitto *et al.*, 1990; Anger, 1990).

Sensitivity, Reliability and Validity of NCTB Tests

Although the NCTB is of recent origin, one or more of the seven tests in the NCTB (or very similar variants) have been employed widely in behavioral neurotoxicology research in occupational settings. Indeed, while the functional areas assessed by the NCTB were selected to represent a range of important nervous system functions, the main criterion applied in selecting individual tests was sensitivity to changes associated with chemical exposures. Practically, this meant that the tests had successfully discriminated between worker groups exposed to chemicals known to be neurotoxic (viz. lead, mercury, carbon disulfide) and unexposed worker groups (Johnson *et al.*, 1987).

Data from a recent comprehensive review (Anger, 1990) of worksite research published from 1966 through 1989 substantiates the sensitivity of the NCTB tests. The NCTB tests are listed in Table 1, along with the number of studies (column 2) in which each has been employed out of a total of 185 worksite studies identified in the review. The Pursuit Aiming II test has not been used in any such study, but the other six tests have been used in between 2 and 60 studies. Column 3 lists the number of studies in which the author reported that the exposed subjects differed significantly from the control subjects. As noted in column 4 of Table 1, the tests selected for the NCTB have identified statistically significant group differences (between chemical-exposed and unexposed referent groups) in 44–62% of the studies in which they have been employed (the number in column 3 divided by the

TABLE 1
NCTB TESTS, NUMBER OF WORKSITE STUDIES (OF 185) EMPLOYING EACH TEST, AND NUMBER OF THOSE STUDIES WHERE THE TEST IDENTIFIED SIGNIFICANT DIFFERENCES

Test	Studies employing test	Significant results	%
Santa Ana	29	18	62
Simple Reaction Time	52	27	52
Benton Visual Retention	20	10	50
Digit Symbol	60	27	45
Digit Span	39	17	44
Profile of Mood States	2	1	50
Pursuit Aiming Test	0	0	—

number in column 2). This compares favorably with an overall mean of 43% significant differences (calculated in the same way) reported for all 250 unique tests administered in the 185 studies surveyed. Within the functional categories selected (e.g., memory, coordination, speed), the NCTB tests are the most frequently employed in worksite research and among the most sensitive (defined as identifying significant differences) to chemical exposures of the tests employed in this field of research (Anger, 1990).

The two primary psychometric requirements of behavioral tests are reliability and validity. Each NCTB test was developed and extensively used in either neuropsychology or experimental psychology (Johnson *et al.*, 1987). It was thus assumed that they were reliable, although test–retest reliability data for these tests are limited (especially in normal adult populations) or inaccessible (unpublished). Since the NCTB was intended to be a screening instrument to identify chemical-induced health effects, traditional measures of content or construct validity (i.e., the tests assessed the functions intended) are of limited importance. Rather, criterion-related validity is the operative validation principle (Murphy and Davidshofer, 1988). Here, the criterion is the ability of the tests to reflect performance deficits in chemical-exposed groups when compared to a group which is not exposed to the chemicals under scrutiny. Since most worksite research is cross-sectional in nature, only one-time assessments are available to establish the validity criterion in any given population. Thus, replication of findings in different populations exposed to the same chemical using the same tests serves to establish the validity criterion for this area of research. An exclusive analysis of the NCTB tests has not been conducted. However, Anger's 1990 review demonstrates that control vs exposed group differences in populations exposed to lead, mercury, and carbon disulfide have been replicated extensively in independent studies using NCTB tests and, more broadly, evaluations of those functions assessed by NCTB tests. Therefore, the NCTB tests satisfy the validity criterion.

NCTB Cross-Cultural Assessment

What was not clear to the original group that proposed the NCTB as a battery was whether it could be employed in the wide array of cultural settings in which neurotoxic chemicals are employed (e.g., pesticides are used in virtually every country). A specific concern was that the NCTB tests were developed primarily in Western European or Western European-derived populations. Further, both construct validity and the criterion validity or sensitivity of the tests to chemical-exposed groups were also determined primarily in such populations. While the evidence is limited, research on cultural differences in performance testing indicates that some motor (e.g., Bernard, 1989) and cognitive processes are affected by racial or cultural differences, although it is difficult to eliminate educational factors in this research. However, most research employing intercountry or interculture comparisons suggests that cognitive processes are fundamentally similar in diverse cultures (D'Andrade, 1989). Nonetheless, the extent to which the NCTB tests would be reliable and sensitive to chemical-induced health effects in very different cultures is largely untested.

Under the auspices of WHO's Office of Occupational Health, a Cross-Cultural Assessment (CCA) was planned. The CCA's primary goal was to administer the NCTB to working populations unexposed to chemicals in a minimum of eight geographically dispersed countries representing diverse cultures. Specific objectives were to evaluate potential problems that could occur in the translation of instructions or test materials into different languages and in test administration to populations unfamiliar with such tests. This effort was thus aimed at assessing the NCTB's feasibility for testing subjects in diverse international populations. On a more utilitarian level, the CCA sought to develop for the NCTB tests baseline or normative data on diverse cultures in five age ranges between 16 and 65.

MATERIALS AND METHODS

Subjects

Listed in Table 2 are cities and countries in which there are verified reports that Cross-Cultural Assessment data have been collected from volunteer subjects who were selected from occupations, jobs, or companies in which there were no known exposures to neurotoxic chemicals. Subjects from most countries came from large (People's Republic of China, France, Hungary), medium (The Netherlands, United States), or small (Austria, Canada, Poland) urban settings (Italian subjects were from settings of all sizes) and lived in working class or entry-level white collar housing for that country. Nicaraguan subjects were peasants who lived and worked in a rural setting. Also listed in Table 2 are primary subject occupations, incentives given to motivate participation, and an estimate of the years of education of the groups studied. Most subjects in all countries had completed 7–9 years of education, or a few years more. U.S. subjects had the highest number of years of education completed (13–15), while most Nicaraguan subjects had completed 3 or fewer years of school.

It was impossible to provide chemical sampling data in the various countries where this study was conducted. Rather, the principals running the study in each country took the responsibility for seeing that subjects met the criterion of being "unexposed" to chemicals. Subjects in most cities are exposed to some level of pollution. Subjects from the one rural area studied, in Nicaragua, were selected from agricultural farms (coffee, cattle) which specifically did not use pesticides, and each worker's occupational history was taken to verify that there were no significant chemical exposures in the prior 2 years. The principal from Poland indicated concern that the control subjects from Lodz may have been exposed to pollutants in this highly industrialized area, but the lack of a defined comparison basis makes it difficult to offer reliable information on this matter.

Subjects from the United States were paid U.S. $25.00 to take the NCTB and seven NES tests (the latter not reported here) in 2 hours of testing. This amount exceeded slightly the subject's hourly wage in most cases; most testing was conducted outside regular working hours. Subjects from the Netherlands were paid Dfl 25 (approximately U.S. $15 in early 1991), and subjects in Nicaragua were paid the equivalent of 2 days wages (approximately U.S. $4 in early 1991) for taking the tests during working hours. Canadian subjects were paid Canadian $75.00 (approximately U.S. $66 in early 1991) to participate in a 6-hr test session which included four NCTB tests. Subjects from Austria, China, France (in the annual medical examination), Hungary (in preemployment or periodic employee medical examinations), Italy, and Poland were not paid for their performance but were tested during regular work hours, which can be considered "time off work."

The number of male and female subjects in each of the five selected age ranges is listed in Table 3 for each country included in the results. (Submitted data were

TABLE 2

STUDY SITES PROVIDING DATA TO THE CROSS-CULTURAL ASSESSMENT, PRINCIPAL'S RESPONSIBLE, PRINCIPAL OCCUPATION OF GROUP STUDIED, INCENTIVES PROVIDED, AND ESTIMATED YEARS OF EDUCATION

Country	Study site	Sponsor	Principal	Study group occupations	Incentive	Years education[e]
Europe (east)						
Hungary[a]	Budapest	Institute Occupational Health	E. Zsögön A. Jóni	Pesticide factory wkrs[f] (not exposed)	Time off work	13
Poland	Lodz	Institute Occupational Health	B. Dudek B. Bazylewicz-Walczak	Power plant wkrs, cooks, drivers, cleaners	Time off work	9–13
Europe (west)						
Austria	Wiesing	Dept. Social Medicine	J. Hörtnagl	Household, office wkrs	No	8–9
France[d]	Paris	University of Paris	L. Fournier	Auto manufacturing wkrs	Time off work	10
Italy	Chieti	Svcs. Occup. Medicine	C. Fanelli	Cleaners	Time off work	8
	Potenza	Svcs. Occup. Medicine	C. Fanelli	Metallurgical wkrs	Time off work	8
	Poggibonsi	Svcs. Occup. Medicine	C. Fanelli	Food market wkrs	Time off work	8
	Bari	Institute Occup. Health	C. Fanelli	Mechanical plant wkrs	Time off work	8
	Milan	University of Milan	M. Cassitto	Job applicants	Time off work	8
The Netherlands	Hague	TNO	J. Hooisma	City residents	Money	11
Asia						
People's Republic of China	Shanghai	Shanghai Medical University	Y. Liang Z. Chen	Embroidery wkrs, inspectors, clerks	Time off work	11–12
	Beijing[c]	Inst. Occup. Medicine	Z. Zhou	Waiters, teachers	Time off work	10
North America						
United States	Salt Lake City	NIOSH	K. Anger	Postal, hospital wkrs	Money	13–15
	Cincinnati	NIOSH	K. Anger	Hospital, insurance wkrs	Money	13–15
	Portland (OR)	NIOSH	K. Anger	Hospital, government wkrs	Money	13–15
Canada[b]	Beauhamois	University of Québec in Montréal	D. Mergler	Utility wkrs, police, manufacturing wkrs	Money	9–12
Central America						
Nicaragua[d]	Leon Matagalpa	Swedish International Aid Agency	R. Amador M. Keifer	Cattle farm, coffee wkrs	Money	3

[a] Due to equipment costs, the Santa Ana and Simple Reaction Time tests were not used in this country.
[b] Used only four NCTB tests (three reported here) in a study involving other tests; the NCTB tests were administered following prescribed NCTB training procedures.
[c] Data not presented in this article.
[d] Due to concern over translation uncertainties, the POMS was not administered.
[e] Approximated in some cases; some subjects in younger age groups were still in school and are thus not reflected in these estimates.
[f] Wkrs, Workers.

TABLE 3

NUMBER OF NCTB SUBJECTS BY COUNTRY, SEX, AND AGE RANGES FOR DATA REPORTED IN ARTICLE

	Male subjects					Female subjects					
Country	16–25	26–35	36–45	46–55	56–65	16–25	26–35	36–45	46–55	56–65	Totals
Europe (east)											
Hungary		24									24
Poland	19	28	25	17	10		16	14	16		145
Europe (west)											
Austria	30	19	15	12	21	31	18	12	19	25	202
France		24	22								46
Italy		192	11				16	14	16		249
The Netherlands	26	35	41		67	32	13	14		27	255
Asia											
People's Republic of China[a]	27	49	19	40		25	53	33	36		282
North America											
United States	80	73	85	50	29	140	163	162	92	41	915
Canada			70	37							107
Central America											
Nicaragua	31	25	28.	12	10						106
Totals	213	469	316	168	137	228	279	249	179	93	2331
			Male	subtotal	1303			Female	subtotal	1028	

[a] Data from Shanghai only.

omitted when the number of subjects was less than 10 in a given age range.) Data exclusions due to equipment malfunctions or test interruptions variably reduced the number of subjects on some tests.

Tests

The NCTB tests are described below in the order the tests were presented. Each test description is brief, since they are thoroughly described in other publications (Johnson *et al.*, 1987; Cassitto *et al.*, 1990). The entire battery of tests was given in all countries except Canada (where only the Santa Ana, Benton, Digit Span, and Profile of Mood States (POMS) tests were administered), Hungary (where the Santa Ana and Simple Reaction Time tests were not given), and Nicaragua (where the POMS was not administered).

Profile of Mood States. This test contains 65 adjectives that describe different moods. Subjects were asked to indicate their mood states during the past week on a five-point scale ranging from "not-at-all" to "extremely."

Simple Reaction Time. The test of reaction time requires the subject to place the index finger of their dominant hand so it is just touching the response button and then to press the button whenever a nearby light is illuminated, in a series of 64 trials of random-length intervals (mean 5.6 sec; range 1.0–11.0 sec).

Digit Span. A series of digits are read to the subject who is asked to recall them immediately, in the order presented, and, subsequently, in the reverse of the order presented. Beginning with three-digit strings forward and two-digit strings backward, two sets of strings at each string length (i.e., number of digits) were presented until the subject missed both strings at a given length, terminating the test.

Santa Ana. This test has a base plate with square depressions and fitted pegs with a cylindrical upper section. The subject takes each peg in succession, lifts it from the depression, turns it 180 degrees, and puts it back as quickly as possible; the subject has 30 sec in which to turn as many pegs as possible.

Digit Symbol. The worksheet on this test contains a list of numbers (1–9) that are associated with nine simple symbols. Below this code list is a string of random digits (1–9) paired with blank squares in which the subject must draw the appropriate symbols (i.e., the symbol paired to their corresponding digit) as fast as possible in 90 sec.

Benton Visual Retention (Recognition form). This test contains 10 test cards, 2 presenting one geometric figure and 8 presenting three horizontally arranged geometric figures. Subjects are shown figure(s) on a card for 10 sec and then asked to recognize the same figure from a set of four similar alternatives shown subsequently on a second card.

Pursuit Aiming II. Subjects are required to use a moderately sharp pencil to place a dot in 2-mm-diameter circles for two 60-sec trials, following a back-and-forth pattern on a single page.

Procedures

Three WHO Collaborating Centres were identified as coordinating centers for the CCA. These were in Helsinki (Institute of Occupational Health) under Dr. H. Hänninen, in Milan (Institute of Occupational Health) under Dr. M. Cassitto, and in Cincinnati (National Institute for Occupational Safety and Health) under Dr. K. Anger. Each coordinating center assembled the NCTB tests into "kits" for distribution (loan), trained test administrators (to assure consistency of test administration), and collated results. An Operational Guide for the NCTB (WHO, unpublished; see References for availability), specifying detailed instructions and procedures for administering the seven NCTB tests in English, was developed and used as a procedural blueprint for each principal investigator.

Translation of the instructions was the responsibility of the host principal investigator, sometimes with assistance of a coordinating center. The people who administered the NCTB in the CCA were trained by Drs. Anger, Cassitto, or Hänninen (one of four testers in the People's Republic of China, Italy; one of three testers in The Netherlands; one of six testers in Nicaragua, Poland, United States) or they were trained by someone trained by Anger, Cassitto, or Hänninen (Austria; Canada; three of four testers in the People's Republic of China; two of three testers in The Netherlands; five of six testers in Nicaragua). The administrators in France and Hungary were self-taught using the NCTB Operational Guide. Training took between 4 and 80 hr, depending on the expertise (and translation needs) of the persons receiving training and their degree of preparation for the training. The instructions and information in the Operational Guide were the basis for the training. NCTB tests were administered in a variety of settings, including laboratories, offices, conference rooms, and clinics in companies, government buildings, hospitals, schools, apartment buildings, and trailers. Minimization of noise and other extraneous factors, as specified in the Operational Guide, was a responsibility of the principal investigator in each country. The degree of compliance with these guidelines was not monitored.

RESULTS

Results from some countries in the Cross-Cultural Assessment have been presented in tabular form in other publications (Cassitto *et al.*, 1990; Liang *et al.*, 1990). Those and other data are collected here to provide a graphic presentation of the two basic descriptive parameters, mean and standard deviation. Inferential statistics were not applied to the data summarized in this article because the primary goal was to provide a descriptive presentation of findings in several countries for the purposes of data exploration and generation of hypotheses, rather than to demonstrate (significant) differences. Thus, comments on group differences or trends are made purely at the descriptive level and are not necessarily statistically reliable.

Mean and standard deviation data from all locations and age ranges where a minimum of 10 subjects was obtained are presented in Figs. 1–10 for six NCTB tests. (POMS data are not presented due to the expectation of high population specificity and thus large between-country variability due to concerns about translating context-sensitive words.) Data were excluded where testing errors were identified (e.g., Digit Span data from Poland are omitted due to the use of a different scoring method, and Simple Reaction Time data from Nicaragua were omitted because the initial finger location on each trial was different than specified in the Operational Guide). Figures 1–6 include data for each country where at least three age ranges are represented for each NCTB test; data from male subjects are in the top panel and data from female subjects are in the bottom panel. Countries that provided data in only one or two age ranges (viz. France, Italy, Hungary, Canada) are represented only in Figs. 7–10. Figure 7 presents results on all test

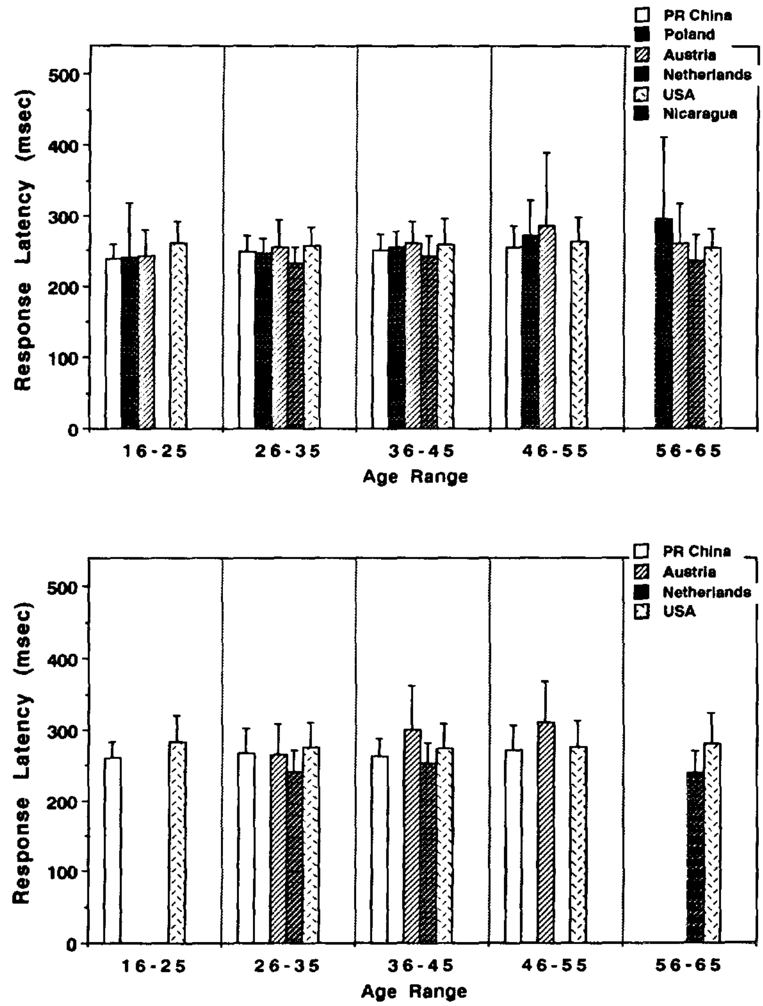


FIG. 1. Mean (±SD) response latency on the Simple Reaction Time test in age ranges between 16 and 65 for males (top) and females (bottom).

data from all countries with female subjects in the 26–35 age range; similarly, data from male subjects in the 26–35, 36–45, and 46–55 age ranges are presented in Figs. 8–10.

From the Simple Reaction Time test, Fig. 1 presents the mean response latency (from onset of light until the button was pressed) in milliseconds. The mean male (top panel) latencies are clustered tightly around 250 msec, and the mean response latencies in females (bottom panel) were between 239 and 311 msec.

Figure 2 presents the mean number of digit strings correctly recalled (sum of strings forward and backward) in the Digit Span test. (For example, if the subject recalled both strings of 3 and 4 digits then failed both of 5 digits forward, recalled one out of two strings of 2 then failed both at 3 backward, the score would be 5.) Male and female subjects remembered between 11 and 19 digit string sequences, with a small reduction in the 56–65 age range in subjects from Austria, The

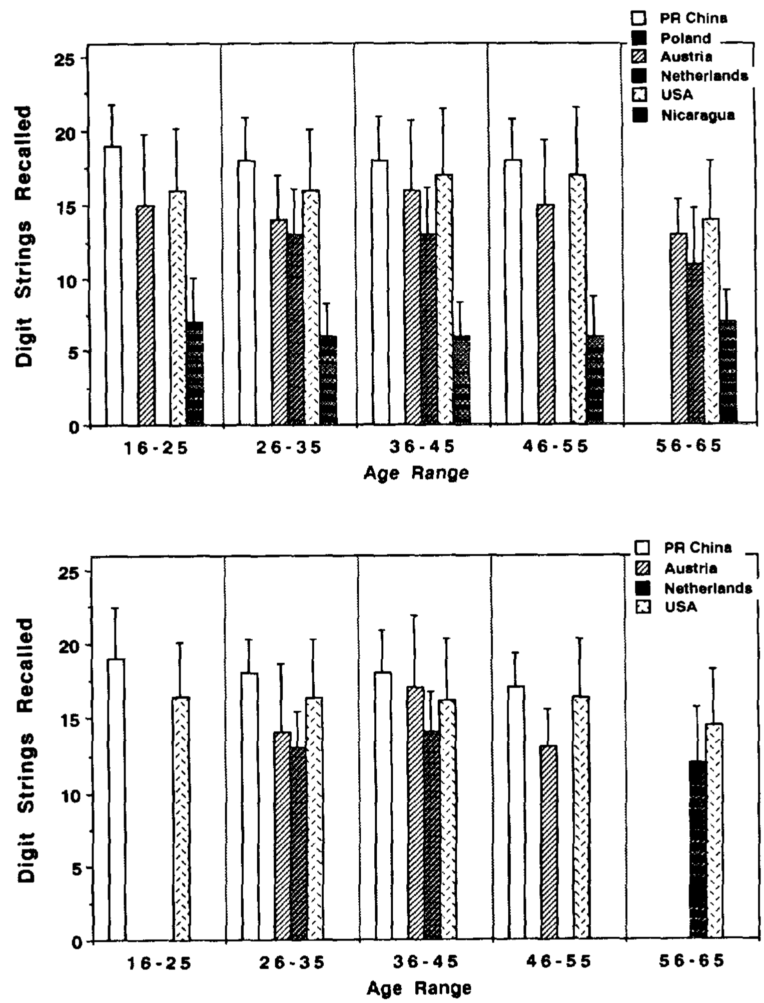


FIG. 2. Mean (±SD) number of digit strings recalled in the Digit Span test in age ranges between 16 and 65 for males (top) and females (bottom).

Netherlands, and the United States (within-country comparisons). Nicaragua (male subjects) is an exception in that the mean number of digit strings recalled was only 6–7 and there was no decline in the older age range. Performance within a given country is relatively consistent (similar) across the four age ranges between 16 and 55 in both male and female subjects.

From the Santa Ana dexterity test, Fig. 3 presents the mean total number of pegs turned by the preferred (dominant) hand on two 30-sec trials. Total pegs turned ranged from 28 to 47 in male and 33 to 45 in female subjects. Within each country, the range was 5–8 across the five age ranges in all cases except Netherlands females who had a range of 11 pegs across the four age ranges they spanned. A country by country comparison suggests a consistent decrement in the number of pegs turned between subjects in the three younger age ranges and subjects in the 56–65 age range. The Santa Ana is the one NCTB test on which performance by subjects from Nicaragua was close (within 1 SD) to that seen in

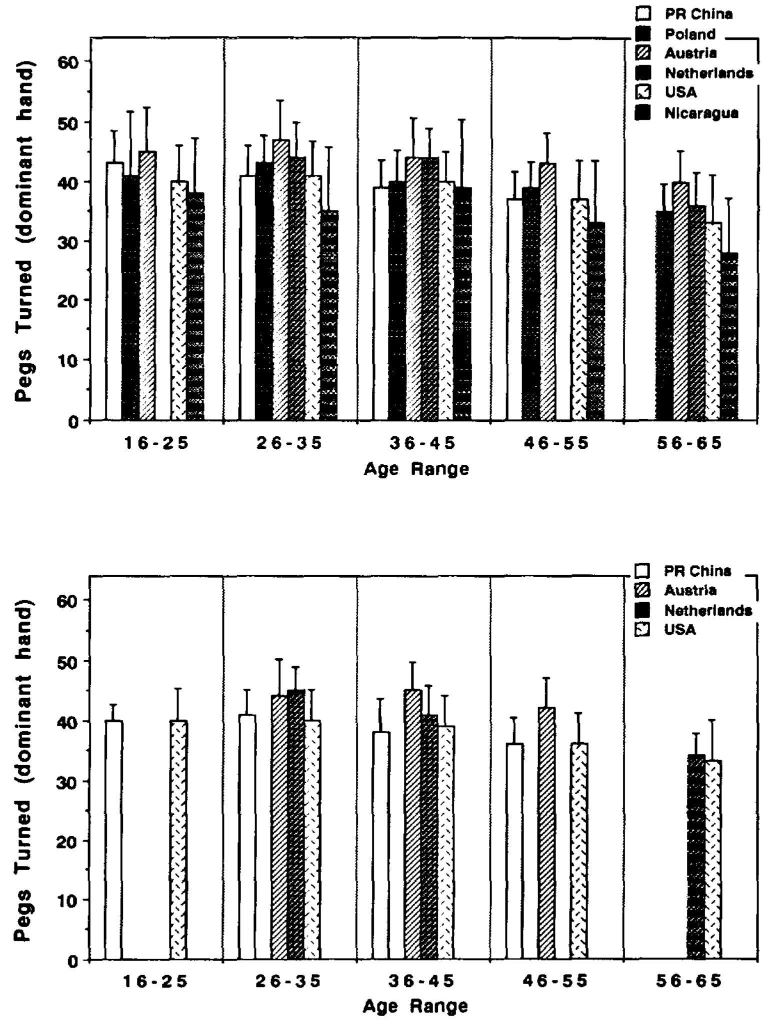
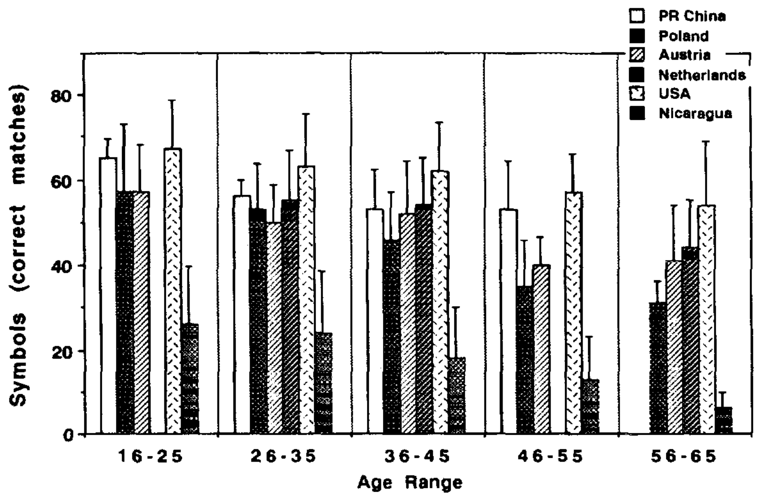


FIG. 3. Mean (±SD) number of pegs turned by the dominant hand in the Santa Ana dexterity test in age ranges between 16 and 65 for males (top) and females (bottom).

other countries, although Nicaraguan subjects turned the smallest number of pegs in every age range except 36–45 where they tied for the smallest number with subjects from China.

Figure 4 presents the mean number of correct symbols drawn in 90 sec in the Digit Symbol test. In Nicaragua (male subjects), the mean number of symbols ranged between 6 and 26. In all countries except Nicaragua, the mean number of symbols drawn by males ranged from 31 to 65; females drew between 44 and 72 symbols. Within-country data follow a consistent decreasing trend from younger to older age ranges and a trend to higher scores in females than in males.

Figure 5 presents the number of correctly recognized geometric figures (of 10 possible) in the Benton Visual Retention test. The mean number of figures remembered was between 7.6 and 9.4 for both male and female subjects. The sole exception is Nicaragua (male subjects) where the range was from just under 4 to



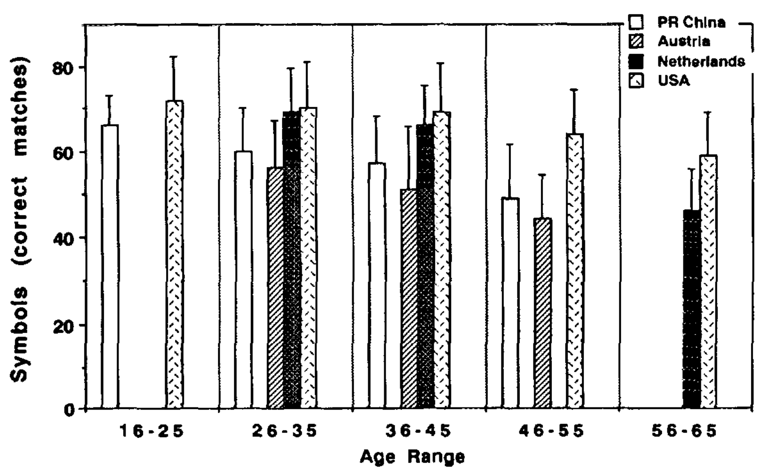


FIG. 4. Mean (±SD) number of symbols (correct matches) drawn in the Digit Symbol test in age ranges between 16 and 65 for males (top) and females (bottom).

just over 4.5. If Nicaragua is excluded, within-country data are highly consistent (similar), as are between-country data.

From Pursuit Aiming II, Fig. 6 records the number of dots placed in circles in two 60-sec trials. Male subjects from all countries except Nicaragua placed a mean number of between 93 and 240 dots, while the range in (male) Nicaraguan subjects was between 18 and 25 dots. Within-country data are relatively more consistent than between-country data, with a tendency to slight declines in the 46–55 and/or 56–65 age group. Females follow the same pattern (range from 108 to 256).

The larger number of countries represented in the 26–35, 36–45, and 46–55 age ranges, shown in figures 7–10, reinforces the evidence of between-country consistency or similarity seen in Figs. 1–6. Between-country data are highly consistent (similar) for the Simple Reaction Time test and, excepting Nicaragua, the

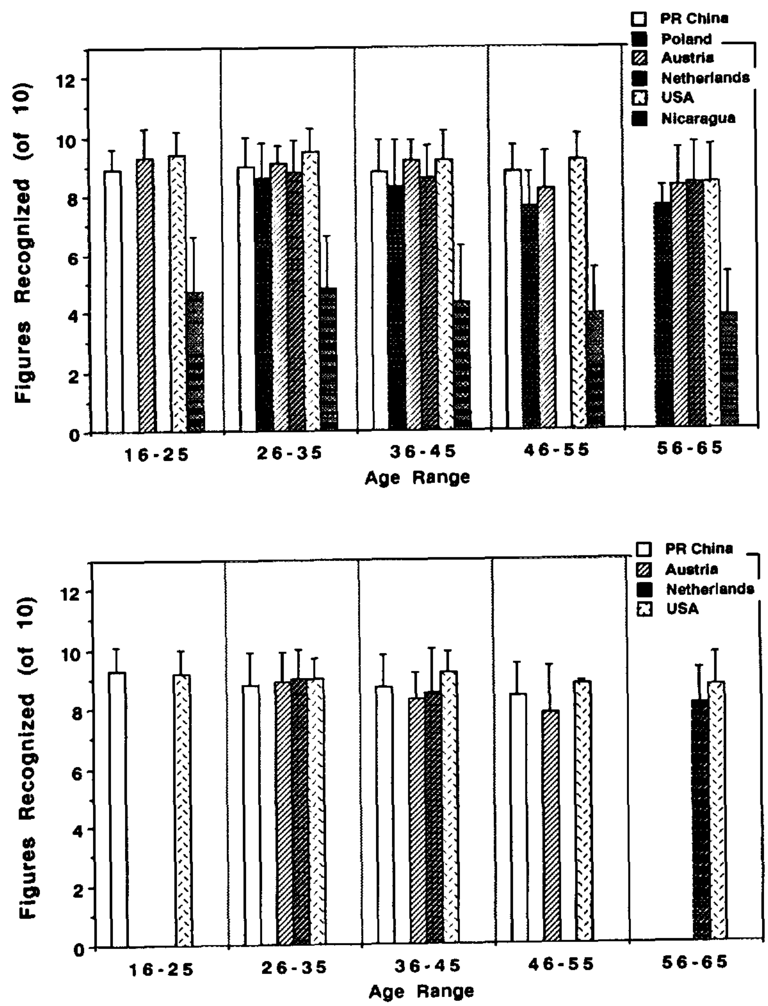


FIG. 5. Mean (±SD) number of figures recognized (of 10) in the Benton Visual Retention test in age ranges between 16 and 65 for males (top) and females (bottom).

Benton test, whereas the data are less consistent for, in declining order, the Santa Ana, Digit Symbol, Digit Span, and Aiming tests.

DISCUSSION

The Cross-Cultural Assessment of the NCTB exceeded by two the goal of collecting data in eight countries, although the distribution did not achieve the degree of cultural diversity originally sought. Data presented here represent performance of over 2300 participants (1303 males and 1028 females) from European (involving, at a minimum, slavic, grecolatin, and germanic cultural groups), North American (predominantly Western European), Central American, and Asian populations. Collection of NCTB data from other populations (e.g., Venezuela and India) also appears achievable in the near future.

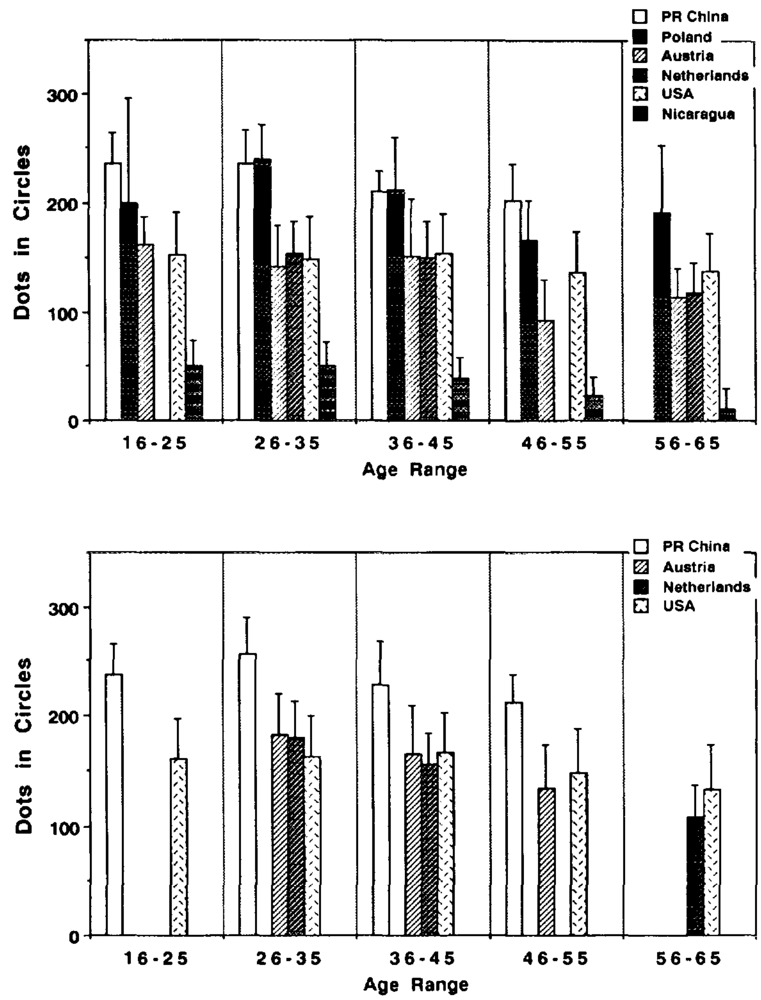


FIG. 6. Mean (±SD) number of dots placed in circles in the Pursuit Aiming II test in age ranges between 16 and 65 for males (top) and females (bottom).

Between-Country Consistency

The data collected on the Simple Reaction Time test demonstrate consistency or similarity in absolute scores across all countries studied. With the exception of Nicaragua, data from the Benton test also demonstrate consistency between countries. Santa Ana, Digit Span, Digit Symbol, and Aiming data demonstrate more diversity between countries, although data on these tests appear relatively consistent within each country when viewed across age ranges (Figs. 1–6).

The data from Nicaragua stand in stark contrast to results of the NCTB in other countries. The explanation for this performance difference does not appear to relate to procedural problems. The test administrators in Nicaragua were trained by Dr. R. Amador, who was trained in NCTB administration by Dr. H. Hänninen. Dr. Amador conducted the study and served as one of the six Nicaraguan test administrators. In addition, Dr. M. Keifer, who received NCTB training from Dr. K. Anger, collaborated on project testing. Drs. Amador and Keifer judged the test

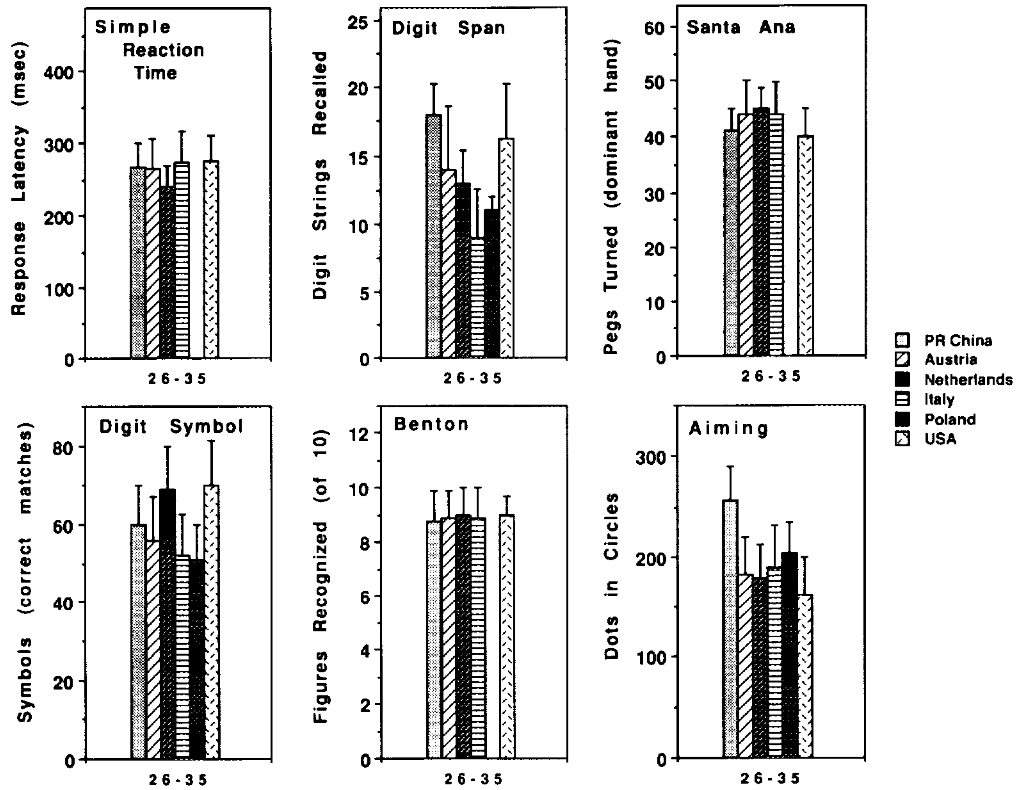


FIG. 7. Mean (±SD) on NCTB tests for female subjects ages 26–35 (including countries which only reported results in limited age ranges).

administrators to be well-trained and motivated. Most subjects had some difficulty with all tests, but test administrators and observers (Amador, Keifer) were convinced that subjects understood the instructions, responded positively to the testers and test setting, were motivated to perform the tests (many appeared competitive, inquiring how they did compared to others), and supported the study's goals (both unions and management supported the study in open meetings in which potential subjects asked many questions about who would benefit and who was running the study). Mean performance by Nicaraguan subjects on the Santa Ana, though slightly lower than that in other countries, was within one standard deviation of mean scores in other countries (Fig. 3). This suggests that the Nicaraguan subjects had a reasonable level of motor competence, motivation, understanding of instructions, and comfort with the test conditions.

Evidence to support this conclusion was sought from the findings of the Simple Reaction Time test which, as noted above, were excluded from the results because the Operational Guide instructions had failed in Nicaragua to produce the appropriate response behavior and were thus altered. The principals changed the procedure to require the subjects to place their fingers to the side of the response button and, when they saw the light, to "press the trigger." While the Nicaraguan subjects had little education, many have had recent military experience. Their mean response latency varied between 385 and 450 msec, or roughly twice that of subjects in the other countries (Fig. 1) who had a far shorter distance to respond

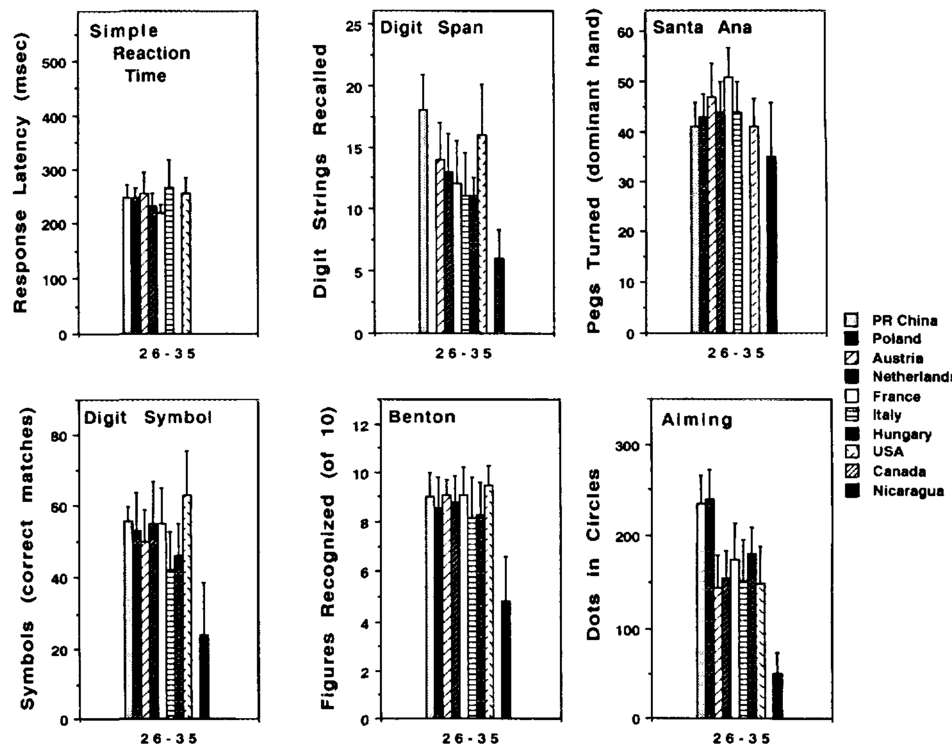


FIG. 8. Mean (±SD) on NCTB tests for male subjects ages 26–35 (including countries which only reported results in limited age ranges).

(viz. 0 mm, as they were touching the button). To further investigate this finding, 32 faculty members (mean 17 years education) of the University of Nicaragua in Leon were given this test using the correct instructions; their mean reaction time latency was 255 msec for subjects aged 16–25 and 258 msec for subjects 26–35; this is in the range found in all other countries (Fig. 1). When administered to the same 32 faculty members using the modified (incorrect) instructions, the mean reaction time latency was 355 and 334 msec for the 16–25 and 26–35 age ranges, respectively. Finally, 45 of the original control subjects were retested on this the Reaction Time test using the correct instructions; the mean time latency was 330–402 msec in the five age ranges, suggesting that performance was slower in Nicaragua than in other countries but within 1–2 SD of performance in other countries. The performance of the faculty demonstrates that a highly educated group of Nicaraguans can perform this test very well. This suggests a hypothesis related to education, explored below.

Aside from the results on the Santa Ana, mean scores by Nicaraguan subjects on the remaining NCTB performance tests are roughly half the level achieved in other countries. The most straightforward explanation for this finding is in factors inherent to the subjects tested. Nicaragua is the one country in the CCA where a predominantly rural population with almost no formal education (approximately 3 years) was tested. Specifically, 32% of the Nicaraguan subjects had no formal education, 42% had 1–3 years, and 26% had 4–19 years of education. Most Nicaraguan subjects responded to a structured question set that they did not write or

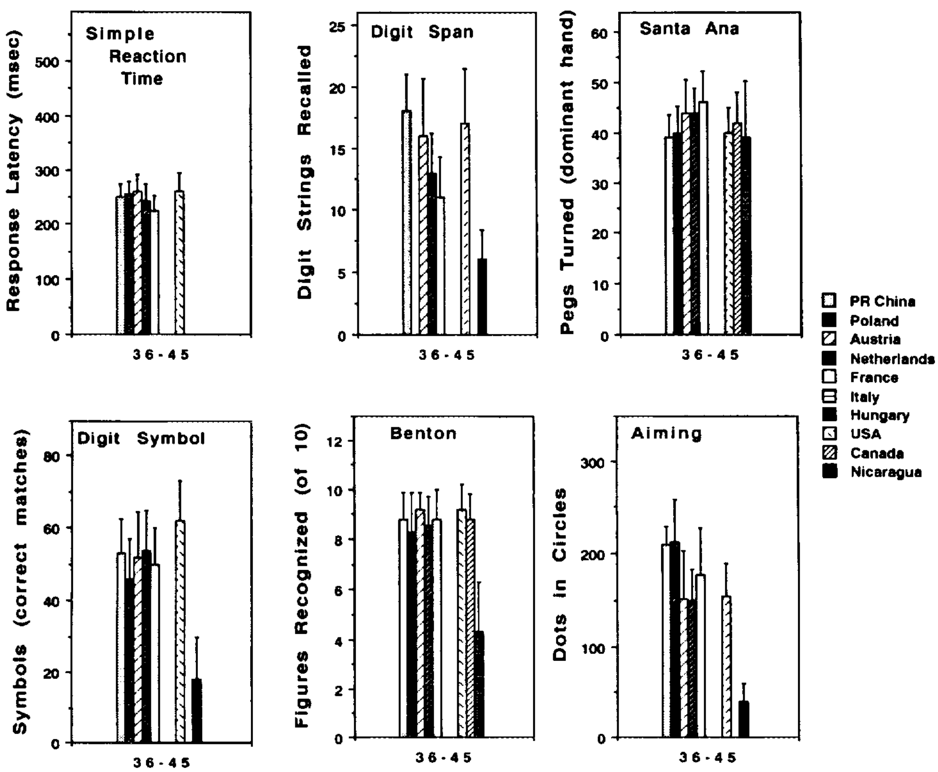


FIG. 9. Mean (±SD) on NCTB tests for male subjects ages 36–45 (including countries which only reported results in limited age ranges).

read in their daily lives. Thus, at least 74% of these subjects can only be characterized as marginally literate or illiterate.

Drs. Amador and Keifer provide enlightening comments on Nicaraguan subject performance relating to the lack of formal education and an associated unfamiliarity with performance tests such as those in the NCTB. On the Benton, the lack of experience with geometric figures is hypothesized as the key problem. Inexperience with writing was a significant handicap in the Aiming and Digit Symbol tests; test administrators noted that about one-third of the subjects appeared unused to having pencils in their hands. The Aiming test may also have been affected by uncorrected vision problems, especially in older subjects. The Digit Symbol test amounted to a symbol–symbol test (and perhaps was further affected by uncorrected vision deficits) in these marginally literate subjects, and the Digit Span may also have been affected by inexperience with numbers.

To assess the possibility that education was a major factor in the Nicaraguan results, Pearson r and Kendall's τ correlations were run between years of education and scores on each test. The nonparametric Kendall's τ correlations were included because the large number of subjects with 0–3 years education produced nonnormal distributions. The results for each test were: Digit Symbol (r = 0.69; τ = 0.64); Digit Span (r = 0.44; τ = 0.52), Benton (r = 0.33; τ = 0.42); Aiming (r = 0.58; τ = 0.58). As would be expected, the lowest correlation was on the Santa Ana (r = 0.29; τ = 0.31) where performance was close to that seen in other countries and the Simple Reaction Time test (r = −0.12; τ = 0.11).

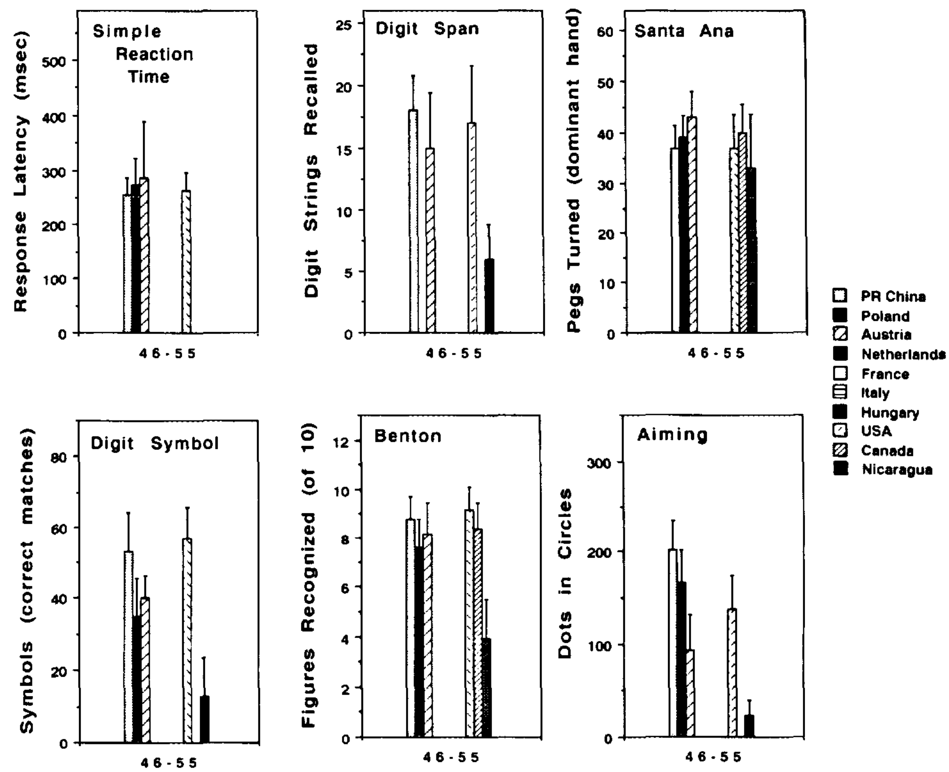


FIG. 10. Mean (±SD) on NCTB tests for male subjects in ages 46–55 (including countries which only reported results in limited age ranges).

These correlations are consistent with other reports. Neuropsychological test performance is positively associated with education, including tasks (e.g., memory) independent of academic performance (Lezak, 1983). There is a 0.44 and 0.42 correlation between years of education and performance on the Digit Symbol and Digit Span, Wechsler Adult Intelligence Scale (WAIS) subtests, respectively. Scores on WAIS "performance" tests, including the Digit Symbol test, are roughly 10–20% lower in subjects with a range of 0–11 years of education than in subjects with 12–15 years of education (Albert and Heaton, 1988). Similar though smaller performance differences related to education have been reported for cognitive tests such as those in the Halstead-Reitan (Leckliter and Matarazzo, 1989) and the WAIS Digit Span (Heaton *et al.*, 1987) and simple motor tests such as tapping (Warner *et al.*, 1987) and grooved pegboard tests (Bornstein, 1985).

While there are large differences in education between Nicaraguan and other CCA subjects, the Nicaraguan population also represents a unique cultural group (American Indian–Spanish) in this study. Cultural differences provide a reasonable competing hypothesis (to education) as the cause of the performance differences between Nicaragua and the other countries. However, the evidence on the educational factor in cultural differences in cognitive performance (D'Andrade, 1989), evidence in many populations of a correlation between performance and education on some NCTB tests (noted above), and the high correlation between years of education and NCTB performance reported in this study reinforce the proposal that educational factors provide the more supportable hypothesis to explain the poorer performance of Nicaraguan subjects.

The huge performance differences between subjects from Nicaragua and those from other CCA countries raises doubts about the NCTB's hoped-for universal utility. The high correlation between years of education and several NCTB tests indicates that this variable must be taken into account (through subject selection criteria or data analysis) in the analysis of NCTB data in any study. However, more research is clearly needed on education as a factor in NCTB performance. A different battery may be needed to provide an adequate baseline (i.e., sufficiently high performance and a sufficiently small standard deviation relative to the mean) to detect behavioral effects (performance decrements) of neurotoxic chemicals in poorly educated or poorly literate populations. *Data from other Latin American countries and from sparsely educated subjects in other countries and in different socioeconomic strata are needed to address these findings.*

Other Subject Factors (Age, Sex)

Data on two NCTB tests demonstrate slight (Santa Ana, Digit Span) or marked (Digit Symbol) age-related performance declines, particularly when viewed in within-country comparisons. This is consistent with other research. WAIS verbal tests remain stable into later decades, while performance tests show a characteristic age-related decline (Nolan *et al.*, 1986). The one WAIS verbal test in the NCTB is the Digit Span and the only performance test in the NCTB is the Digit Symbol (Murphy and Davidshofer, 1991). CCA data are thus consistent with these findings; declines in the 56–65 age range are more apparent in the Digit Symbol (Fig. 4) than in the Digit Span (Fig. 2). Further, age-related memory losses are found on free recall (e.g., Digit Span) but not on recognition tests (e.g., Benton) (Shimamura, 1990; Botwinick, 1981). This is consistent with CCA data in that performance on the Benton is virtually the same in all age ranges (Fig. 5), while there appears to be a slight decline in the 56–65 age range on the Digit Span (Fig. 2). With regard to motor tests, mean scores on the Santa Ana are lowest in the 56–65 age range, when compared to subjects of younger ages within each country, but there are no such differences on the Simple Reaction Time test.

Age-related declines in performance are reported on tests of memory (Nolan *et al.*, 1986), a variety of visual discrimination tasks (Hochanadel and Kaplan, 1984), and motor tests, including Simple Reaction Time (Botwinick, 1967, 1984; Salthouse, 1985; Nolan *et al.*, 1986) and grooved pegboard (Bornstein, 1985). While age-related performance declines do occur in the fifth decade, they are more typical of the sixth decade or beyond (Albert and Heaton, 1988; Nolan *et al.*, 1986), and age-correlated differences are exaggerated by cross-sectional research such as the present study, due presumably to improvements in early health care, nutrition, and education (Albert, 1988). Since the age-related declines seen in this study are minimal, the NCTB would appear relatively resistant to group differences in age, especially below age 55.

There are performance differences between men and women on many behavioral tests. For example, women perform significantly better on the Digit Symbol test than do men (Albert and Heaton, 1988), and the reverse is the case for the grooved pegboard test (Bornstein, 1985). Similar trends are seen in the NCTB's Digit Symbol data in some countries (e.g., United States) and in the Santa Ana (pegboard) test, but the differences are very small and not completely consistent in the latter case. Generally, however, CCA test results of women subjects are remarkably similar to test results of men subjects on most NCTB tests when comparing data country by country (Figs. 1–6).

In summary, trends in the CCA data are generally consistent with findings in the literature with regard to the variables of education, age, and sex. Sex appears to have a smaller impact on NCTB performance, and the changes in NCTB performance over age ranges of 16–65 (noted within each country) are also relatively slight. This suggests that, in worksite research, small differences between the mean ages of referent and exposed groups with only relatively equal distributions of men and women would not seriously jeopardize conclusions from data on most NCTB tests. Rather, concern in a given study should be focused on education and perhaps other potential variables if sex and age differences between groups are not large, and especially when subject ages are below 55 years.

Technical Factors

For those tests where there are substantial differences between countries (e.g., Digit Span, Aiming), the differences could result from a variety of factors, including cultural, educational, nutritional, prenatal, health status, genetic, motivational, or socioeconomic. Performance variability on all tests could also result from a variety of technical factors such as instructions, test administration consistency, and test grading. Relevant technical factors that could have produced between-country variability and within-country consistency include the following: (a) the orally presented Digit Span test is particularly susceptible to vagaries of administration; (b) the relation between subject height/arm length and table height in the Santa Ana dexterity test may have been variably optimal across countries; (c) the Pursuit Aiming test is susceptible to substantial grading criterion differences such that all tests in a given study should be graded by one person; (d) substitute equipment in the Simple Reaction Time test was used after the original equipment became prohibitively expensive; button travel distance, button resistance, and other factors could have differed, despite heroic efforts to assure that alternative devices retained the same timing characteristics as the original source; and, (e) the POMS test is so highly dependent on the existence of precisely equivalent terminology in diverse languages that translation limitations must be expected to be unresolvable (e.g., D'Andrade, 1989).

CONCLUSIONS

The data presented here serve an illustrative purpose. The initial goal of the Cross-Cultural Assessment was to determine if a wide variety of populations could perform in a roughly similar manner on the NCTB tests, thus satisfying a loosely defined "feasibility" criterion. While there are clearly differences in mean performance between countries on many tests (e.g., Aiming, Digit Symbol, Digit Span), within-country performance across age ranges is remarkably consistent in the six countries in which such data have been obtained by the CCA. Feasibility is thus clearly demonstrated in a range of cultural groups, as long as within-country comparisons are planned. The NCTB can therefore be used widely, perhaps universally, in male and female working age populations with 8–12 years of formal education, as long as the education variable is equated among groups or included in the data analysis. However, the degree to which subjects with less than 8 years education can provide baseline performance on most NCTB tests which will allow detection of neurotoxic effects in exposed populations remains an open question.

It was recognized by the original expert group that data from this Assessment in unexposed workers would *not substitute for control or reference groups in future studies* because factors such as living conditions, health status, education, socioeconomic status, genetics, motivation, and culture (e.g., deemphasis on individual achievement) can be dominant variables. The baseline data presented here tend to support that contention with regard to education and perhaps correlated socioeconomic background (Nicaraguan subjects were described as rural peasants while all other subject groups came from towns or cities and were typically not on the bottom socioeconomic rung). However, the data in this article can serve as a *guide to the range within which control data would be expected to fall.* More useful is the relative lack of performance changes across age ranges which suggests that conclusions based on most NCTB tests can be drawn in cross-sectional studies when group ages are relatively similar (especially between the ages of 16 and 45). Similarly, minor differences in group gender composition in cross-sectional studies with the NCTB would not be expected to affect conclusions on most tests.

The Cross-Cultural Assessment was designed to develop baseline data across the working age span in a minimum of eight countries by testing working populations unexposed to chemical substances where they work. The minimum number of countries was met, although the cultural and socioeconomic dispersion is insufficient at present. Further, the feasibility of using the NCTB in poorly educated populations remains an unresolved question that urgently needs an answer before the NCTB can be recommended for use in such populations. Thus, additional data are being sought to further expand the CCA in breadth (across countries) and depth (multiple sites within countries) in order to increase the generality of the baseline data and to answer questions related to feasibility of use in poorly educated populations. Continuation of this effort has been undertaken directly by Drs. Anger and Cassitto and their host institutions (Oregon Health Sciences University and Milan's Institute of Occupational Health, respectively), with tacit but not administrative support of WHO.

ACKNOWLEDGMENTS

The many principal investigators are responsible primarily for data and interpretive comments they provided to the article about their country; the senior author takes responsibility for the conclusions (with which other authors may or may not concur), although no disagreements have come to light. U.S. data were collected when W. K. Anger was employed at NIOSH (United States). Dr. Helena Hänninen participated importantly in the NCTB cross-cultural assessment, and several scientists who are not named as authors collected data presented in this manuscript, including (not ordered) A. Jóni for Hungary data; C. Fanelli, E. Fattorini, and D. Camerino for Italy data; Y. Miranda (National Autonomous University of Nicaragua), I. Lundberg (Karolinska Hospital, Sweden), R. McConnell (Mt. Sinai School of Medicine, United States), K. Claypoole, W. Daniels, L. Rosenstock (University of Washington) for Nicaragua data; Z. Chen (Shanghai Medical University, People's Republic of China) for Shanghai data; F. He and Z. Zhou (Institute of Occupational Medicine, Beijing, People's Republic of China) for Beijing information; B. Bazylewicz-Walczak for Poland data; and S. Brightwell, B. J. Taylor, and S. Robertson (NIOSH) for U.S. data. Appreciation is extended to W. Carlton for word processing.

REFERENCES

Albert, M. S. (1988). General issues in geriatric neuropsychology. *In* "Geriatric Neuropsychology" (M. S. Albert and M. B. Moss, Eds.), pp. 3–10. The Guilford Press, New York.

Albert, M. S., and Heaton, R. K. (1988). Intelligence testing. *In* "Geriatric Neuropsychology" (M. S. Albert and M. B. Moss, Eds.), pp. 13–32. The Guilford Press, New York.

Anger, W. K. (1990). Worksite behavioral research: Results, sensitive methods, test batteries and the transition from laboratory data to human health. *Neurotoxicology* **11,** 629–720.

Baker, E. L., Letz, R., and Fidler, A. (1985). A computer-administered neurobehavioral evaluation system for occupational and environmental epidemiology: Rationale, methodology, and pilot study results. *J. Occup. Med.* **27,** 206–212.

Bernard, L. C. (1989). Halstead-Reitan neuropsychological test performance of black, hispanic, and white young adult males from poor academic backgrounds. *Arch. Clin. Neuropsychol.* **4,** 267–274.

Bornstein, R. A. (1985). Normative data on selected neuropsychological measures from a nonclinical sample. *J. Clin. Psychol.* **41,** 651–659.

Botwinick, J. (1967). "Cognitive Processes in Maturity and Old Age." Springer, New York.

Botwinick, J. (1984). "Aging and Behavior." Springer, New York.

Botwinick, J. (1981). Neuropsychology of aging. *In* "Handbook of Clinical Neuropsychology" (S. B. Filskov and T. J. Boll, Eds.), Vol. 1, pp. 135–171. Wiley, New York.

Cassitto, M. G., Camerino, D., Hänninen, H., and Anger, W. K. (1990). International collaboration to evaluate the WHO Neurobehavioral Core Test Battery. *In* "Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health" (B. L. Johnson, W. K. Anger, A. Durao, and C. Xintaras, Eds.), pp. 203–223. Lewis, Chelsea, MI.

D'Andrade, R. G. (1989). Cultural cognition. *In* "Foundations of Cognitive Science" (M. I. Posner, Ed.), pp. 795–830. The Massachusetts Institute of Technology Press, Cambridge, MA.

Dews, P. (1975). An overview of behavioral toxicology. *In* "Behavioral Toxicology" (B. Weiss and V. Laties, Eds.), pp. 439–445. Plenum Press, New York.

Hänninen, H. (1966). Psychological tests in the diagnosis of carbon disulfide poisoning. *Work Environ. Health* **2,** 16–20.

Hänninen, H., and Lindström, K. [1979, 1989 (revised)]. "Neurobehavioral Test Battery of the Institute of Occupational Health." Helsinki, Institute of Occupational Health.

Heaton, R. K., Grant, I., and Mathews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. *In* "Neuropsychological Assessment of Neuropsychiatric Disorders" (I. Grant and K. M. Adams, Eds.), pp. 100–120. Oxford Univ. Press, New York.

Hochanadel, G., and Kaplan, E. (1984). Neuropsychology of normal aging. *In* "Clinical Neurology of Aging" (M. L. Albert, Ed.), pp. 231–244. Oxford Univ. Press, New York.

Johnson, B. L., and Anger, W. K. (1983). Behavioral Toxicology. *In* "Environmental and Occupational Medicine" (W. R. Rom, Ed.), pp. 329–350. Little, Brown, Boston.

Johnson, B. L., Baker, E. L., El Batawi, M., Gilioli, R., Hänninen, H., Seppäläinen A. M., and Xintaras, C., Eds. (1987). "Prevention of Neurotoxic Illness in Working Populations." Wiley, New York.

Kaufman, A. S. (1990). "Assessing Adolescent and Adult Intelligence." Allyn & Bacon, Needham Heights, MA.

Leckliter, I. N., and Matarazzo, J. D. (1989). The influence of age, education, IQ, gender, and alcohol abuse on Halstead-Reitan neuropsychological test battery performance. *J. Clin. Psychol.* **45,** 484–512.

Letz, R. (1990). The neurobehavioral evaluation system: An international effort. *In* "Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health" (B. L. Johnson, W. K. Anger, A. Durao, and C. Xintaras, Eds.), pp. 189–201. Lewis, Chelsea, MI.

Lezak, M. D. (1983). "Neuropsychological Assessment." Oxford University Press, New York.

Liang, Y-x, Chen, Z-q, Sun, R-k, Fang, Y-f, and Du, J-h. (1990). Application of the WHO Neurobehavioral Core Test Battery and other neurobehavioral screening methods. *In* "Advances in Neurobehavioral Toxicology: Applications in Environmental and Occupational Health" (B. L. Johnson, W. K. Anger, A. Durao, and C. Xintaras, Eds.), pp. 225–243. Lewis, Chelsea, MI.

Matarazzo, J. D., and Herman, D. O. (1984). Relationship of education and IQ in the WAIS-R standardization sample. *J. Consult. Clin. Psychol.* **52,** 631–634.

Murphy, K. R., and Davidshofer, C. O. (1988). "Psychological Testing: Principles and Applications." Prentice-Hall, Englewood Cliffs, NJ.

Nolan, B. H., Swihart, A. A., and Pirozzolo, F. J. (1986). The neuropsychology of normal aging and dementia: An introduction. *In* "The Neuropsychology Handbook: Behavioral and Clinical Perspectives" (D. Wedding, A. M. Horton, Jr., and J. Webster, Eds.), pp. 410–440. Springer, New York.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. *In* "Handbook of the Psychology of Aging" (J. E. Birren and J. W. Schaie, Eds.), 2nd ed., pp. 400–426.

Shimamura, A. P. (1990). Aging and memory disorders: A neuropsychological analysis. *In* "Cognitive and Behavioral Performance Factors in Atypical Aging" (M. L. Howe, M. J. Stones, and C. J. Brainerd, Eds.), pp. 37–65. Springer, New York.

Warner, M. H., Ernst, J., and Townes, B. D. (1987). Relationships between IQ and neuropsychological measures in neuropsychiatric populations: Within-laboratory and cross-cultural replications using WAIS and WAIS-R. *J. Clin. Exp. Neuropsychol.* **9**, 545–562.

World Health Organization. "Operational Guide for the NCTB" (Not published). Available from W. Kent Anger, Ph.D., Center for Research on Occupational and Environmental Toxicology, Oregon Health Sciences University, Portland, Oregon 97201 or Maria Cassitto; University of Milan; Milan, Italy.