Lateralizing Brain Damage with the Luria-Nebraska

Neuropsychological Battery: Diagnostic

Effectiveness as Compared to the Halstead-Reitan

Neuropsychological Test Battery

by

Michael Stambrook

A thesis
presented to the University of Manitoba
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
Psychology

Winnipeg, Manitoba

LATERALIZING BRAIN DAMAGE WITH THE LURIA-NEBRASKA

NEUROPSYCHOLOGICAL BATTERY: DIAGNOSTIC

EFFECTIVENESS AS COMPARED TO THE HALSTEAD-REITAN

NEUROPSYCHOLOGICAL TEST BATTERY

BY

MICHAEL STAMBROOK

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY
✓© 1985

ACKNOWLEDGMENTS

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Michael Stambrook

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Michael Stambrook

LIST OF TABLES

Abstract

Methodological and conceptual concerns have been raised regarding the test construction and validation procedures used in the development of the Luria-Nebraska Neuropsychological Battery (LNNB). The present investigation examined a major neuropsychological validation question, the ability of the LNNB to lateralize brain damage since substantial statistical and methodological issues have been raised regarding the initial LNNB lateralization validation study. A comparison was then made of the ability of the LNNB and the Halstead-Reitan Neuropsychological Test Battery to lateralize brain damage. Both neuropsychological batteries were administered to 30 predominantly left- or right-hemisphere-damaged subjects (15 in each group) with discriminant analysis classification procedures demonstrating similar, above chance, accuracy in lateralizing cerebral dysfunction. Cross-validation of objective clinical rules designed to aid in LNNB test interpretation resulted in classification of brain damage and lateralization at levels below reported values from the test developer's laboratory. The limited role such simplistic rules have in assessment is discussed, as is the need to evaluate dimensions of test usefulness other than

those related to gross diagnostic decisions (e.g., presence
or laterality of brain damage) in determining the instrument
of choice for clinical neuropsychological practice. It is
argued that the demonstration of the ability to make gross
diagnostic decisions provides necessary but minimal evidence
for the usefulness of a neuropsychological instrument. More
research is recommended to fully define the limits of the
clinical utility of the LNNB.

Lateralizing Brain Damage with the Luria-Nebraska

Neuropsychological Battery: Diagnostic

Effectiveness as Compared to the Halstead-Reitan

Neuropsychological Test Battery

The recent development, promoting, and marketing of the
Luria-Nebraska Neuropsychological Battery (LNNB) as a
clinical instrument has precipitated considerable
acrimonious discussion in the literature (e.g., Adams,
1980a, 1980b, 1984; Golden, 1980; Stambrook, 1983) regarding
the methodological and conceptual rigor of the test
construction and validation procedures used by the battery's
principal designer, Charles Golden, and his associates
(Golden, Hammeke, & Purisch, 1980). Numerous substantial
statistical and methodological problems have been identified
with the LNNB research base which should cause potential
clinical users some concern (see Stambrook, 1983 or, for an
expanded analysis and critique of the battery's theoretical
antecedents, development, standardization, reliability, and
validity, see Appendix 1). In spite of this, the battery
continues to be heavily advertised for purchase for clinical
use (Western Psychological Services, 1984),for training at
American Psychological Association sponsored continuing
education seminars (Neuropsychological Associates of

California, 1985), and more recently, is profiting from the introduction of computerized scoring and report writing (Precision People Inc., 1985).

Despite detractors' critiques (Adams, 1980a, 1980b; Crosson & Warren, 1982; Delis & Kaplin, 1982; Spiers, 1981, 1982; Appendix 1), the impressive array of data that Golden and his team have compiled and presented -- which they suggest unequivocally demonstrates the efficacy of the LNNB in detecting the presence, lateralization, and localization of brain damage (for reviews see Golden, 1981a, 1981b; Appendix 1) -- seems to be one reason why the battery is enjoying increasing popularity (Goldberg & McNamara, 1984; Hartlage & Telzrow, 1980; Noonberg & Page, 1982). Reviews of the literature on the battery to date, however, lead to the conclusion that the research base that unambiguously demonstrates the effectiveness of the instrument is not large enough to justify placing confidence in the clinical use of the battery,(e.g., Adams, 1980a, 1980b; Spiers, 1981, 1982; Appendix 1). Nevertheless, it is clear that the efficacy of the LNNB in clinical situations rests not on the developer's and publisher's claims, or on detractor's critiques, but on carefully planned and well executed research. Further, if as is stated in the recently published test manual (Golden, Hammeke, & Purisch, 1980), "the Luria-Nebraska can replace the much more extensive batteries commonly used in American neuropsychology [e.g.

the Halstead-Reitan Neuropsychological Test Battery] that may take two to three times as long to administer" (p.13), it is essential to empirically demonstrate that data relevant to brain-behavior assessment is not lost for the sake of brevity.

Despite documented shortcomings (cf., Adams, 1980a, 1980b; Stambrook, 1983; see also Appendix 1), the LNNB has been shown to be able to detect the presence of brain damage (e.g., Golden, Hammeke, & Purisch, 1978; Kane, Parsons, & Goldstein, 1985). The situation is, however, less clear in regards to the next major issue in the validation of a neuropsychological instrument -- assessing the ability of the LNNB to lateralize cerebral dysfunction -- and this was the primary focus of the present study. In overview, the goals for the present project were to:

1. Ameliorate the interpretive problems that are present in the Osmon, Golden, Purisch, Hammeke, and Blume (1979) investigation of the power of the LNNB to lateralize cerebral dysfunction as documented on independent neurological grounds using discriminative function classification procedures.

2. Provide a 'head to head' test of the comparative diagnostic accuracy of the LNNB and the Halstead-Reitan Neuropsychological Test Battery (HRNTB) in lateralizing cerebral dysfunction in groups of right- and left-hemisphere damaged subjects using

discriminative function classification procedures.
Of interest here as well was a comparison of the
administration times for each battery.

3. Of ancillary interest, this project was designed to
provide a cross-validation of a set of objective
clinical rules that have been developed to diagnose
and lateralize brain damage using the LNNB (Golden,
Moses, Graber, & Berg, 1982; McKay & Golden, 1979a,
1979b).

In the next three sections, each of these will be examined
in turn.

The Osmon et al. (1979) Investigation

The ability of the LNNB to statistically discriminate
lateralized and diffusely brain-damaged subjects has been
investigated by Osmon et al. (1979). Twenty subjects were
assigned to each of three groups (left, right, diffuse)
depending on the locus of brain damage as confirmed by one
or more of the following methods: neurological exam,
arteriogram, electroencephalogram (EEG), computerized axial
tomography (CAT scan), pneumoencephalogram, skull x-rays,
and surgery. While the groups were found not to differ in
age or education, the homogeneity of the groups with respect
to other potentially confounding variables such as sex,
chronicity, and process (Parsons & Prigatano, 1978) was not
reported. Analyses of variance on each of the 14 LNNB

summary scales revealed that the groups differed significantly only on the Left Hemisphere scale with the left-hemisphere group more impaired than the right-hemisphere or diffuse groups. The possibility that one significant analysis of variance (out of 14) could have arisen from the operation of random or chance factors is not considered by the researchers. Osmon et al. (1979) also report that a discriminant analysis classification procedure performed on the 14 summary scales was able to correctly classify 59 out of 60 subjects (98%).

There are, however, substantive problems with the Osmon et al. (1979) analysis that render unambiguous interpretation of the findings difficult. Aside from the shrinkage in classification accuracy that would be expected from the use of a discriminative analysis with a subject (per group) to variable ratio as small as was present (20/14) (Fletcher, Rice, & Ray 1978), there is a lack of appreciation of the statistical and conceptual improbability of achieving near perfect discriminative classificatory accuracy when the univariate tests indicate the essential equality among groups (Adams, 1980a; Appendix 1). This can be demonstrated by the fact that 11 of the 14 univariate F statistics had values less than 1.0, and that the average F value over the 14 tests was .83. This interpretative difficulty is compounded by the absence of a reported Multivariate F statistic, and by the absence of a reported

evaluation of whether all 14 dependent measures added

significantly to the between group separation or, whether

they capitalize on chance variation in the sample (Appendix

1). Fletcher et al. (1978) have presented data on this

issue and have demonstrated that in, for example, a 2-group

design with small subject to variable ratios, discriminant

analysis classification procedures on random numbers can

lead to hit rates that range from 71% to 90% (# subjects/#

variables; 1:1 to 2:1). This clearly causes concern and

raises suspicions that the groups really did not differ and

seriously calls into question the reported high level of

accuracy in the classification of the subjects based on the

discriminant function.

Comparative Effectiveness of the LNNB and HRNTB in Lateralizing Brain Damage

The HRNTB is considered by many to be the preeminent

standardized neuropsychological test battery in clinical use

in North America (Craig, 1979; Golden & Kuperman, 1980;

Hartlage & Telzrow, 1980; Noonberg & Page, 1982). The

battery has a long and extensive history of research to its

credit that has convincingly demonstrated the procedures'

utility in diagnosing the presence, lateralization,

localization, and etiology of brain damage (for reviews see

Boll, 1981; Goldstein, 1974; Hevern, 1980; Klove, 1974;

Lezak, 1976; Reitan, 1962, 1964, 1966a, 1966b, 1975;

Russell, Neuringer, & Goldstein, 1970).

While the degree of accuracy lateralizing brain damage using the LNNB (98%) is somewhat higher than that obtained using discriminant analyses of HRNTB data (e.g., 95%, Golden, 1977; 88% Goldstein & Shelley, 1973; 93%, Wheeler, Burke, & Reitan, 1963; 88%, Wheeler & Reitan, 1963), such comparisions are of limited value because of many known and unknown differences (e.g., age, sex, chronicity, process, extent, etc.) in the composition of the groups under investigation. Certainly, at a minimum, the various groups used by different researchers differ in the stringency with which they are formed because subjects are consistently described as predominantly left- or right-hemisphere damaged because of the complex and not well understood distance effects of intracranial lesions (Lishman, 1978). Hence, since different research groups obviously used different personnel (primarily neurologists) to assign subjects to groups, it is unlikely that the same criteria were used across studies.

This argues strongly in favor of a procedure that involves a contemporaneous or 'head to head' comparison of the HRNTB and the LNNB in lateralizing the same groups of brain-damaged subjects. Although, as noted, there are significant questions regarding the Osmon et al. (1979) investigation, there are good reasons to suspect that the LNNB and the HRNTB are capable of lateralizing brain damage with approximately the same degree of success. For example,

Golden, Kane, Sweet, Moses, Cardellino, Templeton, Vincente, and Graber (1981) have demonstrated that a high degree of shared variance exists between the two batteries. Using a multiple correlation procedure, it was found that the 14 LNNB summary scales were highly correlated with each of 14 HRNTB variables [Category Test; Tactile Performance Test – Time, Memory, and Location; Rhythm Test; Speech Sounds Perception Test; Trail Making Test – A and B; Aphasia Screening Test; Sensory-Perceptual Test; Finger Tapping Test – Dominant and Non-Dominant, Wechsler Adult Intelligence Scale (WAIS) Verbal IQ; and WAIS Performance IQ; for a description of these tests see Reitan & Davison, 1974]. When each of the HRNTB variables was used as a dependent measure with the LNNB summary scales as predictors, the multiple correlations ranged from .71 (dependent measure, Finger Tapping – Dominant) to .96 (dependent measure, WAIS-R Performance IQ), with a mean multiple correlation of .85. When each of the LNNB summary scales was used as a dependent measure with the HRNTB variables as the predictors, the multiple correlations ranged from .77 (dependent measure, LNNB Rhythm) to .94 (dependent measure, Visual), with a mean multiple correlation of .86. Although a canonical analysis (Kerlinger & Pedhazur, 1973) would have been a more appropriate technique to use with these data, the high degree of shared variance (approximately 73%) between the LNNB and the HRNTB suggests that the two batteries overlap considerably in the basic set of skills they each assess

and, hence, should be equally suited in lateralizing brain damage.

Further, 'head to head' comparisons diagnosing the presence of brain damage have demonstrated the essential equality of the two batteries in discriminating between the same groups of control and neurological subjects. Use of a discriminant analysis on the 14 LNNB summary scales was able to achieve a hit rate of 87% in the neurological group (42/48) and a hit rate of 88% in the control group (53/60), while a discriminant analysis on the 14 HRNTB variables achieved a 90% hit rate in the neurological group (43/48) and a 84% hit rate in the control group (50/60) (Golden, Kane, et al., 1981). Comparable hit rates for the two batteries in detecting the presence of brain damage (each approximately 80%) were also obtained in a blind, expert, clinical interpretation of protocols from a mixed psychiatric and brain-damaged sample (Kane, Sweet, Golden, Parsons, & Moses, 1981).

For the clinician who is interested in the use of a standardized battery approach to neuropsychological assessment and is confronted by a choice between the LNNB and the HRNTB, an important consideration, over and above the comparative effectiveness of the batteries in reflecting brain-behavior relationships (e.g., lateralizing brain damage), is the time required for the administration of the tests. While the test manual (Golden, Hammeke, & Purisch,

1980) and the promotional material (Western Psychological Services, 1984) state that the LNNB can be administered in 1.5 - 2.5 hours, a considerably shorter period of time than for HRNTB administration (between 5 - 8 hours, Boll, 1981), there are no data demonstrating that this is so. Although Golden's research publications (e.g., Golden, Hammeke, & Purisch, 1978; Hammeke, Golden, & Purisch, 1978) state that the battery is completed in approximately 2 hours, more recent reports from Golden's laboratory (Moses, Golden, Berger, & Wisniewski, 1981) and elsewhere (Malloy & Webster, 1981) suggest that the battery's administration may take considerably longer. While Malloy and Webster (1981) indicated that administration time was dependent upon the subject's level of impairment and was completed within 4 hours, Moses et al. (1981) state that "the time of testing varied from between 2 hours in the more intact patients to over 5 hours in the more severely impaired patients" (p.96).

Since the demand for neuropsychological assessment clearly exceeds the current available resources (i.e., qualified clinical neuropsychologists and trained test administration technicians) (Craig, 1979; Golden & Kuperman, 1980; Hartlage & Telzrow, 1980; Meier, 1981; Noonberg & Page, 1982), the introduction of a new standardized neuropsychological battery (LNNB) that is purported to perform equally well in a much shorter time frame than the major battery in current use (HRNTB) is very favorably

received. Thus, the major purpose of the current investigation is to examine the comparative diagnostic effectiveness of the LNNB and the HRNTB in lateralizing brain damage using discriminant analysis classification procedures and to provide a comparison of their respective administration times.

Objective Clinical Rules

Golden and his colleagues have developed a series of objective rules which, when applied to LNNB protocols, are aimed toward aiding the clinical neuropsychologist in his or her decision-making regarding questions of the presence, lateralization, and localization of cerebral lesions. While the present investigation examined the clinical rules for detecting the presence and lateralization of brain damage, its scope did not include examining the utility of the rules for localizing brain damage. The later were examined only in so far as they related to the lateralization question.

Detecting the Presence of Brain Damage

Golden, Moses, Graber, and Berg (1981) found that normal subjects rarely perform at a level 10 T score points above (high scores indicate impairment) that which may be predicted from a knowledge of the subjects' age and education (a 'critical level') on any of the LNNB summary scales. Based on this, they developed a rule such that the

presence of brain damage is inferred if two or more summary

scale $\underline{T}$ score is above the critical level (see Appendix 1

for a more complete discussion on the derivation of this

rule). If the Arithmetic or Writing scales are above the

critical level, a subject is not considered to be brain

damaged unless three or more scales are above the critical

level.

Application of this rule to derivation samples and to

cross-validation samples resulted in hit rates of 91% and

84%, respectively (Golden, Moses, Graber, & Berg, 1981). A

similar high hit rate (86%) using the critical level

procedure has been reported by Golden, Moses, Fisburne,

Engum, Lewis, Wisniewski, Conley, Berg, & Graber (1982).

Independent cross-validation of the procedure carried out by

Malloy and Webster (1981) again demonstrated the high degree

of accuracy (78%) that may be achieved in detecting the

presence of brain damage.

Lateralizing Brain Damage

Although the LNNB has two summary scales (Left

Hemisphere and Right Hemisphere scales) thought to be

reflective of the integrity of the cerebral hemispheres,

McKay and Golden (1979a, 1979b) have empirically derived

additional LNNB scales that are considered to be highly

effective in lateralizing brain damage. The alternate sets

of items designed specifically for lateralizing brain damage

were derived by statistically comparing the performance of a group of normal subjects with groups of laterally damaged subjects on each of the 269 items of the battery and retaining those items for the new scales on which performance of the lateralized groups differed from the normal subjects (McKay & Golden, 1979b; see Appendix 1, for a discussion of this derivation procedure). The empirically derived Right- and Left Hemisphere scales, designated R* and L* respectively (Golden, Moses, Fishburne, et al., 1981), were found to discriminate right-hemisphere subjects from left-hemisphere subjects with 100% accuracy in the derivation samples by simply using the arithmetic sign that results from the R* score minus L* score operation (McKay & Golden, 1979b). McKay and Golden (1979b) also report that the R* - L* difference score was able to yield a 87% hit rate in the application of the sign rule to cross-validation samples of right- and left-hemisphere-damaged subjects.

McKay and Golden (1979a) have empirically derived additional LNNB scales to localize brain damage using a procedure similar to that employed in the empirical derivation of the lateralization scales (McKay & Golden, 1979b). The performance of neurologically normal subjects was compared, on each of the 269 LNNB items, to the performance of each of eight groups of subjects with localized brain damage (left and right frontal, sensorimotor, parietal-occipital, and temporal). The eight

scales were derived such that each scale (named after its corresponding group) consisted of items that were found to statistically discriminate a particular locally brain-damaged group from the normal group with as little item overlap across scales as possible. Use of a rule such that the highest score from the eight localization scales suggests the cerebral localization of brain damage, McKay and Golden (1979a) were able to achieve an 89% hit rate in their derivation, neurologically impaired, sample. McKay and Golden (1979a) do not report a cross-validation of this localization rule.

There are significant problems with the derivation of the lateralization and localization scales and, hence, with their associated clinical rules. For example, the very small size of the brain-damaged samples used (mean sample size = 6.6, McKay & Golden, 1979a; mean sample size = 14.5, McKay & Golden 1979b) is compounded by the extravagant use of sequential t-tests (2152, McKay & Golden, 1979a ; 538, McKay & Golden, 1979b) without correction for the consequent inflation of the Type 1 error rate. It is hence surprising that a cross-validation of the lateralization and localization objective clinical rules carried out by Golden, Moses, Fishburne, et al. (1981) was so successful in replicating the high hit rates of the derivation studies. The application of the R* - L* difference clinical rule yielded a 71% hit rate for right-hemisphere-damaged subjects

and an 84% hit rate for left-hemisphere-damaged subjects,
for an overall hit rate of 78%. The highest score on the
localization scales was able to correctly localize 75% of
the brain-damaged subjects, with accuracy for the right-
hemisphere-localized lesions (66%) less than the accuracy
for the left-hemisphere-localized lesions (81%). Of
particular interest to the present investigation is the
finding that the use of the highest score on the
localization scales rule to lateralize subjects resulted in
higher levels of discrimination than that found using R* -
L* difference rule (right-hemisphere-damaged subjects, 95%;
left-hemisphere-damaged subjects, 94%).

Although the data presented by Golden, Moses, Graber
and Berg (1981), Golden, Moses, Fishburne et al. (1981),
Malloy and Webster (1981), and McKay and Golden (1979a,
1979b) appear impressive in indicating the potential of the
LNNB to detect, lateralize, and localize cerebral lesions,
the methodological and statistical problems (e.g. small
samples and multiple, uncorrected t-tests) with the data
base (see also Appendix 1), suggest a great need for
independent research to cross-validate the basic data before
unequivocal statements may be made concerning the efficacy
of the objective clinical rules in making diagnostic
decisions.

In summary, the purpose of this project was to attempt
to replicate the high degree of statistical classificatory

accuracy reported by Osmon et al. (1979) in lateralizing
brain damage with the LNNB and to provide a test of the
comparative diagnostic efficiency of the LNNB and the HRNTB
in determining the laterality of brain damage. Of interest
here as well was a comparison of the LNNB and HRNTB in terms
of their respective administration times. An ancillary goal
of this investigation was to provide a cross-validation of
the objective clinical rules developed by Golden and his
team to determine the presence of brain damage (critical
level proceedure) and the laterality of lesion focus (R* -
L*, highest score on the localization scales).

To achieve these ends, a discriminant analysis
classification procedure was applied individually to the
LNNB data, and to the HRNTB data, generated by the timed
administration of these batteries to a predominantly left-
hemisphere-damaged sample and to a predominantly right-
hemisphere-damaged sample. Subjects were assigned to the
criterion groups based on the laterality of brain damage as
documented on independent neurological grounds. Hit rates
derived from the application of the classification procedure
to the LNNB data and the HRNTB data were then compared for
the examination of the comparative effectiveness of the
batteries in lateralizing brain damage.

The cross-validation of the objective clinical rules
was be accomplished by 'running the rules' on the laterally
brain-damaged samples and comparing the consequent hit rates
with those reported in the literature.

Method

Subjects

Subjects for this investigation were 30 predominantly left- or right-hemisphere-damaged patients (15 in each group) who were referred to the Neuropsychology Service at the Rehabilitation Hospital, Health Sciences Center, in Winnipeg, Manitoba, for neuropsychological assessment. All patients referred for neuropsychological assessment were potentially in the subject pool with the selection of usable subjects being dependent upon neurological evidence of predominantly right- or left-hemisphere damage. This decision and consequent assignment to criterion groups was made by a qualified neurologist following a review of data from all relevant clinical neurological investigations performed on each subject as dictated by clinical medical need and included one or more of the following tests: clinical neurological exam, skull x-rays, EEG, angiogram, CAT scan, and surgery. All of these procedures were performed independently of the administration of the LNNB and HRNTB. The left-hemisphere-damaged group was composed of 11 subjects with closed head injury and 4 with cerebrovascular accidents. This distribution of central nervous system damage did not differ (chi square $(2, \underline{N} = 30)$ = 1.874, $\underline{p} > .30$) from the right-hemisphere-damaged group which was made up of 8 closed head injured subjects, 6 with cerebrovascular accidents, and one who had had a tumor excised.

As Table 1 indicates, the right- and left-hemishere-damaged groups were found not to differ in terms of age, education, and chronicity (time since onset of deficits). The groups did not differ in their breakdown of sex of subject, handedness, or whether they were inpatients or outpatients when they were referred for neuropsychological assessment (see Table 2). The groups also did not differ on overall measures of severity of neuropsychological dysfunction as assessed by what are typically used as indicators of severity (e.g., Impairment Index, Reitan & Davison, 1974; Average Impairment Rating, Russell, Neuringer, & Goldstein, 1970, see Table 5; Pathognomonic Scale, Golden, Hammeke, & Purisch, 1980, see Table 4).

Subjects were assessed by the LNNB and the HRNTB while on the medications they were currently prescribed as there is no evidence that medications that have psychoactive properties differentially effect the performance of laterally brain-damaged subjects on neuropsychological tests (see Dodrill, 1981; and Heaton & Crowley, 1981, for reviews of the neuropsychological consequences of psychoactive medication). At the time of their assessments, 6 subjects were known to be taking medication with psychoactive properties (left-hemisphere-damaged group - 3 subjects taking anticonvulsants; right-hemisphere-damaged group - 2 subjects taking anticonvulsants, 1 subject taking a minor tranquillizer). Of them, no subject was taking medication at

TABLE 1

Age, Education, Chronicity, and Time Between Tests

Variable	Group		
	Left	Right	
	Mean (SD)	Mean (SD)	t(28)
Age (years)	37.60 (12.99)	42.60 (18.49)	-0.86
Education (years)	12.67 (2.35)	11.00 (2.36)	1.94
Chronicity (months)	16.53 (21.45)	17.67 (54.90)	-0.07
Time Between Tests (days)	9.27 (8.53)	7.20 (10.25)	0.60

Note. All t-tests are two-tailed. Left = Left-Hemisphere Damaged; Right = Right-Hemisphere-Damaged.

*p<.05.

TABLE 2

Sex, Handedness, Patient Status, and Test Administered First

	Group		
Variable	Left	Right	Chi-square(1,$\underline{N}$=30)
Sex (# of males)	14	11	0.96
Handedness (# of right-handers)	14	15	0.90
Patient Status (# of inpatients)	5	6	0.14
Test Used First (# with LNNB first)	2	2	0.00

Note. Left = Left-Hemisphere-Damaged
 Right-Hemisphere-Damaged.

*$\underline{p}$<.05.

doses where toxicity was evident or, where medication was expected to significantly alter cognitive function. More detailed clinical data regarding the subjects is available from the author.

Procedure

All patients were administered the full LNNB (269 items) according to the instructions outlined in the current test manual (Golden, Hammeke, & Purisch, 1980) and what is considered to be the full HRNTB using current conventions regarding administration (Boll, 1981; Reitan & Davison, 1974; Russell, Neuringer, & Goldstein, 1970). The items of the LNNB are divided into 11 basic scales of unequal length (Motor Functions, Rhythm, Tactile Functions, Visual Functions, Receptive Speech, Expressive Speech, Writing, Reading, Arithmetic, Memory, and Intellectual Processes) and three additional basic scales which are formed by a recombination of some of the items (Left- and Right Hemisphere scales and the Pathognomonic scale, see Golden, Hammake, & Purisch, 1980; and Appendix 1). The R* and L* scales and the eight localization scales, Right Frontal (RF), Right Sensorimotor (RSM), Right Temporal (RT), Right Parietal-Occipital (RPO), Left Frontal (LF), Left Sensiomotor (LSM), Left Temporal (LT) and Left Parietal-Occipital (LPO), are formed in a similar manner by recombining the items that were found to discriminate the

relevant laterally or locally brain-damaged groups from
normals (McKay & Golden, 1979a, 1979b; see Appendix 1, for a
discussion). The HRNTB consisted of the following tests:
Category Test, Tactile Performance Test, Speech Sounds
Perception Test, Rhythm Test, Finger Tapping Test, Trail
Making Test, Grip Strength Test, Aphasia Screening Test,
Sensory-Perceptual Test and the WAIS-R (Boll, 1981; Reitan &
Davison, 1974; Russell, Neuringer, & Goldstein, 1970).

Both batteries were administered by trained
neuropsychological assessment technicians and there was no
difference between groups in terms of the length of time
between administrations of each battery (see Table 1).
Although it would have been desirable to counterbalance the
order of administration of the batteries, this goal was not
possible due to the demands of clinical practise where the
HRNTB was the primary instrument in clinical use. The groups
did not, however, differ in terms of which battery was
administered first (see Table 2) and there has been no
report of order effects following the counterbalanced
administration of both batteries (Goldstein & Shelly, 1984).

Analysis

LNNB Accuracy Lateralizing Brain Damage.

The ability of the LNNB to lateralize brain damage was
assessed by the use of discriminant analysis classification

procedures (Fletcher, et al., 1978; Huberty, 1975; Tatsuoka, 1971, 1974) using initially, as did Osmon et al. (1979), the $\underline{T}$ scores (see Appendix 1) for the 14 summary scales as the dependent measures and the neurologically determined lateralization of brain damage as the independent variable. Discriminant analysis is a procedure that achieves maximal statistical seperation between groups by forming composite variables based upon the weighted linear combination of individual dependent measures. The DISCRIMINANT subprogram of the $\underline{SPSSX}$ (SPSS Inc., 1983) was employed with, as the first step, all 14 dependent measures being included in the resultant discriminant function. While this full model analysis classification procedure (cf., Lachin & Schachter, 1974) is an analogue of the procedure utilized by Osmon et al. (1979), and thus represents a replication of their analytic technique, statistical problems inherent in such analyses with low subject to variable ratios are present. Although a significant full model multivariate analysis would indicate true separation between the left- and right-hemisphere-damaged groups (such was not reported by Osmon et al., 1979), the analysis would not take into account the possibility that many of the 14 entered variables do not add significantly to the group separation, and thus may adventitiously take advantage of random variation in the samples (Fletcher, et al., 1978; Appendix 1).

In an effort to deal with this, the 14 summary scales were then entered into a step-wise discriminant analysis procedure (SPSSX) with variables being retained for the final, reduced model, solution if they met the following statistical criteria: probability of F-to-enter <.05, and probability of F-to-remove <.05. Simply, variables were added to, or removed from, the discriminant function if, in combination with variables already in the equation at any given step, their unique variance provided group separation according to a specified statistical criterion (here, p<.05). While this strategy prevents dependent variables being added to the discriminant analysis that do not add significantly to the group separation, it affords a conservative estimate of the battery's discriminatory power. Thus, the results from both the full model analysis and the reduced model analysis serve to bracket the discriminatory ability of the battery in the sense that a liberal and a more conservative estimate has been provided. Since the left- and right-hemisphere groups were of equal size, a prior probability of .5 was used for the classifications.

The 2 (neurologically lateralized) X 2 (LNNB lateralized) classification matrix for both the full and reduced models was then analyzed by the kappa statistic (Cohen, 1960; Fleiss, Cohen, & Everitt, 1969) to determine the amount of non-chance concordance between the two methods of classification, with a one-tailed proportional chance

criterion test (Huberty, 1984) being used to determine whether the observed classification accuracy was significantly greater than what would be expected by chance.

Comparative Effectiveness of the LNNB and HRNTB.

To effect the comparison of the ability of the LNNB and the HRNTB to lateralize brain damage, two sets of discriminant function classification procedures were employed using the neurologically documented lateralization of brain damage as the independent variable; one set on the LNNB summary scales (carried out above) and one set on HRNTB measures (see Table 3).

The discriminant analysis classification procedure on the 14 HRNTB variables was carried out exactly as was the procedure for the LNNB variables. The DISCRIMINANT subprogram was again be used with all 14 dependent measures being included in the resultant discriminant function for the full model analysis. As before, to provide a more conservative estimate of the battery's discriminative power, the 14 HRNTB variables were entered into a stepwise discriminant analysis procedure with those retained if they met a probability of F-to-enter $<.05$, and a probability of F-to-remove $<.05$ (reduced model solution). This ensured that variables that entered the discriminant function did add significantly to the discrimination between groups and were not taking advantage of chance variation. A prior

TABLE 3

LNNB and HRNTB Dependent Variables

LNNB Variables	HRNTB Variables
Motor Functions	Category Test
Rhythm	Tactile Performance Test (RH-LH)
Tactile Functions	Tactile Performance Test-Memory
Visual Functions	Tactile Performance Test-Location
Receptive Speech	Finger Tapping Test (RH-LH)
Expressive Speech	Trail Making Test-A
Writing	Trail Making Test-B
Reading	Grip Strength Test (RH-LH)
Arithmetic	WAIS-R Verbal IQ
Memory	WAIS-R Performance IQ
Intellectual Process	Aphasia Screening Test
Pathognomonic	Sensory-Perceptual Test (Right-Left)
Left Hemisphere	Speech Sounds Perception Test
Right Hemisphere	Seashore Rhythm Test

Note. RH = right hand, LH = left hand.

probability of .5 was used for classification for both the full and reduced model analyses. The kappa statistic was used to analyze the 2 (neurologically lateralized) X 2 (HRNTB lateralized) classification matrix to assess the amount of the agreement between the two classification methods and a proportional chance criterion (one-tailed) was used to determine whether the classification accuracy obtained was significantly greater than chance rates.

The hit rates obtained by the LNNB and the HRNTB were then compared using a chi-square goodness-of-fit test (Siegal, 1956; Hays, 1973) to examine the comparative accuracy of the batteries in lateralizing brain damage. Times for administration for each battery were examined with t-tests.

Objective Clinical Rules.

1. Critical Level Procedure. To determine if a subject was brain-damaged using the critical level procedure (Golden, Moses, Graber, & Berg, 1981), each subject's critical level was determined by the following formula:

Critical Level= 68.8 + .214 (age in years) - 1.47 (education in years).

As outlined in Golden, Moses, Fishburne, et al. (1981), postsecondary education is credited by the

number of years of education with 16 years for a
bachelor's degree or equivalent, 18 years for a
master's or equivalent, and 20 years for a Ph.D. or
equivalent. Once the critical level is established,
the number of elevated T scores on the summary scales
is determined (excluding Left- and Right Hemisphere
summary scales). Zero or 1 summary scale T scores
above the critical level represents normal
performance and 3 or more summary scale T scores
above the critical level is indicative of brain
damage. A score of 2 is considered borderline: if
either of the elevated summary scales is Writing or
Arithmetic, the subject is considered normal;
otherwise the subject is considered to be brain
damaged (Golden, Moses, Fishburne, et al., 1981).

2. <u>Lateralizing Brain Damage</u>. Lateralizing brain damage
by use of the R* - L* difference involves calculating
the T score difference between the R* and L* scales
(McKay & Golden, 1979b) and then classifying the
subject right-hemisphere damaged if the difference is
positive (R* higher) or left-hemisphere damaged if
the difference is negative (L* higher). The
lateralization of subjects using the eight
localization scales involves determining the T scores
for each of the localization scales (McKay & Golden,
1979a) and classifying subjects as left- or right-
hemisphere damaged based on the lateralization

indicated by the highest $\underline{T}$ score. For example, if the highest localization $\underline{T}$ score for a particular subject was on the RF scale, this subject would be classified as being right-hemisphere damaged.

Results

The data for this investigation are reported by presenting analyses on the LNNB initially and then, in a similar format, the data for the HRNTB are presented before the batteries are compared as to their effectiveness in lateralizing brain damage. Following this, the objective clinical rules are evaluated.

LNNB Accuracy Lateralizing Brain Damage

Full Model Analysis. When all 14 summary scales were submitted to multivariate analysis of variance for a full model test of significance (cf., Lachin & Schachter, 1974), the left- and right-hemisphere-damaged groups were found to be significantly different (Wilks' Lambda = .293, equivalent Multivariate $\underline{F}(14,15) = 2.585$, $\underline{p}< .04$). Univariate $\underline{t}$-tests carried out on the 14 summary scales revealed significant differences on only 2 scales, the Left and Right Hemisphere Scales. The means and standard deviations for the summary scales are presented in Table 4.

TABLE 4

Means and Standard Deviations for the LNNB Summary Scales

Variable	Group		
	Left Hemisphere	Right Hemisphere	
	Mean (SD)	Mean (SD)	t(28)
Motor	59.80 (22.05)	60.33 (22.12)	−0.07
Rhythm	53.60 (20.68)	56.87 (21.51)	−0.42
Tactile	54.87 (13.61)	59.60 (14.99)	−0.91
Visual	57.67 (15.72)	56.80 (13.45)	0.16
Receptive	58.13 (24.57)	51.00 (11.29)	1.02
Expressive	58.00 (27.78)	54.20 (10.14)	0.46
Writing	63.40 (18.73)	57.93 (12.83)	0.93
Reading	60.67 (20.28)	53.33 (13.73)	1.16
Arithmetic	65.07 (32.61)	59.00 (18.59)	0.63
Memory	61.53 (15.30)	57.33 (10.40)	0.88
Intellectual	56.00 (19.76)	59.33 (14.20)	−0.53
Pathognomonic	60.73 (23.02)	56.33 (13.36)	0.64
Left Hemisphere	68.13 (29.22)	50.93 (10.45)	2.15*
Right Hemisphere	51.13 (13.11)	74.93 (33.43)	−2.57*

*$p < .05$.

A discriminant analysis classification procedure using the resultant classification functions was able to correctly classify 14 of 15 left-hemisphere-damaged subjects (93%) and 15 of 15 right-hemisphere-damaged subjects (100%), for an overall hit rate of 96.7%. The hit rates for each group were significantly greater than chance expectation (left group, z = 3.11, $p<.001$, and right group, z = 3.62, $p<.001$). The overall hit rate corresponded to a Kappa of .93 indicating a significant (z = 5.11, $p<.0001$) improvement in classification accuracy over chance.

Reduced Model. Application of the reduced model procedure (probability of F-to-enter and to-remove$<.05$) yielded a 2-variable solution with the Right Hemisphere scale entering at step 1, and the Left Hemisphere scale entering at step 2. No other variable was found to contribute significantly to the discrimination and thus, variables that were forced to enter past this using the full model solution were exploiting random variation. With this 2-variable solution the right- and left-hemisphere-damaged groups were significantly different (Wilks' Lambda = .591, equivalent Multivariate $F(2,27)$ = 9.35, $p<.0008$). A classification procedure based on the resultant 2-variable solution was able to correctly classify 13 of 15 left-hemisphere-damaged subjects (86.7%, z = 2.59, $p<.005$), and 10 of 15 right-hemisphere-damaged subjects (66.7%, z = 1.04, $p>.05$). As indicated, while the hit rate for left-

hemisphere-damaged subjects was significantly greater than chance, the hit rate for right-hemisphere-damaged subjects did not exceed chance expectation. The combined hit rate of 76.7 (Kappa = .53) represents classification accuracy at a level significantly greater than the chance level (z = 2.921, p<.002). This 2-variable solution was also obtained when the probability of F-to-enter <.10, and the probability of F-to-remove <.10.

Comparative Effectiveness of the LNNB and HRNTB in Lateralizing Brain Damage

As an initial step in examining the comparative effectiveness of the batteries in lateralizing brain damage, a similar series of analyses were conducted on HRNTB data as were conducted on the LNNB summary scales.

Full Model Analysis. With all 14 HRNTB variables entered in a full model test of significance, the left- and right-hemisphere-damaged groups were found to be significantly different (Wilks' Lambda = .2829, equivalent Multivariate F(14,15) = 2.716, p<.0322). As indicated in Table 5, univariate t-tests revealed significant differences between the 2 groups on the Tactile Performance Test-Time per Block, Tactile Performance Test-Location, Finger Tapping, Sensory-Perceptual Examination, and in terms of Grip Strength.

TABLE 5

Means and Standard Deviations for the HRNTB Variables

	Group		
Varible	Left	Right	t(28)
	Mean (SD)	Mean (SD)	
WAIS-R Verbal IQ	91.33 (16.11)	91.60 (11.13)	-0.05
WAIS-R PIQ	88.33 (12.83)	81.20 (11.14)	1.63
Category Test	50.87 (28.79)	57.00 (27.31)	-0.60
TPT (RH-LH)	204.23 (240.12)	-114.53 (206.46)	3.90**
TPT Memory	6.53 (1.41)	5.40 (2.17)	1.70
TPT Location	3.73 (2.40)	1.47 (1.51)	3.09*
Rhythm	6.47 (4.52)	8.80 (4.63)	-1.40
Speech	11.53 (11.67)	11.93 (10.70)	-0.10
Finger Tap (RH-LH)	-11.20 (18.98)	23.67 (20.55)	-4.83**
Trails A	75.40 (49.55)	97.93 (144.91)	-0.57
Trails B	229.40 (216.70)	234.67 (179.94)	-0.07
Aphasia	13.60 (15.26)	7.27 (6.50)	1.48
Sensory (RH-LH)	8.93 (16.69)	-14.87 (17.21)	3.84**
Grip (RH-LH)	-11.87 (20.07)	10.73 (21.43)	-2.98*
Impairment Index	0.55 (0.27)	0.67 (0.25)	-1.26
Average Impairment Rating	2.14 (0.84)	2.39 (0.89)	-0.81

*p<.01.   **p<.001.

The classification procedure based on the discriminant analysis classification functions was able to correctly classify 13 of 15 left-hemisphere-damaged subjects (86.7%, z =2.59, $p<.005$), and 14 of 15 right-hemisphere-damaged subjects (93%, z = 3.11 ,$p<.001$). The overall hit rate of 90% represents a Kappa of .80, and this classification function is significantly more accurate than a chance classification (z = 4.38, p <.0001).

Reduced Model Analysis. Similar to the LNNB data set, application of the reduced model procedure resulted in a 2-variable solution obtaining with Finger Tapping (right hand - left hand) entering at step 1 and the Location score from the Tactile Performance Test at step 2. With this 2-variable solution, the right- and left-hemisphere-damaged groups were significantly different (Wilks' Lambda = .4332, equivalent Multivariate F(2,27) = 17.66, $p<.0001$) and application of the discriminant analysis classification procedure yielded correct classification of 12 of 15 left-hemisphere-damaged subjects (80%, z = 2.07, $p<.02$), and 13 of 15 right-hemisphere-damaged subjects (86.7, z = 2.59, $p<.005$),for an overall hit rate of 83.3% (Kappa = .67). This rate of classification accuracy is significantly greater than that expected by chance (z = 3.651, p <.0002).

Comparative Effectiveness. Using rates of classification based on the discriminant functions from the full model (14 variable) analyses, the LNNB and the HRNTB

were found not to differ (chi square (1, $\underline{N}$ = 30) = .138, $\underline{p}$>.50) in their classification of left-hemisphere subjects (LNNB, 14/15; HRNTB, 13/15) and right-hemisphere subjects (LNNB, 15/15; HRNTB, 14/15). Also, no differences were observed (chi square(1, $\underline{N}$ = 30) = .823, $\underline{p}$>.30) between the two batteries in their ability to lateralize brain damage when stringent statistical criteria were used such that only variables that significantly contributed to the group separation were retained in the solution (i.e., reduced model). The 2-variable solution for both batteries yielded no difference in the classification of left-hemisphere subjects (LNNB, 13/15; HRNTB, 12/15), and in the classification of right-hemisphere subjects (LNNB, 10/15; HRNTB, 13/15).

As is shown in Table 6, despite the no differences in accuracy classifying subjects as right- or left-hemisphere-damaged, the LNNB was found to take less than half the time to administer as compared to the HRNTB, and this holds for test administration to both left- and right -hemisphere groups. While administration time for the LNNB did not differ for both groups (see Table 6 for means and standard deviations, $\underline{t}$(28) = .36, $\underline{p}$>.50), the HRNTB required significantly more administration time for the left-hemisphere group than for the right-hemispere group ($\underline{t}$(28) = 2.33, $\underline{p}$<.03).

TABLE 6

Administration Times for the LNNB and HRNTB

Administration Time

in Minutes

Group	LNNB		HRNTB		t(14)
	Mean	(SD)	Mean	(SD)	
Left-Hemisphere	164.40	(47.02)	433.00	(157.89)	7.36*
Right-Hemisphere	159.00	(35.32)	332.40	(54.39)	12.10*
Combined Groups	161.70	(40.95)	382.70	(126.81)	10.43*

Note. All t-tests reported are correlated sample tests. The t-test for the combined groups had df = 29

*p<.001.

Objective Clinical Rules.

1. Critical Level Procedure. Use of the critical level
    procedure, as outlined in the Procedure Section, to
    determine if a subject was brain-damaged resulted in
    the overall correct classification of 20 of 30
    subjects (67%) with an equal number being correctly
    classified in each laterally damaged group (10/15).
    Three subjects in the left-hemisphere group, and 2
    subjects in the right-hemisphere group had no LNNB
    summary scale T-scores above the critical level.

2. Lateralizing Brain Damage. Lateralizing brain damage
    by using the arithmetical difference between the T-
    scores on the R* and L* scales (R* - L*) resulted in
    the correct lateralization of 11 of 15 (73.3%, z =
    1.81, p<.04) right-hemisphere-damaged subjects and 13
    of 15 (86.7%, z = 2.59, p<.005) left-hemisphere-
    damaged subjects, for a combined hit rate of 80%
    (Kappa = .60). This observed hit rate was
    significantly greater than expected by chance (z =
    3.29, p <.001). Use of the localization scale with
    the highest T-score to indicate lateralization
    resulted in correct classification in 12 of 15 (80%,
    z = 2.59, p<.005) of the right-hemisphere group and
    11 of 15 (73%, z = 1.81, p<.04) of the left-
    hemisphere group. The combined hit rate of 76.7%
    (Kappa = .53) was found to be significantly higher

than chance expectation (z = 2.92, p <.002). The classification accuracy for the use of the R* - L* rule did not differ from the hit rate obtained using the highest localization scale to lateralize (chi square (1, N = 30) = .265, p>.50). When the rules for lateralizing brain damage were applied only to those subjects who were correctly classified as brain damaged, the R* - L* difference was able to lateralize 85% of the identified brain-damaged subjects (left-hemisphere group, 10/10; right-hemisphere group, 7/10), and the highest localization scale was able to lateralize 80% of the identified brain-damaged subjects (left-hemisphere and right-hemisphere groups both with 8/10). Both of these classification hit rates differ significantly from chance expectation(R* - L* rule, Kappa = .70, z = 3.13, p<.001 ; highest-localization-scale rule, Kappa = .60, z = 2.68, p<.02).

Discussion

Overall, this investigation has provided an independent replication of the ability of the LNNB to lateralize brain damage in terms of finding significant differences between right- and left-hemisphere-damaged subjects, and in terms of being able to correctly classify subjects as to the hemisphere of lesion at rates greater than chance using

discriminant analysis classification procedures. The LNNB
was able to achieve overall hit rates comparable to the
HRNTB in lateralizing brain damage using the discriminant
analysis and this was achieved in approximately one half the
time it was found to take to administer the HRNTB. Despite
these generally positive findings, application of the set of
clinical rules that have been offered to diagnose the
presence of brain damage, and to lateralize it, yielded hit
rates below those reported by Golden and his associates;
this points to the potentially limited utility such rules
may have in clinical practice.

LNNB Accuracy Lateralizing Brain Damage

Following the demonstration that the left- and right-
hemisphere-damaged groups did not differ on important
demographic variables known to influence neuropsychological
function (Parsons & Prigatano, 1978) or on overall indices
of severity of neuropsychological impairment, this study has
provided a cross-validation of the ability of the LNNB to
lateralize cerebral dysfunction. The Osmon et al. (1979)
report of 98% classification accuracy in placing subjects in
either left- (20/20), right- (20/20), or diffuse-brain-
damage (19/20) groups while impressive, can not be
unquestioningly accepted. The failure to report the
Multivariate F value to document statistically reliable
group differences, and the failure to establish that all

variables entered into the discriminant equation added significant increments to discriminatory power has caused concern and has pointed to the need for cross-validation of the original findings (Adams, 1980a; Appendix 1). The present findings, however, provide a clearer assessment of the battery's ability to lateralize brain damage. Multivariate analyses do demonstrate that right- and left-hemisphere-damaged subjects are significantly different on the LNNB, and that application of classification functions derived from the discriminant analyses can produce hit rates that do differ from chance expectation.

Use of the full model analysis yielded an overall hit rate (96.7%) comparable to the hit rate (98%) reported by Osmon et al. (1979), but when this current full model hit rate is compared to the reduced model hit rate, a 20% reduction is evident. Given that the reduced model is based on only those LNNB summary scales that significantly add to group separation, the observed 76.7% hit rate is felt to be more representative of the power of the battery in lateralizing brain damage. Based on the reduced model hit rates, there is a suggestion that the LNNB is more effective in determining lesion location in left-hemisphere-damaged subjects (13/15, 86.7%) than in right-hemisphere-damaged subjects (10/15, 66.7%). While this speculation will need to be more thoroughly examined in large sample data sets, it is consistant with the Lurian theoretical antecedants of the

LNNB which focussed on the assessment of left-hemisphere

functioning (Luria, 1980, p. 587), statements made in the

LNNB administration guide regarding greater sensitivity to

left-hemisphere lesions (Golden, Hammeke, & Purisch, 1980,

p. 56), and empirical work demonstrating the high verbal

loading on the items of the battery (see Appendix 1).

Comparative Effectiveness of the LNNB and HRNTB in Lateralizing Brain Damage

Administration of the HRNTB resulted in a full model

multivariate significant difference between the left- and

right-hemisphere groups thereby attesting to the ability of

the battery to be sensitive to the lateralization of

cerebral damage. A full model hit rate of 90% based on the

discriminant analysis classification procedure compares

favorably with what have been full model analyses previously

reported in the literature in differentiating left- and

right-hemisphere-damaged subjects (e.g., 95%, Golden, 1977;

88%, Goldstein & Shelley, 1973; 93%, Wheeler, Burke, &

Reitan, 1963; 88%, Wheeler & Reitan, 1963). The overall hit

rate of 83% obtained using the reduced 2-variable model also

compared well with those reported in the literature.

The contemparaneous, 'head-to-head', comparison of how

effective the LNNB and HRNTB are in lateralizing cerebral

damage demonstrated the practical equivalence of the

batteries in making this diagnostic determination. This held

true for both the comparison based on the full model analyses (LNNB, 96.7%; HRNTB, 90%) and the comparison based on the reduced model analyses (LNNB, 76.7%; HRNTB, 83%). This is an important finding because, as the LNNB test developers (Golden, Hammeke, & Purisch, 1980) claim that the battery can be as effective as the HRNTB, there has of yet not been a published report demonstrating that the two batteries do not differ in their ability to statistically discriminate laterally brain-damaged subjects.

Thus, decisions regarding which battery (HRNTB vs LNNB) to endorse for clinical practice will rest on considerations other than on whether the batteries produce sufficient data regarding brain-behavior relationships to statistically diagnose the presence of brain damage (Golden, Kane, et al., 1981), or determine the side of predominant lesion (this study). Although there are many dimensions along which any psychometric device may be evaluated [e.g., suitability for providing data on referral question, ease of administration and scoring, reliability, validity, etc., (see Stambrook, 1983, and more recently, Bryant, Maruish, Sawicki, & Golden, 1984; Delis & Kaplan, 1983)], time for administration is an important consideration given the demand for neuropsychological services. Based on the present data, overall, the HRNTB (mean administration time, 383 min) does take over 2 times longer to administer than the LNNB (mean administration time, 162 min) which is consistent with the

claim for the LNNB made in the test manual (Golden, Hammeke, & Purisch, 1980). However, it was evident that the claims that the battery can be administered in "only" 1.5 to 2.5 hours (Golden, Hammeke, & Purisch, 1980; Western Psychological Services, 1984) are somewhat misleading in that 50% of the current sample required more than 2.5 hours of administration time.

The administration time for the LNNB was found to be equivalent for subjects with left- and right-hemisphere-damage while the HRNTB took significantly longer to give to left-hemisphere-damaged subjects than those with right-hemisphere damage. The reason for this is that since the LNNB administration calls for adherance to specified time limits for individual items (approximately 1/3 of all items have specific time limits that range from 10-90 sec, and the rest are discontinued if no response is initiated within 15 sec), the time for administration is more controlled than with the HRNTB where no such limits are enforced. That is, having individual item time requirements serves to ensure that the LNNB can be given in a relatively short time as compared to the HRNTB and not be susceptible to such influences as, for example, the effects any form of aphasic disturbance may have on the time it takes to administer tests.

The severity of the aphasic impairment, as assessed by the Aphasia Screening Test, was found to be related to the

administration time for the HRNTB (r(28) = .65, p<.001) but not for the LNNB (r(28) = .43, p>.001). This suggests that there is a degree of flexibility in the HRNTB administration that can accomodate the longer latency of responding that is frequently apparent with aphasic patients or, patients with any disabling condition that can reduce response speed. In contrast, the specified time limits for the LNNB can prevent the use of the standardization norms to evaluate patient performance in situations where the requisite skills for a particular item are present, albeit slowly expressed. This issue has prompted some (Crosson & Warren, 1982; Delis & Kaplan, 1982) to question the appropriateness of the LNNB in the examination of aphasic patients and, more generally, has caused Spiers (1981) to suggest that response speed requirements confound all items on the battery.

Objective Clinical Rules

Use of the critical level procedure as developed by Golden, Moses, Graber, and Berg (1981), and advocated in the LNNB manual (Golden, Hammeke, and Purisch, 1980), to diagnose the presence of brain damage resulted in the correct classification of only 67% of the current brain-damaged subjects. This level is substantially lower than accuracy rates reported for the initial derivation (91%) and cross-validation samples (84%, Golden, Moses, Graber, & Berg, 1981), and rates reported in an independent

replication from Golden's laboratory (86%, Golden, Moses, Fishburne, et al., 1981), and from independent replications conducted in other laboratories (78%, Mallory & Webster, 1981; 92%, Sears et al., 1984). Recently, Sawicki and Golden (1984) have reported a more modest classification accuracy using the critical level procedure (77%) and have reiterated the earlier caution (Golden, Moses, Graber, & Berg, 1981) that "no single criterion from the LNNB should be used by itself to make a diagnostic decision" (p. 218). Although sample differences among studies can account for differences in hit rates, a overall hit rate of 77%, which represented a misclassification of approximately 20% of the brain-damaged subjects and 24% of normals (Sawicki & Golden, 1984) and, a hit rate of 67% in the present study, which represents a misclassification of 33% of brain-damaged subjects, is entirely unsatisfactory in the clinical setting for a test-manual-endorsed procedure that purports to be able to detect the presence of brain damage -- given what can be the tremendous human cost for misclassification.

The necessity of using multiple levels of inference (Reitan & Davison, 1974) and not simply a single level of performance criterion in the interpretation of neuropsychological test data was apparent as well for the HRNTB measures of impairment. Using established cutoff points for the determination of brain damage (Impairment Index > .40, Halstead, 1947; Average Impairment Rating >

1.55, Russell et al., 1970), 63% of the current sample were correctly classified as brain-damaged using the Impairment Index, while 73% were correctly classified using the Average Impairment Rating.

Application of the objective clinical rules for lateralizing brain damage in the present study resulted in an 80% hit rate for the R* - L* rule, and a 77% hit rate using the highest score from the localization scales to lateralize brain damage. The current hit rate for the R* - L* rule is consistent with the 78% hit rate reported by Golden, Moses, Fishburne, et al. (1981), but their high hit rate for lateralizing with the highest localization scale (94%) was not replicated here. Results on the same order of magnitude as found here were recently reported by Sears et al. (R* - L*, 80% accuracy; highest localization scale, 70% accuracy, 1984). Thus, independent laboratory cross-validation of the ability of the objective clinical rules to lateralize brain damage has indicated that the R* - L* rule can produce hit rates at approximately the 80% level, and the accuracy of the highest localization scale rule is more modest than has been previously reported (Golden, Moses, Fishburne, et al., 1981) and may also be at the 80% level.

Conclusions

In a recent major review of the literature on the LNNB, Stambrook (1983) argued that while research is suggesting that the battery may have potential as a standardized, quantitative neuropsychological instrument, the research base was not large enough to justify placing confidence in its clinical use. As was stated, "reporting inconsistencies, inaccuracies, and omissions that pervade the LNNB literature cause serious concern and tend to overshadow the positive results that have been reported" (Stambrook, 1983, p. 265). Although there is strong opinion to the contrary (Adams, 1984), like others (e.g., Sears et al., 1984; Shelly & Goldstein, 1983; Spiers, 1984; Stanley & Howe, 1983), my position is that carefully planned and well executed research is needed to replicate the major LNNB validation studies and to extend them so as to fully define the limits of the battery's clinical utility (Stambrook, 1983, 1985). This present study, together with the results reported by Sears et al. (1984), provide independently collected data that the LNNB can, based on the use of objective clinical rules, lateralize cerebral dysfuntion with hit rates that range from 70-80%, a more modest degree of accuracy as compared to the 94% hit rate reported by Golden, Moses, Fishburne, et al. (1981). This study also independently replicated the Osmon et al. (1979) finding regarding the ability of the LNNB to be sensitive to brain-behavior

relationships in discriminating right- and left-hemisphere-
damaged subjects. As well, the LNNB was found to be able to
can classify subjects as to laterality of lesion at levels
that did not differ from those achieved with the HRNTB.

However, the kind of validation study such as was
reported here (and that pervades much of the literature on
neuropsychological assessment) designed to examine
criterion-related (diagnostic) validity, provides only a
minimal, low level type of validation for the utility of an
instrument that is currently reported "to assess a broad
range of neuropsychological functions" (Western
Psychological Services, 1984, p. 82). While
neuropsychological assessment continues to play an important
role in neurodiagnosis (Wedding & Gudeman, 1980; Bigler &
Steinman, 1981), its major role is to accurately describe
brain-behavior relationships and to provide accurate
statements regarding the behavioral consequences of lesions
for treatment planning, rehabilitation counselling, and
litigation (Statz & Fletcher, 1981; Wedding & Gudeman,
1980). The implication here is that while questions
regarding a battery's ability to detect the presence of
brain damage, to provide differential diagnosis between
brain damage and psychiatric disorders, and to lateralize
and localize brain lesions are important, the instrument's
clinical merit is dependent upon its ability to reflect the
behavioral effects of brain functioning.

It is in this vein that the continuing controversy regarding the content validity of the LNNB is important. Simply, if a test instrument does not adequately sample the target domain(s) of behavior, as has been suggested (Appendix 1; Delis & Kaplan, 1983; Spiers, 1981), regardless of its ease of administration or its ability to diagnose, lateralize, and localize brain damage, it is of limited use in measuring neuropsychological functions. Thus, despite the demonstration here that the LNNB can lateralize brain damage, this is a necessary but minor demonstration in the determination and assessment of the battery's usefulness in clinical deployment.

What should be said of the objective clinical rules? The current finding that only 67% of the subjects could be classified as brain damaged despite that fact that all 100% had unequivocal neurological evidence of brain damage is disturbing. While appropriate cautions regarding "cook book rules" (p. 618) were provided in the original paper published in a professional journal (Golden, Moses, Graber, & Berg, 1981), no such caution is present in the test manual (Golden, Hammeke, & Purisch, 1980) other than an appeal that individual test items themselves be examined. I would submit that having a so-called "objective clinical rule" that is purported to diagnose brain damage, and is strongly endorsed in the test manual as a first step in the objective interpretation of the battery, contributes significantly to

the possible misuse of the instrument by those who would,

without questioning it, use such a procedure. This concern

is similar in nature to concerns raised by Delis and Kaplan

(1983) regarding the content, and then construct, validity

of the LNNB summary scales and their possible misuse. They

argue that because of low summary scale content validity,

achievement on any given "scale may be erroneously

associated with the integrity [or impairment] of the

designated cognitive or motor function" (p. 396). Spiers

(1984)has gone so far as to state:

> The LNNB is alluring because of Luria's reputation
> and the manner in which it purports to provide a
> standardized profile of neuropsychological
> functions. Unfortunately it has been marketed to,
> and most enthusiastically adopted by, precisely
> those clinicians who are least qualified to
> evaluate independently and objectively whether the
> LNNB is an adequate, reliable or valid
> neuropsychological assessment instrument. (p. 551)

Although Adams (1984) has argued that the LNNB "may be

removed from clinical practice without invoking an

obligation on the part of neuropsychology" (p. 455), my

position has been and remains that well formulated and

executed research is needed from independent laboratories

such that the checks and balances inherent in the peer-

reviewed scientific enterprise can, over time, serve to

fully delimit the clinical utility of the LNNB. Efforts such

as the present study and those reported by Delis and Kaplan

(1982), Sears et al. (1984), and Stanley and Howe (1983)

provide models for this necessary research endeavor.

REFERENCES

Adams, K. M. 1980. An end of innocence for behavioral neurology? Adams replys. Journal of Consulting and Clinical Psychology, 48(4), 522-524. (a)

Adams, K. M. 1980. In search of Luria's battery: A false start. Journal of Consulting and Clinical Psychology, 48(4), 511-516. (b)

Adams, K. M. 1984. Luria left in the lurch: Unfulfilled promises are not valid test. Journal of Clinical Neuropsychology, 6(4), 455-458.

Bigler, E. D., & Steinman, D. R. (1981). Neuropsychology and computerized axial tomography: Further comments. Professional Psychology, 12(2), 195-197.

Boll, T. J. (1981). The Halstead-Reitan Neuropsychology Battery. In S. B. Filskov & T. J. Boll (eds.), Handbook of Clinical Neuropsychology. New York: John Wiley.

Bryant, E. T., Maruish, M. E., Sawicki, R. F., & Golden, C. J. (1984). Validity of the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 52 (3), 445-448.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Craig, P. L. (1979). Neuropsychological assessment in public psychiatric hospitals: The current state of practice. Clinical Neuropsychology, 1(4), 1-7.

Crosson, B. & Warren, R. L. (1982). Use of the Luria-Nebraska Neuropsychological Battery in aphasia. A conceptual critique. Journal of Consulting and Clinical Psychology, 50(1), 22-31.

Delis, D. C. & Kaplan, E. (1982). The assessment of aphasia with the Luria-Nebraska Neuropsychological Battery: A case critique. Journal of Consulting and Clinical Psychology, 50(1), 32-39.

Delis, D. C., & Kaplan, E. (1983). Hazards of a standardized neuropsychological test with low content validity: Comment on the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 51(3),396-398.

Dodrill, C. B. (1979). Neuropsychology of epilepsy. In S. B. Filskov & T. J. Boll (eds.), Handbook of clinical neuropsychology. New York: Wiley.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72, 323-327.

Fletcher, J. M., Rice, W. J., & Ray, R. M. (1978). Linear discriminative function analysis in neuropsychological research: Some uses and abuses. Cortex, 14, 564-577.

Goldberg, A. L., & McNamara, K. M. (1984). Internship training in clinical neuropsychology. Professional Psychology: Research and Practise, 15(4), 509-514.

Golden, C. J. (1977). Validity of the Halstead-Reitan Neuropsychological Battery in a mixed psychiatric and brain injured population. Journal of Consulting and Clinical Psychology, 45(6), 1043-1051.

Golden, C. J. (1980). In reply to Adams "In search of Luria's battery: A false start". Journal of Consulting and Clinical Psychology, 48(4), 517-521.

Golden, C. J. (1981). The Luria-Nebraska Neuropsychological Battery: Theory and research. In P. McReynolds (ed.), Advances in psychological assessment, (Vol. 5). San Francisco, Jossey-Bass. (a)

Golden, C. J. (1981). A standardized version of Luria's neuropsychological tests. In S. B. Filskov & T. J. Boll (eds.), Handbook of clinical neuropsychology. New York: Wiley. (b)

Golden, C. J., Hammeke, T. A., & Purisch, A. D. (1978). Diagnostic validity of a standarized neuropsychological battery derived from Luria's neuropsychological tests. Journal of Consulting and Clinical Psychology, 46, 1258-1265.

Golden, C. J., Hammeke, T. A., & Purisch, A. D. (1980). The Luria Nebraska Neuropsychological Battery Manual. Los Angeles: Western Psychological Services.

Golden, C. J., Kane, R., Sweet, J., Moses, J. A., Cardellino, J. P., Templeton, R., Vincente, P., & Graber, B. (1981). Relationship of the Halstead-Reitan Neuropsychological Battery to the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 49(3), 410-417.

Golden, C. J., & Kuperman, S. K. (1980). Graduate training in clinical neuropsychology. Professional Psychology, 11, 55-63.

Golden, C. J., Moses, J. A., Fishburne, F. J., Engum, E., Lewis, E. P., Wisniewski, A. M., Conley, F. K., Berg, R. A., & Graber, B. (1981). Cross-validation of the Luria-Nebraska Neuropsychological Battery for the presence, lateralization, and localization of brain damage. Journal of Consulting and Clinical Psychology, 49(4), 491-507.

Golden, C. J., Moses, J. A., Graber, B., & Berg, R. (1981). Objective clinical rules for interpreting the Luria-Nebraska Neuropsychological Battery: Derivation, effectiveness, and validation. Journal of Consulting and Clinical Psychology, 49(4), 616-618.

Goldstein, G. (1974). The use of clinical neuropsychological methods in the lateralization of brain lesions. In S. J. Dimond & J. G. Beaumont (eds.), Hemispheric function in the human brain. New York Wiley.

Goldstein, G., & Shelley, C. H. (1973). Univariate and multivariate analysis in neuropsychological test assessment of lateralized brain damage. Cortex, 9, 205-216.

Goldstein, G., & Shelly, C. (1984). Discriminative validity of various intelligence and neuropsychological tests. Journal of Consulting and Clinical Psychology, 52(3), 383-389.

Halstead, W. C. (1947). Brain and Intelligence. Chicago: University of Chicago Press.

Hammeke, T. A., Golden, C. J., & Purisch, A. D. (1978). A standardized short and comprehensive neuropsychological test battery based on the Luria-Nebraska Evaluation. International Journal of Neuroscience, 8(3), 135-141.

Hartlage, L. C., & Telzrow, C. F. (1980). The practice of clinical neuropsychology in the U.S. Clinical Neuropsychology, 4, 200-202.

Hays, W. L. (1973). Statistics for the social sciences (2nd ed.).New York: Holt, Rinehart, & Winston.

Heaton, R. K., & Crowley, T. J. (1979). Effects of psychiatric disorders and their somatic treatments on neuropsychological test results. In S. B. Filskov & T. J. Boll (eds.), Handbook of Clinical Neuropsychology. New York Wiley.

Hevern, V. W. (1980). Recent validity studies of the Halstead-Reitan approach to neuropsychological assessment: A critical review. Clinical Neuropsychology, 2(2), 49-61.

Huberty, C. J. (1975). Discriminant analysis. Review of Educational Research, 45(4), 543-598.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95(1), 156-171.

Kane, R. L., Parsons, O. A., & Goldstein, G. (1985). Statistical relationships and discriminative accuracy of the Halstead-Reitan, Luria-Nebraska, and Wechsler IQ scores in the identification of brain damage. Journal of Clinical and Experimental Neuropsychology, 7(3), 211-223.

Kane, R. L., Sweet, J. J., Golden, C. J., Parsons, O. A., & Moses, J. A. (1981). Comparative diagnostic accuracy of the Halstead-Reitan and standardized Luria-Nebraska Neuropsychological Batteries in a mixed psychiatric and brain-damaged population. Journal of Consulting and Clinical Psychology, 49(3), 484-485.

Kerlinger, F. N., & Pedhazur, C. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Klove, H. (1974). Validity studies in adult neuropsychology. In R. M. Reitan & L. A. Davison (eds.), Clinical Neuropsychology, New York: Wiley.

Lachin, J. M., & Schachter, J. (1974). On stepwise discriminant analyses applied to physiological data. Psychophysiology, 11, 703-709.

Lezak, M. D. (1976). Neuropsychological assessment. New York: Oxford.

Lishman, W. A. (1978). Organic psychiatry: The psychological consequences of cerebral disorders. Oxford: Blackwell.

Malloy, R. R., & Webster, J. S. (1981). Detecting mild brain impairment using the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 49(5), 768-770.

Matarazzo, J. D. (1976). Wechsler's measurement and appraisal of adult intelligence. Baltimore: Williams & Wilkins.

McKay, S. E., & Golden, C. J. (1979). Empirical derivation of experimental scales for localizing brain lesions using the Luria-Nebraska Neuropsychological Battery. Clinical Neuropsychology, 1(4), 19-23. (a)

McKay, S. E., & Golden, C. J. (1979). Empirical derivation of neuropsychological scales for the lateralization of brain damage using the Luria-Nebraska Neuropsychological Battery. Clinical Neuropsychology, 1(2), 1-5. (b)

Meier, M. J. (1981). Educatin for competency assurance in human neuropsychology: Antecedents, models, and directions. In S. B. Filskov & T. J. Boll (eds.), Handbook of Clinical Neuropsychology. New York Wiley.

Moses, J. A., Golden, C. J., Berger, P. A., & Wisniewski, A. M. (1981). Neuropsychological deficits in early, middle, and late stage Huntington's disease as measured by the Luria-Nebraska Neuropsychological Battery. International Journal of Neuroscience. 14(1\2), 95-100.

Neuropsychological Associates of California. (1985). [Advertisement for wprkshops on the Luria-Nebraska Neuropsychological Battery]. APA Monitor, 16, 39.

Noonberg, A. R., & Page, H. A. (1982). Graduate neuropsychology training: A later look. Professional Psychology, 13(2), 252-257.

Osmon, D. E., Golden, C. J., Purisch, H. D., Hammeke, T. A., & Blume, H. G. (1979). The use of a standardized battery of Luria's tests in the diagnosis of lateralized cerebral dysfunction. International Journal of Neuroscience, 9(1), 1-9.

Parsons, O. A., & Prigatano, G. P. (1978). Methodological considerations in clinical neuropsychological research. Journal of Consulting and Clinical Psychology, 46(4), 608-619.

Precision People Inc.. (1985). [Advertisement for computer software]. APA Monitor, 16, 53.

Reitan, R. M. (1962). Psychological Deficit. Annual Review of Psychology. 13, 415-444.

Reitan, R. M. (1966). A research program on the effects of brain lesions in human beings. In N. R. Ellis (ed.), International Review of Research in Mental Retardation: Vol. 1. Academic Press, New York. (a)

Reitan, R. M. (1966). Diagnostic inferences of brain lesions based on psychological test results. Canadian Psychologist, 7a(4),368-383.

Reitan, R. M. (1975). Assessment of brain-behavior relationships. In P. McReynolds (ed.), Advances in psychological assessment. San Fransisco: Jossey-Bass.

Reitan, R. M. (1964). Psychological deficits resulting from cerebral lesions in man. In J. M. Warren & K. Akert (eds.), The frontal granular cortex and behavior. New York: McGraw-Hill.

Reitan, R. M., & Davison, L. A. (1974). Clinical Neuropsychology. New York: Wiley.

Russell, E. W., Neuringer, C., & Goldstein, G. (1970). Assessment of brain damage. New York: Wiley.

Satz, P., & Fletcher, J. M. (1981). Emergent trends in neuropsychology. Journal of Consulting and Clinical Psychology, 49, 851-865.

Sawicki, R. F., & Golden, C. J. (1984). Examination of two decision rules for the global interpretation of the Luria-Nebraska Neuropsychological Summary Profile. International Journal of Neuroscience, 22, 209-220.

Sears, J. D., Hirt, M. L., & Hall, R. W. (1984). A cross-validation of the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 52(2), 309-310.

Shelly, C., & Goldstein, G. (1983). Discrimination of chronic schizophrenia and brain damage with the Luria-Nebraska Battery: A partially successful replication. Clinical Neuropsychology, 5(2), 82-85.

Siegal, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Spiers, P. A. (1981). Have they come to praise Luria or bury him? The Luria-Nebraska Battery controversy. Journal of Consulting and Clinical Psychology, 49(3), 331-341.

Spiers, P. A. (1982). The Luria-Nebraska Neuropsychological Battery revisited: A theory in practice or just practicing? Journal of Consulting and Clinical Psychology, 50(2), 301-306.

Spiers, P. A. (1984). What more can I say ? In reply to Hutchinson, one last comment from Spiers. Journal of Consulting and Clinical Psychology, 52(4), 546-552.

SPSS Inc.. (1983). SPSSX users guide. New York: Mcgraw-Hill.

Stambrook, M. (1983). The Luria-Nebraska Neuropsychological Battery: A promise that may be fulfilled. Journal of Clinical Neuropsychology, 5(3), 247-269.

Stambrook, M. (1985). What to do ? Neuropsychology and obligation in the Luria-Nebraska controversy. Unpublished manuscript.

Stanley, B., & Howe, J. G. (1983). Identification of multiple sclerosis using double discrimination scales derived from the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 51(3), 420-423.

Tatsuoka, M. M. (1971).Multivariate analysis. New York Wiley.

Tatsuoka, M. M. (1974). Classification procedures. Champlaign, Ill.: IPAT.

Wedding, D., & Gudeman, H. (1980). Implications of computerized axial tomography for clinical neuropsychology. Professional Psychology, 11, 31-35.

Western Psychological Services. (1984). 1985-86 Catalog. Los Angles: Manson Western Corporation.

Wheeler, L., Burke, C. J., & Reitan, R. M. (1963). An application of discriminant functions to the problem of predicting brain damage using behavioral variables. Perceptual and Motor Skills, 16, 417-440.

Wheeler, L., & Reitan, R. M. (1963). Discriminant functions applied to the problem of predicting cerebral damage from behavioral tests: A cross-validation study. Perceptual and Motor Skills, 16, 681-701.

Appendix I

The Luria Nebraska Neuropsychological Battery:

A Promise that May Be Partly Fulfilled

The Luria-Nebraska Neuropsychological Battery:

A Promise That <u>May</u> Be Partly Fulfilled

Over the past seven years Golden and his associates have reported the development a standardized version of Luria's Neuropsychological Investigation. This instrument, currently called the Luria-Nebraska Neuropsychological Battery (Golden, Hammeke, & Purisch, 1980; Golden, Purisch, & Hammeke, 1979), is believed to be capable of providing a comprehensive and extensive evaluation of all major neuropsychological functions such that the battery can be used effectively for both diagnosis and rehabilitation planning (Golden, Hammeke, & Purisch, 1980; University of Nebraska Press, 1979). Further, in the recently published manual (Golden, Hammeke, & Purisch, 1980), it is stated that "the Luria-Nebraska can replace the much more extensive batteries commonly used in American neuropsychology that may take two to three times as long to administer and yet not yield any more extensive information on the patient's performance or disorder" (p. 13).

Although Golden and his team have provided an impressive array of data which they suggest unequivocally demonstrates the efficacy of the Luria-Nebraska Neuropsychological Battery (LNNB) in detecting the presence, lateralization, and localization of brain damage (for reviews see Golden, 1979, 1981a, 1981b; Golden Hammeke, & Purisch, 1980; and see below), serious concerns have arisen that question many of the claims that have been made concerning the battery's clinical utility (Adams, 1980a, 1980b; Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, 1982,

Note 1). That the literature on the LNNB has generated vigorous controversy is without doubt and the tenor of this debate is amply demonstrated by Adams (1980a, 1980b) who, in a thinly veiled critique, questions Golden's veracity as a researcher. It has been suggested that the LNNB was marketed and promoted as a clinical, diagnostic instrument prematurely, and that a consideration of the concerns that have been voiced should result in caution being taken by those who would use the battery in clinical settings (Adams, 1980a, 1980b; Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, 1982, Note 1).

It is the purpose of this paper to critically review the literature on the LNNB in terms of the instrument's development, reliability, and validity. In doing so, this paper will serve to highlight areas of strengths and weaknesses of the battery, as well as provide an indication of where future research efforts may most fruitfully be carried out.

THEORETICAL ANTECEDENTS

The LNNB was designed and developed to be a standardization and operationalization of Luria's (1966, 1980) neuropsychological investigation technique. As such, the battery was based on Luria's (1964, 1970, 1973, 1975, 1980) complex and multifaceted conceptualization of the functional organization of the central nervous system. The most important aspect of Luria's theoretical structure for the LNNB is the concept of the functional system (Golden, Ariel, McKay, Wilkening, Wolf, MacInnes, 1982). In rejecting both the strict localization and equipotential views of cortical functioning, Luria posited that overt behavior is a

result of the coordinated activity of many cortical and subcortical
structures, each of which makes a highly specific contribution to
the performance of the behavior. In Luria's framework, a func-
tional system for a particular behavior is seen as the totality
of the anatomical components whose functioning results in the
completion of the behavior. An important part of this func-
tional system idea is the notion that not only can there be many
different functional systems underlying a given behavior, but
also, that areas of the brain participate in many different func-
tional systems.

The functional system concept has far-reaching implications
for neuropsychological assessment. Since behavior, especially
complex behavior such as speaking or writing, results from the
interaction of many zones of the brain, it is not possible to
topographically localize cerebral lesions when a symptom of
impaired behavior is demonstrated. In Luria's framework, the
goal of neuropsychological assessment is not only to describe
"the symptoms of disturbances of higher cortical functions, but
also [the] qualification of the defects and an analysis of the
factors underlying these behavioral defects" (Luria, 1975, p. 7).
The crux of Luria's neuropsychological investigation is this
qualification of the symptom. Because the functional system
underlying a given behavior may be disrupted in many ways, by
lesions in any or all of the cooperating zones of the brain, it
is necessary to qualitatively analyze the nature of the symptoms
to determine which component(s) of the functional system is
impaired. A cortical lesion may cause the collapse of the

functional system for a behavior but, because each of the many zones of the brain makes a highly specific contribution to the overall system, the qualitative analysis allows the delineation of the specific impaired link in the chain of contributing processes. An assumption of this qualification of the symptom approach is that other behaviors which utilize the same impaired link will also be disrupted, while behaviors which do not involve this link will remain intact.

The role of the neurodiagnostician in this approach is to systematically examine the quality of performance on a variety of tasks, each requiring the operation of functional systems dependent on different sets of cortical zones, so that the precise defect responsible for the disruption of the functional system(s) can be isolated. This can be clarified by using the simplistic but concrete illustration provided by Golden, Ariel, McKay, et al. (1982). If three tasks are performed in a manner suggesting impairment, and if performance on these tasks is thought to be a result of functional systems which involve skills A, B, and C, skills B, F, and G, and skills B, D, and E, respectively, it could be hypothesized that a deficit in skill B underlies the impaired performance on all three tasks. It would then be possible to attempt to make some prediction concerning the location of the lesion in the brain which results in impairment of performance in behaviors dependent on skill B.

It is clear from the above example that the form of analysis that derives from a qualification of the symptom approach is highly dependent on the ability of Luria's theory to accurately

reflect the component skills that make up a behavior, and to suggest the anatomical locus for each molecular skill. Although Luria's theoretical position reflects over 40 years of clinical and experimental work in the Soviet Union and many laboratories around the world, it is clear from his writings (primarily Luria, 1973, 1980) that the theory is an approximation which has been continually evolving as new data came to light. The practical consequence of this is that, while the qualification of the symptom approach to neuropsychological assessment may remain unchanged, constant revision of the hypothesized component skills for each functional system is a necessity as knowledge increases about the functional organization of the brain.

In actual practise Luria's neuropsychological investigation is an extension of the classical neurological examination. The investigation,which has been most comprehensively described in Luria's Higher Cortical Functions in Man (1966, 1980), consists of two separate, but conceptually related parts, the preliminary interview and the detailed, selective examination. The preliminary interview involves obtaining a history of the present condition of the patient and conducting an initial series of very short tests which, in a very general sense, assess various aspects of the patient's mental activity and the integrity of all major neuropsychological processes. Since the complexity of the questions and tests is such that they do not give intact patients difficulty, Luria contends that the examiner must attend not only to the accuracy of the responses, but also, to their form or quality. The purpose of the preliminary interview is to

detect the presence of areas of impairment (identification of the symptom(s)) so that they may be subjected to more detailed and molecular analysis in the second, selective phase (qualification of the symptom). As Luria (1980) states:

> In contrast, therefore, to the preliminary stage, the second stage of the investigation must be strictly individualized. It must be built up, firstly, on the basis of results obtained in the first stage and, secondly, by taking into account the facts that are obtained in the second phase of the investigation itself. It is clear, therefore, that the second stage of the neuropsychological investigation is the more complex, yields richer results, and calls for greater flexibility in the conduct of the experiments [i.e. tests]. (p. 395)

The primary objective of the second, highly individualized, qualitative phase is to identify and isolate the fundamental defect(s) that underlies the collapse of one or more functional systems. Luria believes that it is only through this approach that it is possible to suggest a local lesion that is responsible for the observed symptom.

Luria (1966, 1980) has described, in a fairly comprehensive manner, a sample of the many tasks or tests that may be administered to patients suspected of suffering from brain lesions. The clear implication is that what is important is not the adherence to any particular set of tasks, but that the diagnostician, limited only by his or her creativity and imagination, use tasks

that facilitate the isolation of the factors responsible for impaired functioning. Since Luria (1966, 1980) did not present the various tasks used in assessment in a manner conducive for actual administration, Christensen (1975) compiled the items into a framework to assist the clinician in administration and "to ensure that the process of investigation would be as thorough and exhaustive as it was designed to be" (p. 9). While Luria had an opportunity to examine and correct the Christensen manuscript, he was less than enthusiastic in his endorsement stating, "of course it is a vulgarization - but I have always wanted someone to do what you have done" (Christensen, 1975).

Christensen (1975) shares with Luria (1960, 1980) the commitment to the concept of the functional system, and the qualification of the symptom approach to the neuropsychological examination. The primary difference between Luria's and Christensen's work is that Christensen has taken the items that Luria presented as being only a sample, admittedly a select sample, of those that could be used, and in presenting them as is done, implies that these items are the primary ones that Luria would have used in his neuropsychological examination.

Christensen's (1975) adaption of Luria's neuropsychological investigation consists of information on the preliminary interview and 253 items that are organized, primarily for didactic purposes, into 10 major content areas. These areas consist of items involving motor functions, acoustico-motor organization, higher cutaneous and kinesthetic functions, higher visual functions, impressive (receptive) speech, expressive speech, writing and

reading, arithmetical skills, mnestic (memory) processes, and intellectual processes. An important point that has not been attended to in the development of the LNNB, or in the subsequent research literature on the battery, is that Luria's investigation is considered to primarily evaluate the functions of the left hemisphere (Christensen, 1975, p. 11; Luria, 1980, p. 587).

Luria (1966, 1980) and Christensen (1975) strongly oppose the standardized neurodiagnostic assessment techniques embodied in the Halstead-Reitan Neuropsychological Test Battery (HRNTB), which is considered to be one of the most widely used neuropsychological assessment instruments in North America (Craig, 1979; Golden & Kuperman, 1980; Hartlage & Telzrow, 1980; Reitan & Davison, 1974). They posit that great variability and flexibility are required of the clinician in conducting an examination and that "any attempt to apply static standardized experimental psychological techniques must be entirely discouraged. Only if this condition is satisfied (which, to do so, requires great experience) will the clinical psychological investigation prove effective" (Luria, 1980, p. 392).

It has been stated (Adams, 1980b; Luria, 1980; Luria & Majovski, 1977; Reitan, 1976a, 1976b) that there are major differences between the basic approaches used in North American and Soviet clinical neuropsychology. In the main, the American approach, represented by the HRNTB, has been typified as being standardized, quantitative, and atheoretical, while the Soviet approach, represented by Luria's procedures, is individualized, qualitative, and theoretically based. Although this generalization

is partially true, it misrepresents the actual nature of North
American clinical neuropsychology. This has been most recently
pointed out by Satz and Fletcher (1981) who suggest that is is
incorrect and unfair to characterize North American clinical
neuropsychology by using the HRNTB. There is a sufficient
amount of diversity among approaches used in American clinical
neuropsychology to render any generalization based on just one
approach virtually meaningless. As Satz and Fletcher (1981)
suggest, there are many approaches in American clinical neuro-
psychology that rely heavily on theoretical models. The assess-
ment of language disorders (Goodglass & Blumstein, 1973),
memory disorders (Butters & Cermak, 1980), movement disorders
(Heilman, 1979), and the long-term effects of closed head
injuries (Levin, Grossman, Rose, & Teasdale, 1979), for example,
use theoretical models in an effort to describe the functional
deficits that occur when the brain is damaged. Heilman and
Valenstein (1979) provide other examples of the impact theory
has on North American clinical neuropsychology. In a similar way,
Goodglass and Kaplan (1979) and Lezak (1976) have provided North
American clinical neuropsychology with approaches that are
neither standardized nor necessarily quantitative. In these
approaches, neuropsychological assessment is individualized to
meet the needs of each patient and referral question, and
includes an integration of both qualitative and quantitative data
in an effort to best describe the clinical picture presented.

Thus, while the HRNTB differs in many respects from Luria's
procedures, it is not accurate to suggest that this reflects a

difference between North American and Soviet clinical neuropsychology. The apparent differences between North American and Soviet approaches fade when one considers that some of the emergent trends in North American neuropsychology include the rapid development of conceptual models that attempt to account for the functional organization of the brain in sickness and in health (Galin, 1976; Hellige, Cox & Litvac, 1979; Moscovitch & Klein, 1980; Satz and Fletcher, 1981; Sergent and Bindra, 1981; Tucker, 1981), and the growing appreciation of the importance of systematic individual case studies (Parsons & Prigatano, 1978; for examples see Golden, Strider, Strider, Moore, & Gust, 1979; Newlin & Tramontana, 1980).

Luria's approach to neuropsychological assessment has several advantages over the HRNTB approach. Luria's procedures afford the opportunity for an extensive and systematic breakdown of complex behavior into component skills, while many of the individual tests of the HRNTB are so complex (e.g. Category Test) that it is difficult to determine which component skills are responsible for impaired performance (Golden, Hammeke, & Purisch, 1978; Luria, 1980). Although the use of Reitan's (1974) four methods of inference (level of performance, left-right differences, pathognomonic sign, pattern analysis) allow predictions to be made regarding the nature of the impairment and it's possible anatomical locus, the basic data are derived from complex tests which makes specification of specific skill deficits difficult. Because of the ability of Luria's procedures to identify specific skill deficits, it has been suggested that the

method yields information directly relevant to diagnosis and treatment planning (Hammeke, Golden, Purisch, 1978; Luria, 1980). The other major advantages of Luria's procedures are that they require a relatively short administration time (under 3 hours), little, if any, equipment, and may be administered at bedside, while the HRNTB typically requires 6 to 8 hours administration time and a laboratory setting with expensive equipment.

Despite these advantages, Luria's procedures have been harshly criticized (Reitan, 1976a, 1976b). Reitan (1976b) contends that the individualized, qualitative approach used by Luria,

> is in direct opposition to the development and
> use of standardized procedures which offer
> quantitative results. . . This type of approach,
> in brief, shows little respect for the difficulties
> implicit in devising an experiment which would have
> general significance. . . One can conclude that
> Luria's methods represent the ultimate in disregard
> to the time-honored concept of cross-validation
> inasmuch as Luria's method itself precludes the
> prospect of objective cross-validation.(p. 199)

There's no doubt that much of what Luria claims for his procedures must be taken on faith because, as of yet, there has been no systematic presentation of his data that would permit critical examination of his assertions regarding the efficacy of the techniques or the adequacy of the theoretical structure (Adams, 1980b).

DEVELOPMENT OF THE LNNB

The stated goal in the development of the standardized version of Luria's Neuropsychological Investigation was to create a battery which would have the advantages of the qualification of the symptom approach to assessment as well as having the advantages of an approach which is standardized and quantitative (Golden, 1981a, 1981b; Golden, Ariel, McKay et al., 1982; Golden Hammeke, & Purisch, 1978; Hammeke et al., 1978). The battery "was intended to provide a basis for quick and reliable collection of empirical data, while allowing for qualitative analysis as exemplified by Luria's work" (Golden, Ariel, Moses, Wilkening, McKay, MacInnes, 1982, p. 40-41).

Item Selection

The initial items for the LNNB were derived exclusively from Christensen's (1975) adaption of Luria's procedures (Golden, Hammeke, & Purisch, 1978). There is almost a one-to-one correspondence between the LNNB and Christensen's items. Golden and his colleagues (Golden, Hammeke, & Purisch, 1978; Hammeke et al., 1978), however, added items so that there would be an examination of the right and left sides of the body for an analysis of tactile and motor functions. Items were excluded from the pool when it was discovered that normal subjects were unable to perform them, and when standardized administrative and scoring techniques could not be developed.

This procedure yielded a pool of 282 items (Hammeke, Note 2) which were then administered to 50 neurologically intact medical patients and 50 patients with mixed neurological diagnoses

(Golden, Hammeke, & Purisch, 1978; Hammeke, Note 2). Based on
the ability to discriminate between the two groups of patients,
269 items were retained and are the items that are currently in
use (Golden, Hammeke, & Purisch, 1980). The items are thought
to differ in terms of their familiarity, complexity, intellectual
demands, verbal demands, response methods, input sources, atten-
tion and concentration demands, and integration demands (Golden,
Ariel, Moses et al., 1982). These items are organized into 11
"primary" behavioral summary scales based almost exclusively on
the organizational framework provided by Christensen (1975).
Although Christensen (1975) maintained that her organization was
artificial with respect to neuropsychological functioning, and
was presented for didactic reasons, Golden (Golden, Hammeke, &
Purisch, 1978, 1980) used Christensen's descriptors to label the
scales. This resulted in the following summary scales: Motor
Functions, Rhythm, Tactile Functions, Visual Functions, Receptive
Speech, Expressive Speech, Writing, Reading, Arithmetic, Memory,
and Intellectual Processes. Examples of the types of items
that appear on each summary scale can be found in Golden (1981a,
1981b) and Golden, Hammeke, and Purisch (1978). Since the
scale structure is considered to be flexible (Golden, Ariel, Moses
et al., 1982), numerous additional scales have been developed by
recombining the 269 items in different ways, in a manner similar
to the way additional MMPI scales can be developed (Butcher &
Tellegen, 1978). While this procedure has resulted in the crea-
tion of empirically derived lateralization (McKay & Golden, 1979b),
localization (McKay & Golden, 1979a), and factor analytic scales

(McKay & Golden, 1981a), there are three scales that are basic to the LNNB. These are the Left and Right Hemisphere Scales and the Pathognomonic scale. The Left and Right Hemisphere Scales are made up of items from the Tactile and Motor Scales that assess the function of the contalateral arm and hand, while the Pathognomonic Scale is made up of items that best discriminate neurological and control patients (Golden, Hammeke, & Purisch, 1980).

Administration

Despite the fact that the intent in developing the LNNB was to provide a standardized method of performing Luria's investigation, the administration instructions for the battery (Golden, Hammeke, & Purisch, 1980) are a curious blend of admonishments to maintain a standardized format and invitations to improvise and test the limits (Adams, 1980a). The administration manual (Golden, Hammeke, & Purisch, 1980) does provide detailed instructions for each item, but also suggests that items may be modified to meet the needs of the individual patient and referral question, if the original intent of the item is preserved. However, it is difficult to know what the authors mean when they state that the "standardized instructions are flexible" (Golden, Hammeke, & Purisch, 1980, p. 13).

While the intent of introducing some degree of flexibility is useful because it allows the use of the battery for cases which may have been traditionally untestable, it is unknown how the flexibility would effect the scoring of the items since, presumably, the scoring system is based on a standardized

administration. Recently, Golden and his associates (Golden, 1981a, Golden, Ariel, Moses et al., 1982) have indicated that a new form of the battery is currently being developed which will more fully indicate the options for administration. As Adams (1980b) suggests, however, the concept of a standardized Luria "seems to be a logical impossibility. . . The need to be consistent, rigorous and public in the application and development of proto-cols seems antithetical to the approach that Luria described" (p. 514).

According to the manual (Golden, Hammeke, & Purisch, 1980), the battery takes from 1-1/2 to 2-1/2 hours to administer and requires the use of six published stimulus cards, a 15 minute prerecorded audio tape (cards and tape available from Western Psychological Services), a pocket comb, a rubber band, a paper clip, a compass, an erasor, a key, a pin, a quarter, a ruler, a blindfold, and a stopwatch. The battery may be given at bed-side and, although the battery is designed to be given in it's entirety, the length of sessions may be manipulated depending on the patient's tolerance. While the battery is presently designed for patient's 15 years of age and older, a children's version is currently being developed (Golden, 1981a; Wilkening, Golden, MacInnes, Plaistead, & Herman, Note 3).

Scoring

Items are scored along dimensions that include accuracy and adequacy of response, number of errors, time for performance, number of trials to correct performance, and number of trials completed. These dimensions are thought to reflect the

qualitative factor that the item theoretically is assessing.
The manual (Golden, Hammeke, & Purisch, 1980) clearly specifies
how the scoring procedure is to be applied for each item. After
the raw scores for each item are generated, they are converted
into scale scores of 0, 1, and 2 (95 items are scored 0 or 2).
Initial formulas for converting the raw scores to scale scores
were based on cutoff points chosen to maximize the effectiveness
of each item to discriminate a group of 75 subjects as brain-
damaged or normal. A scale score of 0 was respresentative of
the performance of normals, while a scale score of 2 represented
the performance characteristic of brain-damaged subjects. A
scale score of 1 was intermediate or borderline.

While an extremely brief summary of the scale score deriva-
tion has appeared in many of Golden's publications (e.g., Golden,
1981a, 1981b; Golden, Hammeke, & Purisch, 1978, 1980), it is
unfortunate that the work has not been published either in more
detail, or in it's entirety. Since the scale score system pro-
vides the basic data for the battery in both research and clinical
interpretation (Golden, Ariel, Moses et al., 1982), it is
essential to evaluate it's adequacy. Of prime importance is the
nature of the group on which the scale score system was derived.
The only data that has been provided is that the group either
consisted of 75 subjects classified as normal, schizophrenic,
and brain-damaged (Golden, 1981a; Purisch, Golden, & Hammeke, 1978)
or that it consisted of 37 normal subjects and 38 brain-damaged
subjects (Golden, Hammeke, & Purisch, 1978, 1980).

It has been reported that the scale score conversions have

been cross-validated and that only minor adjustments were needed
on 16 of the 269 items. As with the data on the initial conver-
sions, the cross-validation data has only been briefly presented
in secondary sources, and is particularly confusing because the
sample size changes from one report to the next (n = 200, Golden,
Moses, Fishburne, Engum, Lewis, Wisniewski, Conley, Berg, &
Graber, 1981; n = 210, Golden, 1981a; n = 233, Golden, 1981b).
Other than the fact that it is reported that the sample consists
of both normal and neurological subjects, no other data is pre-
sented. Since the composition and demographic characteristics
of the samples used in the derivation and cross-validation of the
raw to scale score conversions are unknown, the generalizability
of the conversions for clinical and research work is uncertain.
The effects of brain damage differ dramatically depending on a
great many variables (e.g., premorbid condition, etiology, locus,
chronicity, severity, extent, etc., Reitan & Davison, 1974).
Thus, to apply data gathered from samples for which little is
known, to samples that may differ in many ways, is a hazardous
procedure.

After the scale scores for the items are determined, they
are summed within each of the 11 primary behavioral scales and
the Left Hemisphere, Right Hemisphere, and Pathognomonic scales
to yield 14 summary scale scores (high scores indicate greater
impairment). These summary scale scores are then transformed
into T-scores (standardized scores with a mean of 50 and a
standard deviation of 10) based on the means and standard devia-
tions obtained from a sample of 50 medical patients hospitalized

for conditions not effecting the brain (Hammeke et al., 1978).

Although the T-score transformation formulas have been cross-validated (Moses & Golden, 1979) in an additional sample of 50 medical control patients with only minimal differences occurring, as with the raw to scale score conversions, the representativeness of this data is not known. In fact, it is a mystery why the developers of the LNNB attempted to norm the instrument on subjects who were chosen, not because of their representativeness to the population of the non-neurologically impaired, but because they were the most appropriate comparison group in terms of assessing whether the battery could statistically discriminate brain-damaged from intact subjects. Since the brain-damaged subjects were hospitalized, neurologically intact subjects were chosen to partially control for the effects of the hospitalization experience.

Although the administration and scoring manual for the battery (Golden, Hammeke, & Purisch, 1980) does caution that the norms have been developed primarily on a Caucasian "lower to middle socioeconomic group", insufficient data are presented to allow users to make rational judgements regarding the usefulness of the norms as is deemed essential by the standards set forth by the American Psychological Association (1974). This deficit is demonstrated by Delis and Kaplan's (1982) interpretation of a LNNB examination involving a patient whose initial language was Spanish. In a rejoiner, Golden, Ariel, Moses, et al., (1982) criticize the interpretation stating that since the norms for the battery were derived on patient's whose first language was

English, they may not be applicable for patient's with other language backgrounds. This is representative of a serious problem because it is not known what other restrictions may be present that limit the use of the norms. Golden (1981a; Golden, Moses, Fishburne et al., 1981) has recently acknowledged this deficit by calling for extensive cross-validation work to "fully expand the test's normative base (Golden, 1981a, p. 231). It is unfortunate that this was not mentioned in the promotional statements made about the battery (University of Nebraska Press, 1979; Western Psychological Services, 1981), or considered before the test was marketed for clinical use.

The psychometric properties of the scale scores (0,1,2) and the summary scales have been questioned. Adams (1980b) and Spiers (1981) have suggested that the limited range of the scale scores reduces the sensitivity of the measures in describing behavior because current neuropsychological knowledge would permit greater precision. In rebuttal, Golden (1980; see also Golden, Hammeke, & Purisch, 1978) has claimed that other scale scoring schemes were investigated (e.g., a 0, 1, 2, 3 system and a 0, 1, 2, 3, 4 system) but they failed to add to the ability of the items in discriminating neurological from control groups. This argument, however, misses the point because the clinical utility of the items is not just determined by their ability to detect brain damage. It is also determined by the ability of the items to adequately describe behavior such that treatment and rehabilitation programs can be devised (Lezak, 1976; Luria, 1980; Parsons, 1970). The scoring system that most accurately reflects behavior

would be most amenable for this purpose. Further, the scale score system does not permit the identification of the precise component of behavior that is impaired (Spiers, 1981). Since individual items do not measure "pure" skills (Golden, Ariel, Moses et al., 1982), it is unclear as to what a particular scale score for an item would mean.

The summation of individual item scale scores to produce an overall score for the 14 summary scales has also been criticized. Russell (1980) has suggested that because the summary scales are heterogeneous with regard to the type of functions assessed (Golden, Hammeke, & Purisch, 1980), summing scale scores across items is a meaningless procedure. While it has been posited that each summary scale assesses a general skill area, named in the scale title (Golden, Ariel, McKay, et al., 1982), the heterogeneity has led Golden, Hammeke, and Purisch (1980) to stress that, in interpreting summary scale elevations, individual items must be examined to gain insight into the particular nature of impairment. Together, Russell's (1980) criticism and Golden, Hammeke, and Purisch's (1980) caution point to the limited value of the summary scales in neuropsychological assessment.

Russell (1980) has also suggested that the level of measurement inherent in the scale score system precludes the use of the act of summation. Here Russell (1980) is clearly in error because he submits that nominal scaling is used when, in fact, the scale score system is at an ordinal level of measurement (Stevens, 1951). While, in strict sense, the act of summation requires that interval level assumptions be met, it must be noted that the major

instrument for the assessment of adult intelligence, the Wechsler Adult Intelligence Scale, summates ordinal item scores to derive raw scores for subtests such as Comprehension, Similarities, and Vocabulary (see Matarazzo, 1976).

RELIABILITY

Scorer Reliability

The reliability or consistency with which the scoring criteria for each item may be applied was investigated as part of a dissertation carried out by Hammeke (Note 2) and reported by Golden, Hammeke, and Purisch (1978, 1980). From a sample of 50 neurological patients and 50 medical control patients, five subjects were randomly chosen for study. The LNNB was then administered by one examiner, in the presence of a second examiner, with each examiner independently scoring the performance. The percentage of agreement between the two examiners over the scale scores of the then 282 items on the battery ranged from 92% to 98%, with a mean of 95%. The correlations between the scores for each examiner ranged from .97 to .99 for the five subjects.

Although there are some inconsistencies among the reports of this investigation (i.e., number of pairs of examiners used and whether the correlations presented were calculated on raw or scale scores) that make evaluation of results difficult, the greatest shortcoming is that there is no indication of the composition of the sample. It is likely that it would be easier for examiners to agree on scoring if the protocols were unambiguous, as is likely if neurologically intact subjects were used. Given the very small sample size, it is unlikely that there would

be sufficient variability among responses to allow a good test of the adequacy of the scoring criteria. Since there has been some discussion of the ability of the scoring criteria to provide reliable scoring (Adams, 1980b), more data are required before Golden's (1980) claim that "the scoring criteria are highly reliable" (p. 517) can be accepted.

One important aspect of reliability that has not been investigated is the consistency with which the battery can be administered, given that the examiner is permitted considerable flexibility. This is a question of whether two examiners will generate the same data if they use their judgement as to how the items are to be administered.

Test-Retest Reliability

Using patients described as having chronic, static neurological impairment, Golden, Berg, and Graber (1980) reported test-retest correlations ranging from .77 (Right Hemisphere) to .96 (Arithmetic), with a mean of .88. The test-retest interval ranged from 10 to 489 days ($\underline{M}$ = 167 days). Partialling out the variance attributable to the test-retest interval resulted in essentially no change in the test-retest correlations suggesting that "the length of the test-retest interval had a negligible influence on the results" (Golden, Hammeke, & Purisch, 1980, p. 3).

While the magnitude of the test-retest correlations reported for the LNNB are roughly comparable to those reported for the HRNTB (Matarazzo, Matarazzo, Wiens, & Gallo, 1976), the finding that the length of the test-retest interval was not related to the size of the correlations is incongruous with psychometric

lore which holds that test-retest correlations progressively
decrease as the interval increases (Anastasi, 1976; Cronbach,
1970). Since the patients for this study were psychiatric
(increases instability, Matarazzo et al., 1976) as well as neuro-
logical (Golden, 1981a), and neither floor nor ceiling effects
were operative (see Table 2, Golden, Hammeke, & Purisch, 1980),
it seems wise to suggest that endorsement of these results await
independent replication.

Split-Half Reliability

Split-half reliability provides a measure of the degree of
consistency with regard to content sampling (Anastasi, 1976).
Using an odd-even split, Golden, Fross, and Graber (1981) report
that the split-half correlations for the summary scales of the
LNNB ranged from .89 (Memory) to .95 (Reading), with a mean of
.92. The magnitude of the correlations would seem to indicate
that the summary scales are quite homogeneous which is in direct
contrast to the suggestion (Golden, Hammeke, & Purisch, 1980)
that the summary scales are heterogeneous in regard to the func-
tions examined. This anomaly can be reconciled when it is realized
that an odd-even split on the LNNB spuriously inflates estimates
of internal consistency because items which deal with a single
set of skills appear on the battery in a consecutive fashion.
Examples of this appear on the Motor and Tactile Scales where a
task is performed on one item by the right hand and then, on the
next item, by the left hand, and on other scales (e.g. Arithmetic)
where two consecutive items assess the same skill (e.g., item 214;
solve 7-4; item 215: solve 27-8). In this type of arrangement,

Anastasi (1976) has suggested that splits other than an odd-even split be used. It would be interesting to examine the data that would be generated when other measures of internal consistency (e.g., the alpha coefficient, Anastasi, 1976; Cronbach, 1970) are applied that do not rely on any particular planned split between items. As of yet, no data has been presented using other forms of splits between items.

VALIDITY

There has been considerable confusion in the literature on the LNNB over what is meant by the general term "validity", and over the precise useages and relevance of the various kinds of validity data that have been presented (Golden, Ariel, McKay, et al., 1982; Spiers, 1981). In an attempt to provide an organization that is consistent with current psychometric conventions regarding validity, this review will adhere to the classification and useage of the various forms of validity as suggested by the American Psychological Association's (1974) Standards for Educational and Psychological Tests and Manuals, and Anastasi (1976), Cronbach (1970), and Thorndike and Hagen (1977). Although the validity of the LNNB will be discussed under the headings content validity, criterion-related (diagnostic) validity, and construct validity, it is acknowledged that these forms of validity are not distinct or independent and that both content validity and criterion-related validity are intimately related to construct validity (Anastasi, 1976).

Content Validity

The ability of the LNNB to provide a sample of behavior thought to be representative of the domain of functions that should be assessed in a comprehensive neuropsychological evaluation has recently been questioned (Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, Note 1). Although both Golden and his colleagues (Chmielewski & Golden, 1980; Golden, Hammeke, & Purisch, 1980; Moses & Golden, 1979; Purisch et al., 1978), and the publishers (University of Nebraska Press, 1979; Western Psychological Services, 1981), have claimed that the battery can provide a comprehensive and extensive evaluation of all major neuropsychological functions, Spiers (1981) posits that the battery "does not adequately or comprehensively assess any major neuropsychological functions" (p. 337).

Two major issues are apparent in the discussion of the content validity of the battery, one relating to the selection of items and the other relating to the possible contamination of items by skills unrelated to those primarily assessed by the items. Crosson and Warren (1982) and Delis and Kaplan (1982) have been critical of the fact that items were chosen to be included in the battery, not on the basis of current knowledge of brain-behavior relationships, but on whether the items were able to statistically discriminate neurological patients from medical controls (Hammeke, Note 2; Golden, Hammeke, & Purisch, 1978). While such a procedure may make the battery relatively efficient in diagnosing brain damage, it is not consistent with the goal of developing an instrument that is capable of providing a

comprehensive and molecular assessment of neuropsychological functions.

The second point is that most of the items are heavily contaminated by requiring the intact use of expressive and receptive language skills (Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, Note 1). Since it is possible that cerebral lesions can impair language functions while leaving other higher mental functions relatively unimpaired, it is important that the evaluation of non-language skills not be consistently contaminated by the presence of the language impairment (Crosson & Warren, 1982; Parsons & Prigatano, 1978). Crosson and Warren (1982) suggest that this contamination issue is most apparent on the Tactile scale where 20 of the 22 items require a verbal response. This renders the distinction between aphasic disturbances and tactile imperceptions difficult. Comparatively, it should be noted that this verbal contamination issue (receptive and expressive) applies equally to many of the tests that make up the HRNTB (see Crosson & Warren, 1982).

The heavy reliance on verbal functions needed for the processing of item instructions and responding has prompted Crosson and Warren (1982) and Delis and Kaplan (1982) to suggest that while the battery may be effective in diagnosing brain damage, it may not be appropriate for examining the functional deficits of patients with language disorders. As Crosson and Warren (1982) have pointed out, the recognition of a heavy verbal weighting on items that tap other functions may be the reason why severely aphasic patients were not included in LNNB research. In fact,

Lewis, Golden, Moses, Osman, Purisch, and Hammeke (1979) have stated that severely aphasic patients were excluded from their investigation because of "the difficulty such patients have in taking most of these tests [the LNNB]" (p. 1007).

The heavy verbal loading is, however, not surprising in the context of the Luria (1966, 1980) or Christensen (1975) approach to neuropsychological evaluation. Here the concern is with the assessment of primarily left hemisphere functioning through a flexible and systematic application of tasks that permit the qualification of the symptom. In this framework there are no requirements for the presentation of standardized instructions, or for verbal responses, as the goal is not to maintain a standardized approach in which a fixed set of items are administered. The goal is to isolate the deficit that resulted in the collapse of the functional system using any means possible.

In rebutting the criticism that verbal contamination is pervasive in the battery, Golden, Ariel, Moses, et al. (1982) state that the item instructions may be changed to suit the needs of the individual patient, and that verbal responding (particularly on the Tactile scale) is not necessary since any type of response that can communicate the same information is sufficient. Although this suggestion moves the battery closer in spirit to Luria's (1966, 1980) investigation, it is unknown how far one may go in being flexible without compromising the standardization necessary for the valid use of the scale score conversions or the norms.

A related point is that in instances where the contamination is minimal, there are too few items available to adequately sample a particular skill (Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, Note 1; See also Golden, Ariel, McKay et al., 1982; Spiers, 1982). This could result in misinterpretation of the deficits or, more importantly, diagnostic errors when the impairment is mild. While this problem is present for all summary scales (see Spiers, 1981, Note 1, for a comprehensive discussion), it is most notable for the Reading Scale which does not assess reading comprehension (Crosson & Warren, 1982), and for the Memory Scale which does not assess recent or remote memory (Spiers, 1981).

While it is apparent that the content validity of any instrument is not dependent upon anyone estimation of how adequately the relevant domain of behavior has been sampled, the concerns raised here should sound a note of caution to those who attempt to use the LNNB. It would be incumbent upon clinicians who use the battery to be aware of the potential limitations in the sample of behavior that will be generated and to supplement the evaluation with other instruments in areas that the battery does not adequately assess. Although Golden has demonstrated his facility in providing, what appear to be, detailed and comprehensive evaluations of neurological patients (see Golden, Ariel, Moses et al., 1982), it is obvious that the administration and interpretation of the battery (as with all such batteries), requires an in-depth knowledge of brain-behavior relationships and their expression on the LNNB. Hopefully clinicians will be

motivated to report difficulties they encounter in using the
battery such as has been done by Crosson and Warren (1982) and
Delis and Kaplan (1982) in their work with aphasic patients.

<div align="center">Criterion-Related (Diagnostic) Validity</div>

Since criterion-related (diagnostic) validity should be
evaluated in terms of the types of decisions that the battery
is suggested to be effective in making (American Psychological
Association, 1974), the following sections will be organized to
reflect this. Discussion will initially focus on the ability of
the battery to discriminate neurological from intact subjects and
schizophrenics, and then will turn an examination of the ability
of the battery to lateralize and localize brain damage.

Presence of Brain Damage

Brain-damaged vs. normal. The initial investigation demon-
strating the ability of the LNNB to discriminate neurological
from control subjects was carried out as a dissertation by
Hammeke (Note 2) and was subsequently reported in the literature
by Golden, Hammeke, and Purisch (1978) and Hammeke et al. (1978).
While it is clear from an inspection of the demographic and LNNB
assessment data, and from statements made in the LNNB manual
(Golden, Hammeke, & Purisch, 1980), that the same subjects and
data were used for the reports, other publications have implied
that the data presented in the published reports (Golden, Hammeke,
& Purisch, 1978; Hammeke et al., 1978) were generated from inde-
pendent validation studies (e.g., Golden, 1981b; Moses & Golden,
1979; Osmon, Golden, Purisch, Hammeke, & Blume, 1979; see Adams,
1980a). The implication that they are independent validation

studies becomes strengthened when one finds that the diagnoses for the neurological group differ depending on whether the primary source (Hammeke, Note 2) or a secondary source (Golden, 1981b) is referenced.

Much greater precision and accuracy needs to be applied in the reporting of the subject characteristics of the samples used in LNNB validation research. It is also essential that the investigators clearly state whether a report represents the independent collection of data or whether it is based solely on a reorganization of data such that a different emphasis is high-lighted. Only the more recent reports have approximated these goals (e.g., Golden, Fross, & Graber, 1981; Golden, Moses, Fishburne et al., 1981) leaving it unclear as to how often the same basic group of subjects has been used.

The initial validation study (Hammeke, Note 2) compared the LNNB performance of 50 hospitalized, neurological patients to the performance of 50 patients hospitalized for conditions with no central nervous system involvement. Although the criteria for diagnosis was not indicated in the original reports, subsequent publications (Golden, 1980; Golden, Hammeke, & Purisch, 1980) have stated that the diagnoses for all neurological patients involved in LNNB research were made by one or more of the following techniques, depending on clinical need: computerized axial tomography (CAT scan), electroencephalogram (EEG), angiogram, pneumoecephalogram, surgery, skull x-rays, neurological examination, and neurological history. While there were no differences between groups in sex distribution, age, chronicity,

age of onset, or the number of previous hospitalizations, the control group had significantly more education than the neuro-logical group.

Sequential t-tests revealed that the neurological group was significantly (p <.05) more impaired than the control group on 250 (89%) of the 282 test items (Hammeke, Note 2). While Hammeke (Note 2) did not attempt to control for the education difference between groups, Golden, Hammeke, and Purisch (1978) report that the use of multiple two-group analyses of covariance, with education as the covariate, yields essentially no change over the results obtained by the multiple t-tests. Although education has been consistently found to relate to neuropsychological test performance (e.g., Heven, 1980; Parsons & Prigatano, 1978), the lack of relationship here is not surprising since each item taps such a small sample of behavior.

It is difficult, however, to accept the results of the analysis on the individual items at face value because of incon-sistencies between reports of the data. Golden, Hammeke and Purisch (1978) state that 253 of the 285 items significantly discriminated the groups which is at variance with what Hammeke (Note 2) reports. There is no indication where the additional three items came from. Further, while Hammeke (Note 2) indicates that, using a discriminant analysis, the linear combination of 60 items was able to correctly classify all subjects, Golden, Hammeke, and Purisch (1978) claim that it was possible to achieve such perfect accuracy with only 30 items. Although the reasons for this discrepancy are unknown, any discriminant analysis

that has such low subject to variable ratios (per group) presents
serious problems for interpretation because of the instability
of the discriminant function coefficients (Fletcher, Rice, &
Ray, 1978). It is almost a certainty that most of the items
used in the classification procedure added little discriminating
ability (independent variance) over that provided by the items
which were included in the first few steps of the stepwise
discriminant analysis. Since there is no indication of how many
items added a significant amount of independent variance to the
discrimination, the value and appropriateness of the analysis is
in doubt.

The value of performing 282 (285) t-tests must also be
questioned. Irrespective of how many significant **differences were**
observed, the use of significance tests to such a degree, without
any consideration of the inflation of the type 1 error rate,
represents a flagrant disregard for current statistical and
research techniques (Adams, 1980a, 1980b; Glass & Stanley, 1970;
Hays, 1973). This abuse of a statistical technique would not
have had such serious consequences if the individual items had
not been interpreted. It is unfortunate, however, that this
multiple t-test procedure was used as the basis for deciding
which items would be retained in the battery (Golden, 1981a).

The Hammeke (Note 2) dissertation also examined the ability
of the 14 summary scales to differentiate the neurological and
control groups. As reported by Hammeke et al. (1978), the two
groups were found <u>not</u> to differ in amount of education, which is
at odds with what was found by Hammeke (Note 2) and Golden,

Hammeke, and Purisch (1978), although the data presented on the means and standard deviations of the summary scales in the Hammeke (Note 2) dissertation were identical to those reported by Hammeke et al. (1978). The explanation offered in subsequent publications (Golden, 1981a; Golden, Hammeke, & Purisch, 1980) to resolve this apparent contradiction was that five of the original neurological patients (Hammeke, Note 2) were replaced such that the five new patients, while being matched for diagnosis and severity with the original, yielded a sample of neurological patients which did not differ from the sample of controls in amount of education. Further, by oversight, the original data from Hammeke (Note 2) were published rather than the revised data from the sample containing the new subjects. Examination of the revised data presented in Golden (1981a) reveals that it differs only very slightly (no more than 2%) from the original (Hammeke, Note 2).

The logic of the revision procedure is, at best, unclear. One important question that remains unanswered is why would Hammeke et al. (1978) feel justified in manipulating the sample composition to achieve equality between groups along one dimension, using a methodologically suspect strategy, when the equality of the groups could have been statistically derived as in the manner reported by Golden, Hammeke, and Purisch (1978) using what would have been the same subjects? It is possible that if analyses of covariance, with education as the covariate, were used to investigate the two groups on the summary scales, no significant differences would result. That this is plausible

is demonstrated by the finding that the amount of education was significantly related to scores on all the LNNB summary scales (Marvel, Golden, Hammeke, Purisch, & Osmon, 1979).

Since there are no differences in the actual results between the original (Hammeke, Note 2) and revised samples (in Golden, 1981a), the data from the original sample will be discussed. Significant differences were found between the neurological and control groups on all the summary scales. Using optimal cutoff points designed to maximize the percentage of correct classifications, it was found that the accuracy of the individual scales to correctly classify subjects ranged from 74% (Expressive Speech) to 96% (Memory) in the control group ($\underline{M}$ = 85%), and from 66% (Rhythm) to 86% (Expressive Speech) in the neurological group ($\underline{M}$ = 74%). A discriminant analysis using all the summary scales as dependent measures was able to correctly classify 86% of the neurological group and 100% of the control group, to yield an overall hit rate of 93%. There are no data presented to indicate how many dependent variables added a significant amount of independent variance to the discrimination. Thus, it is not possible to determine how robust the function is, or conversely, how much it takes advantage of random variation in the sample.

The initial results on the ability of the battery to discriminate neurological from control patients have been cross-validated by Duffala (1978-79) and Moses and Golden (1979). Although Duffala found that neurological patients differed significantly from controls on all 14 summary scales, the study

is flawed because the control group was both younger, and more educated, than the neurological group. This could result in differences in neuropsychological functioning between groups irrespective of the presence or absence of brain damage (Hevern, 1980; Parsons & Prigatano, 1978).

Using 50 neurological patients and 50 medical controls, Moses and Golden (1979) employed the optimal cutoff points derived by Hammeke et al. (1978) and achieved similar, but marginally lower, hit rates. The groups were similar in terms of age, education, and sex distribution. The hit rates ranged from 72% (Writing) to 98% (Motor and Pathognomonic) in the control group ($\underline{M}$ = 84%), and from 62% (Memory) to 82% (Receptive Speech) in the neurological group ($\underline{M}$ = 70%). A classification procedure using the Hammeke et al. (1978) discriminant function yielded an 88% hit rate for the neurological group and a 98% hit rate for the control group, resulting in an overall hit rate of 93%. This overall hit rate is identical to that found when deriving the discriminant function (Hammeke et al., 1978). Use of a discriminant analysis based on the Moses and Golden (1979) sample resulted in an overall hit rate of 96%.

Moses and Golden (1979) have suggested that the shrinkage in hit rates that occurs when discriminant functions and optimal cutoff points are cross-validated (Anastasi, 1976; Fletcher, Rice, & Ray, 1978; Heaton, Baade, & Johnson, 1978; Kerlinger & Pedhazur, 1973; Parsons & Prigatano, 1978) was offset by the fact that their neurological sample was more impaired than the sample used in the derivation study (Hammeke et al., 1978). Examination

of the data in both reports, however, suggests otherwise. In
fact, it appears that the derivation sample (Hammeke et al., 1978)
was more impaired (higher summary scale scores) which would have
the effect of leading to greater shrinkage than would have been
the case if the groups were equally impaired. An important
difference between the two studies which is neither discussed nor
mentioned is the use of a battery consisting of 282 items in the
derivation of the optimal cutoff points, and the discriminant
function, and the use of a battery consisting of 269 items in
the "cross-validation" attempt. This should have had the effect
of making it harder for subjects to be classified as brain-
damaged and easier to be classified normal because there would be
less items on each summary scale, a fact which was not reflected
in the results. It would have been expected that the net effect
of using a battery with fewer items, on a less impaired sample,
would be to produce a considerable amount of shrinkage in hit rates.

The nearly identical hit rates found in the derivation
(Hammeke et al., 1978) and cross-validation (Moses & Golden, 1979)
studies, while impressive at first glance, cause concern
because they are seriously at odds with what would normally be
expected in cross-validation attempts. Monte Carlo research
reveals that, in a two-group situation with a subject to vari-
able ratio of between 3:1 and 4:1 (50 subjects and 14 variables
per group), the amount of shrinkage in discriminant analysis
hit rates can range from 12% to 17% (Fletcher, Rice & Ray, 1978).
A recent investigation by Leli and Filskov (1981) provides an
illustration of the magnitude of shrinkage that can be expected

in neuropsychological research. Averaged across four discriminant functions, they found that the hit rates decreased almost 18% from when applied to the derivation samples to when applied to the cross-validation samples. A similar degree of shrinkage in classification accuracy can be noted in Wheeler and Reitan's (1963) cross-validation of discriminant functions derived by Wheeler, Burke, and Reitan (1963). Heaton, Baade, & Johnson (1978) have observed that the median hit rates using optimal cutoff points were 14% higher than the median hit rates found using cross-validated cutoff points. This suggests that the cross-validation of optimal cutoff point results would be expected to result in a considerable amount of shrinkage.

While there is a need for replications of the Hammeke et al. (1978) and Moses and Golden (1979) studies to examine whether the high hit rates can be approximated in other laboratories, the results of these investigations (including Golden, Hammeke, & Purisch, 1978) can not be summarily dismissed as resulting from methodological or statistical artifacts as had been done by Adams (1980a, 1980b) and Spiers (1981). Over and above the concerns regarding the statistical adequacy of some of the techniques used, the major criticism of the Golden, Hammeke, and Purisch (1978), Hammeke et al. (1978), and Moses and Golden (1979) studies is that "the essential discrimination being made is probably between very seriously neurologically impaired (and, perhaps, psychiatrically disturbed) patients and medical controls" (Adams, 1981b, p. 513-514). If this were the case, large differences between groups and high hit rates would be expected

and the test provided of the battery's ability to be sensitive to brain damage would not really reflect the type of decision that the clinician is typically faced with. The criticism, however, is unfair because it ignores the claim that the neurological sample was weighted with cases of "mild" to "moderate" impairment (Golden, 1980; Hammeke, Note 2; Hammeke et al., 1978; Moses & Golden, 1979) and that no patient in the neurological group had a psychiatric history (Golden, 1980). Since there are no data presented as to the severity of the neurological samples, or their psychiatric status (other than the above claims), it is an unwise practise to state that the reports offer "over-dramatized claims of diagnostic accuracy" (Adams, 1980a, p. 522) unless attempts to replicate prove this to be so.

While not attempting to replicate the high hit rates using the optimal cutoffs or the discriminant functions, Malloy and Webster (1981) have demonstrated the validity of the LNNB in detecting brain damage in subjects specifically chosen because they were "mildly" impaired. Subjects were selected because they presented diagnostic dilemmas for the referral source since there were no blatent neurological signs such as hemiparesis, aphasia, or visual field deficits. Three groups of 12 subjects each were formed. The pseudoneurological group was composed of patients who presented with a variety of neurological symptoms but for whom EEG, CAT scan, and neurological exam were negative. The borderline group was similar to the pseudoneurological group except that while the CAT scan and neurological exam were negative, the EEG was "mildly" to "moderately" abnormal. The

brain-damaged group consisted of patients with unequivocally positive neurological exams, EEGs, and CAT scans. There were no differences among groups in age or education.

Malloy and Webster (1981) classified subjects as brain damaged or not brain damaged based on objective clinical rules for interpretating the LNNB derived and cross-validated by Golden, Moses, Graber, and Berg (1981). Golden, Moses, Graber, and Berg (1981) determined that the average summary scale $\underline{T}$-score for normal subjects could be predicted by knowledge of age and education ($\underline{R}$ = .74), and that almost all of the individual summary scale $\underline{T}$-scores fell within 10 $\underline{T}$-score points of this predicted average. Conversely, in a neurological sample, the majority of the summary scale $\underline{T}$-scores were found to be more than 10 $\underline{T}$-score points from their predicted average. The predicted average $\underline{T}$-score for the neurological patients was thought to serve as a rough estimate of premorbid status to which the current level of performance could be compared. A rule was developed such that a subject was classified as brain damaged if more than one summary scale $\underline{T}$-score was above a "critical level" defined as the predicted average $\underline{T}$-score plus 10. If the Arithmetic or Writing scales were above the critical level, a subject was not considered to be brain damaged unless three or more scales were above the critical level. Application of this rule to the 60 normal and 60 neurological patients used in the derivation of the rule resulted in a combined hit rate of 91%. Cross-validation using an additional 60 normal and 60 neurological patients resulted in an 85% hit rate for the neurological patients and an

83% hit rate for the controls, for a combined hit rate of 84% (Golden, Moses, Graber, & Berg, 1981).

Malloy and Webster's (1981) use of a three-scale-elevation-above-the-critical-level rule resulted in the correct classification of 75% of the pseudoneurological group and 83% of the brain-damaged group. Seventy-five percent of the borderline group were correctly classified as brain damaged if the EEG data can be considered to reflect brain abnormality. This is by no means certain as Filskov and Goldstein (1974) suggest. Since the brain-damaged and borderline groups were found not to differ significantly on any of the summary scales, they were combined and compared to the pseudoneurological group. The combined "brain-damaged" group was significantly more impaired than the pseudoneurological group on the Rhythm, Receptive Speech, Expressive Speech, Writing, Memory, Intellectual Processes, and Pathognomonic scales.

The results of the investigations that have examined the ability of the LNNB to detect the presence of neurological impairment, in the context of discriminating neurological patients from medical controls, are comparable to those reported using other neuropsychological assessment techniques (for reviews see Hevern, 1980; Klove, 1974; Lezak, 1976; Reitan, 1975). While this suggests that the battery has the potential to be clinically useful, there is a great need for additional research to corroborate the results that have been reviewed. Although the objective clinical rules for determining the presence or absence of brain damage appear valid, significant questions arise regarding

the cutoff points and the discriminant functions. It would be unwise to attempt to clinically use the cutoff points reported by Hammeke et al. (1978) and Moses and Golden (1979) until their validity is unambiguously demonstrated.

Brain-damaged vs. schizophrenic. Purisch et al. (1978) have examined the effectiveness of the battery in discriminating between hospitalized chronic schizophrenics and hospitalized neurological patients. The demographic data clearly indicates that the neurological sample was composed of the 50 neurological patients reported in Golden, Hammeke, and Purisch (1978) and Hammeke et al. (1978) although this fact is not stated in the Purisch et al. (1978) paper. The chronic schizophrenic group was composed of 50 patients diagnosed primarily as having paranoid or undifferentiated schizophrenic disorders. No patient was included in the schizophrenic group for whom the medical history suggested the possibility of neurological impairment. While the groups did not differ significantly in sex distribution, age, education, or length of current hospitalization, they differed significantly on variables that play a role in diagnosing and defining chronic schizophrenia (chronicity of illness, age of onset, and number of previous hospitalizations.

Sequential t-tests revealed that the neurological sample was significantly more impaired than the schizophrenic sample on 72 (26%) of the 282 items. It was reported in the text that a discriminant analysis was able to achieve 100% classification accuracy using the linear combination of 40 items, but this is contradicted by the report appearing in the paper's abstract

that such accuracy was achieved using 60 items. While the significance of this contradiction is unclear, the profligate use of t-tests and the questionable use of discriminant function techniques is again apparent. The reader is referred to the previous section for a discussion of methodological deficiencies of the use of multiple t-tests, and discriminant analysis when the sample size is small relative to the number of dependent measures.

Purisch et al. (1978) also found that the neurological group was significantly more impaired than the chronic schizophrenic group on all summary scales with the exception of the Rhythm, Receptive Speech, Memory, and Intellectual Processes scales. A discriminant analysis on the 14 summary scales was able to achieve a hit rate of 84% for the neurological group and 92% hit rate for the schizophrenic group, yielding an overall hit rate of 88%. These hit rates are substantially higher than those that result from using cutoff points that maximize the ability of each summary scale to classify subjects. Hit rates using optimal cutoff points ranged from 38% (Right Hemisphere) to 70% (Expressive Speech and Pathognomonic) in the neurological group (M = 58%), and from 34% (Rhythm) to 92% (Right Hemisphere) in the schizophrenic group (M = 61%).

A comparison of the optimal cutoff points used in the discrimination of normal from neurological patients (Hammeke et al., 1978) and chronic schizophrenic from neurological patients (Purisch et al., 1978) reveals that chronic schizophrenics need to be substantially more impaired on the battery than normals to

be considered brain damaged. While it is common practise to use optimal cutoff points in research (Golden, 1980), the use of these cutoffs in clinical practise as the administration and scoring manual (Golden, Hammeke, and Purisch, 1980) suggests would be an error. Not only would such use require a prior diagnostic decision (normal vs. schizophrenic), but may also result in mildly neurologically impaired schizophrenics being classified as unimpaired.

The Purisch et al. (1978) results have been "cross-validated" by Moses and Golden (1980) using the same neurological group as Moses and Golden (1979) and an additional sample of 50 chronic schizophrenic patients (primarily paranoid and undifferentiated). More stringent screening procedures were employed than in the Purisch et al. (1978) investigation as no patient was included in the schizophrenic group for whom the neurological exam, skull x-rays, or EEG was abnormal. Further, where a patient's status was in doubt, CAT scanning was used to rule out cerebral atrophy. While the groups did not differ in age, education or sex distribution, the schizophrenic group was more chronic than the neurological group.

As in the initial investigation, the neurological group was significantly more impaired than the chronic schizophrenic group on all summary scales except the Rhythm, Receptive, Speech, Memory, and Intellectual Processes scales. Application of the Purisch et al. (1978) optimal cut off points resulted in essentially similar hit rates for the neurological group, but higher hit rates for the schizophrenic group. The hit rates

ranged from 46% (Memory) to 74% (Pathognomonic) for the neuro-
logical group ($\underline{M}$=59%), and from 46% (Rhythm) to 92% (Tactile
for the schizophrenic group ($\underline{M}$=72%). Use of the discriminant
function derived by Purisch et al. (1978) on the 14 summary
scales yielded an 88% classification accuracy for the neuro-
logical group and an 86% classification accuracy for the schizo-
phrenic group, for an overall hit rate of 87%.

Although the cross-validation discriminant function hit
rates are impressive due to their magnitude and correspondence to
the hit rates found in the initial derivation study (Purisch et
al., 1978), there are major interpretative problems that pre-
clude uncritical acceptance of the data. As in the cross-
validation attempt with normal and neurological samples (Moses
& Golden, 1979), Moses and Golden (1980) used a battery consis-
ting of 269 items while the discriminant function and optimal
cuttoff derivation study (Purisch et al., 1978) used a battery
consisting of 282 items. Again, this fact was not discussed or
mentioned. While Moses and Golden's (1980) more stringent
screening procedures might have resulted in the schizophrenic
sample appearing less impaired on the battery, thus making the
schizophrenic/neurological discrimination somewhat easier and
possibly offsetting the effect of the reduction in the number of
items, the correspondence in hit rates between the derivation
and cross-validation studies is not in accord with what is known
about cross-validation shrinkage in hit rates (see earlier
discussion).

The major difficulty with the Moses and Golden (1980)

study, however, is that, despite the fact that secondary sources (Golden, 1981a, 1981b; Golden, Hammeke, & Purisch, 1980) state that the neurological patients used were the same patients as used in the earlier Moses and Golden (1979) study, the summary scale T-scores for this group are markedly different in each publication (compare Figure 1s in Moses & Golden, 1979, 1980). There is no mention of this fact in either of the primary reports (Moses & Golden, 1979, 1980) or in the LNNB literature subsequently published. This obvious difference in the data reported for the same subjects deserves comment from the authors. While it may be that the scoring criteria for some of the items were changed and that the difference in the reports reflect this fact, it is not possible to explore this possibility because the relevant reports have not been published (see Scoring section). One is left with a situation in which the data just do not make sense and there is a complete absence of discussion that could clarify the situation.

The overall hit rates from the discriminant functions (87%, Moses & Golden, 1980; 88%, Purisch et al., 1978) "grossly exceed" (Adams, 1980a, p. 523) the hit rates that have been found using other neuropsychological tests in discriminating chronic schizo-phrenics from neurological patients. Based on an extensive review of the literature, Heaton et al. (1978) found that the median hit rate for discriminating chronic schizophrenics using instruments other than the LNNB was 54%. The overall finding that chronic schizophrenics perform much as do neurological patients on neuropsychological tests has prompted Heaton et al.

(1978) to suggest that many chronic schizophrenic patients are in fact neurologically impaired. This suggestion, combined with an increasing body of data based on neurological exams, EEGs, pneumoencephalograms, CAT scans, and histopathological studies that is correlating structural changes in the brain with chronic schizophrenia (Heaton et al., 1978; Reitan, 1976b; Weinberger, Torrey, & Neophytides), strongly implies that despite the screening procedure, many of the chronic schizophrenics used by Purisch et al. (1978) and Moses and Golden (1980) were likewise neurologically impaired. Golden, MacInnes, Ariel, Ruedrich, Chu, Coffman, Graber, and Bloch (1982) have verified this by suggesting that, upon detailed evaluation of the LNNB results, 50% of the chronic schizophrenics used in the research (Moses & Golden, 1980; Purisch et al., 1978) could be classified as brain damaged, and 16% classified as borderline. Golden, MacInnes et al. (1982) further suggest that if more sensitive neurological screening techniques had been used (e.g., ventricular to brain ratio, Weinberger et al., 1979), the neurologically impaired chronic schizophrenics would have been identified. Given this, the distinction between the chronic schizophrenic groups and the neurological groups (Moses & Golden, 1980; Purisch et al., 1978) becomes blurred, and the meaning of the reported high degree of classificatory accuracy unclear. This is particularly troublesome when it is considered that the "chronic schizophrenic" group differed from both the neurological group, and the medical control group (Lewis, Golden, Purisch, & Hammeke, 1979). Clinical interpretation is thus rendered

difficult when the decision to be made is between diagnosing the chronic schizophrenic neurologically normal or impaired.

Golden, MacInnes, et al. (1982) have indicated that although Purisch et al. (1978) and Moses and Golden (1980) were able to achieve great accuracy in discriminating what were called chronic schizophrenics from neurological patients, there are no data to justify suggesting that the schizophrenic groups, as a whole, were normal. While the discrepancy in classification accuracy between research using the LNNB (Moses and Golden, 1980; Purisch et al., 1978) and other instruments (see Heaton et al., 1978) clearly points to the need for independent replication studies, especially given the concerns raised here, a series of well executed studies have provided striking evidence that the battery can discriminate between chronic schizophrenics with, and without, cerebral ventricular enlargement.

Based on a sample of 42 chronic schizophrenics (20-to 40-years-old), Golden, Moses, Zelazowski, Graber, Zatz, Horvath, and Berger (1980) obtained a multiple correlation of .72 between LNNB summary scale T-scores and an objective and reliable measure of cerebral ventricular size, the ventricular to brain ratio (VBR). The VBR was calculated by determining the area of the ventricular space which was then divided by the total cross-sectional area of the brain on the same horizontal CAT scan slice. Many CAT scan slices were made with the slice containing the largest ventricular area used for the calculations (usually at the level of the lateral ventricles). Experimenters determining the VBRs were blind to the LNNB results and the

technicians administering the battery were blind to the VBRs. No relationship was found between the VBR or the summary scales and age, education, chronicity, current medications, length of hospitalization or type of schizophrenia (paranoid vs chronic undifferentiated). Using a complex set of decision rules for interpreting the battery (see original article), Golden, Moses, et al. (1980) were able to correctly classify 23 of 25 (92%) subjects with ventricular enlargement (VBR > .10), and 15 of 17 (88%) subjects without ventricular enlargement (VBR < .10), for an overall hit rate of 90%. The .10 VBR cut off was employed because normative research had indicated that the occurrence of VBRs greater than this was extremely rare (1%) in neurologically normal subjects (Synek & Reuben, 1975; Weinberger et al., 1979). While Golden, Moses, et al. (1980) used a VBR greater than .10 to define the presence of ventricular enlargement, it is stated elsewhere (Golden, MacInnes, et al., 1982) that radiologists require a VBR to exceed .25 before a patient would be classified as neurologically impaired. Since Golden, Moses, et al. (1980) report that the VBRs ranged from .01 to .24 ($\underline{M}$ = .12), the sample must be considered only mildly impaired.

The ability of the LNNB to predict ventricular enlargement in chronic schizophrenics has been replicated by Golden, Graber, Moses, and Zatz (1980) and Golden, MacInnes, et al. (1982). Use of the LNNB decision rules generated by Golden, Moses, et al. (1980) enabled Golden, Graber, et al. (1980) to correctly classify 14 of 20 (70%) subjects with ventricular enlargement (VBR > .10), and 20 of 22 (91%) subjects without ventricular enlargement

(VBR $<$.10), for an overall hit rate of 81%. The range of VBRs
was from .01 to .30 ($\underline{M}$ = .11). Using the same rules, Golden,
MacInnes, et al. (1982) were able to use the LNNB to correctly
classify all the 15 subjects with ventricular enlargement
(VBR $>$.10), and 18 of 28 (64%) subjects without ventricular
enlargement (VBR $<$.10), for a combined hit rate of 77%. VBRs
ranged from .01 to .16 ($\underline{M}$ = .08). The multiple correlation
between the VBR and the summary scales was .76, comparable to
that found in the Golden, Moses et al. (1980) study even though
the summary scales that demonstrated the largest contributions
to the total shared variance between the battery and VBR
differed. The ability of the LNNB to predict VBR's has also
been demonstrated by Golden, Moses, and Graber (Note 4) and
Zelazowski, Golden, Graber, Blose, Bloch, Moses, Zatz, Stahl,
Osmon, & Pfefferbaum (1981) using samples of chronic alcoholics.

While the ability of the LNNB to predict ventricular en-
largement (VBR) in chronic schizophrenics on an acturial basis
(multiple regression) and a clinical basis (empirically derived
objective decision rules) is impressive, the clinical utility of
this is not clear. Neuropsychological performance deficits may
be correlated with many types of structural cerebral changes
(and many other variables for that matter), only some of which
result in changes in VBR. As Golden, MacInnes, et al. (1982)
caution, it is necessary to rule out many acute and chronic
neurological dysfunctions before the absence of ventricular
enlargement can be interpreted as the absence of neurological
impairment. In this light, it is noteworthy that all the chronic

schizophrenics used in the VBR studies (Golden, Graber et al., 1980; Golden, MacInnes, et al., 1982; Golden, Moses et al., 1980) were carefully screened such that no patient was included who had a history of seizures, head trauma, drug or alcohol abuse, or any acute or chronic neurological condition. There remains a great need to examine the ability of the LNNB to discriminate between other forms of psychopathology (e.g., anxiety and affective disorders) and brain damage. Although it is likely that the battery will prove effective in discriminating less severely psychopathologically disturbed patients from brain-damaged patients, given that the discrimination between chronic schizophrenics with, and without, ventricular enlargement is possible, this must be empirically demonstrated. Further, it would be important to empirically test the battery's ability in discriminating chronic schizophrenics without ventricular enlargement from neurological patients.

Lateralization and Localization of Brain Damage

The ability of the LNNB to discriminate among lateralized and diffusely brain-damaged subjects has been investigated by Osmon et al. (1979). Twenty subjects were assigned to each of three groups (left, right, diffuse) depending on the locus of brain damage as confirmed by one or more of the following methods: neurological exam, arteriogram, EEG, CAT scan, pneumoencephalogram, skull x-rays, and surgery. The groups were found not to differ in age or education. Analysis of variance on each of the 14 summary scales revealed that the groups differed significantly only on the Left Hemisphere scale with the left-hemisphere group

more impaired than the right-hemisphere or diffuse groups. The possibility that one significant analysis of variance (out of 14) could have arisen from the operation of random or chance factors is not considered by the researchers. Osmon et al. (1979) also report that a discriminant analysis on the 14 summary scales is able to correctly classify 59 of the 60 subjects (98%).

While the degree of accuracy is somewhat higher than that obtained lateralizing brain damage using discriminant analyses of HRNTB data (e.g., 88%, Goldstein & Shelley, 1973; 93%, Wheeler, Burke, & Reitan, 1963), there are substantial problems with the Osmon et al. (1979) analysis. Aside from the shrinkage in classification accuracy that would be expected from the use of discriminant analysis with a subject (per group) to variable ratio as small as is present (20/14) (Fletcher et al., 1978), there is a lack of appreciation of the statistical improbability of achieving near perfect discriminative classificatory accuracy when the univariate tests indicate the essential equality among groups (Adams, 1980a). This can be demonstrated by the fact that 11 of the 14 univariate F statistics had values less than 1.0, and that the average F value over the 14 tests was .83. This clearly suggests that the groups really did not differ and seriously calls into question the reported high level of accuracy in the classification of subjects based on the discriminant function.

McKay and Golden (1979b), and other secondary sources (e.g., Golden, 1981a, 1981b) have stated that Osmon et al. (1979) were able to classify 75% of their cases as right-hemisphere, left-

hemisphere, or diffuse brain damage using the simple arithmetic difference between the Left Hemisphere and Right Hemisphere scales. Unfortunately, there is no mention of this in the Osmon et al. (1979) report. Although the LNNB has these two scales intended to lateralize brain damage, McKay and Golden (1979b) have empirically identified two alternate sets of items that are considered to be more effective in discriminating lateralized brain damage. The alternate sets of items were identified by comparing the performance of 73 normal subjects with 14 left-hemisphere-damaged subjects, and with 15 right-hemisphere-damaged subjects, on each of the 269 items in the battery. The empirically derived Right Hemisphere scale (designated R*, Golden, Moses, Fishburne, et al. 1981) was generated by retaining those items that significantly discriminated (p <.05) the right-hemisphere group from the normal group based on 269 sequential t-tests. The Left Hemisphere scale (designated L*) was generated in a similar manner by comparing the left-hemisphere group with the normal group using 269 sequential t-tests. To make the new scales similar to each other in length, the "least effective" items of the L* scale were eliminated, as were items that appeared on both scales. The methodological inadequacies of the extravagant use of t-tests without correction for the inflation of the Type 1 error rate is again apparent and will not be belabored here.

McKay and Golden's (1979b) data indicate that the simple use of the arithmetic sign that results from the R* score minus L* score operation was able to discriminate the right-hemisphere

subjects from left-hemisphere subjects with 100% accuracy in the derivation samples. Cross-validation using an additional 30 right-hemisphere-damaged subjects and 41 left-hemisphere-damaged subjects resulted in 88% accuracy in the classification of left-hemisphere damage, and 87% accuracy in the classification of right-hemisphere damage, yielding an overall hit rate of 87%. Although McKay and Golden (1979b) claim that the arithmetic difference between R* and L* Scores would be of "great clinical value in discriminating lateralized groups from each other as well as from schizophrenics and other brain-damaged groups" (p.5), other data in the report contradict this claim. While there is marked separation between the lateralized groups in the difference score (R* - L*), there is much variability in the data reported for normals, schizophrenics, and diffusely brain-damaged sub-jects. This suggests that unless there is a priori knowledge that a particular subject or patient has lateralized brain damage, the simple difference between R* and L* scores would be mislead-ing.

The ability of the LNNB to discriminate among subjects with brain damage localized to one of four "major" areas (frontal, sensorimotor, temporal, and parietal-occipital) of each hemis-phere has been investigated by Lewis, Golden, Moses et al. (1979). Based on examination by CAT scan, angiogram, or surgery, subjects were assigned to one of eight groups depending on the locus of maximum involvement of the lesion in one of the quad-rants of either hemisphere. There were nine subjects in each of the left-hemisphere groups and six subjects in each of the right-

hemisphere groups. The groups did not differ on demographic or treatment variables. A neurologically normal group composed of the intact subjects reported by Hammeke et al. (1978) and Moses and Golden (1979) was used as the control group in this study. One-way analyses of variance on the eight experimental groups and the normal group revealed significant differences on all the 14 summary scales. Based on the finding of 167 significant ($p < .05$) pairwise differences among the groups, out of a possible 504 nonredundant pairwise t-test comparisons, Lewis Golden, Moses, et al. (1979) contend that they have demonstrated the potential of the battery to discriminate among subjects with localized brain damage. They further suggest that lesions in "each area of each hemisphere can be distinguished by a specific pattern of LNNB scores" (p. 1010; see Figures 2-5, Lewis, Golden, Moses, et al., 1979). Spiers (1982) has taken issue with these claims by suggesting that because Lewis, Golden, Moses, et al. (1979) did not take into account the subcortical extent of the lesions, or the difficulty in making precise statements about lesion location due to "mass and contracoup effects" (p. 303), the validity of the conclusions is in doubt.

While the overuse of uncorrected t-tests is again apparent, the major difficulty with the Lewis, Golden, Moses et al. (1979) claims is that an examination of the means and standard deviations of the summary scales for each localized brain-damaged group suggests that the report of significant differences among groups is somewhat in error. For example, despite the fact that Lewis, Golden, Moses et al. (1979) report that there were

13 nonredundant, significant pairwise differences among the eight brain-damaged groups on the Motor Scale (28 possible nonredundant pairwise comparisons, see Table 5, Lewis, Golden, Moses, et al., 1979), a re-calculation of the 13 t-tests using the means and standard deviations reported for the groups (Tables 3 and 4, Lewis, Golden, Moses et al., 1979) resulted in only four significant differences (p $<$ .05, two-tailed). Similar calculation problems are present for the data on the other summary scales. The small sample sizes, use of uncorrected t-tests, and errors in calculation present serious problems that impede the acceptance of the Lewis, Golden, Moses, et al. (1979) study as demonstrating the ability of the battery to localize brain damage.

McKay and Golden (1979a) have empirically derived additional LNNB scales to localize brain damage using a procedure similar to that employed in the empirical derivation of the lateralization scales (McKay & Golden, 1979b). The performance of 77 neurologically normal subjects was compared, on each of the 269 items, to the performance of each of eight groups of subjects with localized brain damage. The procedure used to localize lesions was less adequately outlined than in Lewis, Golden, Moses et al. (1979; See Spiers, 1982). The locally brain-damaged groups were as follows: right frontal (n=5), right sensorimotor (n=3), right parietal-occipital (n=7), right temporal (n=4), left frontal (n=12), left sensiomotor (n=6), left parietal-occipital (n=10), and left temporal (n=6). Eight scales were derived such that each scale consisted of items that were found to statistically discriminate a particular locally brain-damaged group from the normal

group with as little item overlap across scales as possible. Using the highest score from the eight localization scales, McKay and Golden (1979a) were able to correctly localize the lesion in 47 of the 53 (89%) neurologically impaired subjects.

Although there are problems with the derivation of the eight localization scales regarding the very small size of the samples employed and the great number of t-tests used (2152), Golden, Moses, Fishburne, et al. (1981) have demonstrated the ability of the scales to localize brain damage using an additional sample of subjects with focal brain lesions. This study also provided a cross-validation of the ability of the critical level procedure (Golden, Moses, Graber, & Berg, 1981) to detect brain damage, and a cross-validation of the ability of the empirically derived lateralization scales (R* and L* scales, McKay and Golden, 1979b) to lateralize brain damage. Depending on the locus of a single cerebral lesion, 87 neurological subjects were assigned to one of eight groups (left and right frontal, sensorimotor, parietal-occipital and temporal). Subjects with subcortical lesions were excluded from the investigation. Sample sizes of the brain-damaged groups ranged from 6 subjects (left sensorimotor) to 17 subjects (left parietal-occipital), with a mean of 11 subjects per group. A neurologically normal group (n=30) was also used, and there were no differences reported among the nine groups in age or education.

Use of the critical level procedure (three or more summary scales above the critical level indicate brain damage, Golden, Moses, Graber & Berg, 1981) resulted in the correct classification

of 77 of 87 (88%) neurological subjects and 24 of 30 (80%) normal subjects, for an overall hit rate of 86%. The R*-L* difference was able to correctly localize 27 of 38 (71%) right-hemisphere-damaged subjects and 41 of 49 (84%) left-hemisphere-damaged subjects, yielding an overall hit rate of 78%. Use of the highest score on the localization scales to lateralize subjects resulted in higher levels of discrimination (right-hemisphere-damaged subjects, 95%; left-hemisphere-damaged subjects 94%). The highest score on the localization scales was able to correctly localize lesions in 65 of the 87 (75%) neurological subjects, with accuracy for the right-hemisphere-localized lesions (66%) less than the accuracy for left-hemisphere-localized lesions (81%).

The Golden, Moses, Fishburne, et al. (1981) study indicates the potential of the LNNB to detect, lateralize, and localize focal brain lesions. Given the concerns raised here, however, regarding much of the research supposedly demonstrating the battery's diagnostic validity, it is obvious that much more data are required before the instrument can be considered clinically useful. In particular, there is a great need for research to replicate the basic data such that unequivocal statements can be made concerning the efficacy of the battery in making diagnostic decisions.

Construct Validity

Although the data reviewed examining the battery's criterion-related (diagnostic) validity, and the discussion of the content validity, speak to the issue of the construct validity of the instrument, it is necessary to consider other components

and indicators of construct validity such as the internal con-
sistency of the summary scales, factor analysis of the battery,
and correlations of the battery with other instruments. Brain
damage is a multifaceted construct (Davison, 1974; Matarazzo,
1976) and, as such, it would be most appropriate to discuss the
construct validity of the behavioral indices (summary scales)
for the component processes that are thought underlie neuro-
psychological impairment.

Item-Scale Consistency

An important component of the construct validity of the
summary scales that make up the battery, is the internal consi-
stency of each scale. This is a question of whether each item
on a summary scale, supposedly tapping one facet of the con-
struct represented by the scale, is highly correlated with the
total score for the scale. While it has been stated that the
summary scales are not strictly homogeneous with respect to item
content (Golden, Fross, & Graber, 1981; Golden, Hammeke, &
Purisch, 1980), Golden, Ariel, McKay, et al. (1982) posit that the
items of each individual summary scale tap the same general
construct. For example, although items of the Motor scale may
differ in terms of the molecular skills they assess, they all
have as their major component, the requirement that a motor act
be completed.

Using the same sample as employed in investigating the split-
half reliability of each summary scale, Golden, Fross, and Graber
(1981) found that of the 269 items in the battery, 250 were more
highly correlated with the scale on which they were placed than

other summary scales. Although Golden, Fross, and Graber (1981)
interpret this to suggest that the battery has a high degree of
item-scale consistency, an evaluation of the homogeneity of the
scales in terms of their underlying constructs (Anastasi, 1976)
is not possible because the magnitude and patterns of the corre-
lations are not reported. This demonstration that items are
placed on summary scales they are most highly correlated with,
provides a prerequisite for the meaningful interpretation of
each summary scale.

Factor Analysis and Item Intercorrelation

Factor analysis is a powerful technique that can be used to
reduce the dimensionality of data in an effort to uncover the
underlying structure. As such, it can be used to support or
generate theoretical models, and to aid in interpretation when
one is confronted by an array of many variables. Unfortunately,
a series of investigations carried out by Golden and his colle-
agues (Golden, Hammeke, Osmon, Sweet, Purisch, & Graber, 1981;
Golden, Osmon, Sweet, Graber, Purisch, & Hammeke, 1980; Golden,
Purisch, Sweet, Graber, Osmon, & Hammeke, 1980; Golden, Sweet,
Hammeke, Purisch, Graber & Osmon, 1980), attempting to factor
analyze each of the summary scales, falls far short of achiev-
ing these goals. Although the reports suggest that, with the
exception of the Receptive Speech scale, the factor structures
of the remaining summary scales conform to what would be pre-
dicted by Luria's (1966, 1980) theory, there are methodological
problems with the factor analyses which tend to seriously question
the factor analytic results.

There are major difficulties in using variables in a factor analysis which are measured on ordinal scales with few categories (Comrey, 1978; Kim & Mueller, 1978). Recall that the items on the LNNB are scored by a 0, 1, and 2 system and that, of the 269 items, 95 are scored in a purely dichotomous fashion (see Scoring section). Under some conditions it is possible to use factor analysis on data with such a limited number of categories, particularly when the goal of the analysis is strictly exploratory (Kim & Mueller, 1978). The factor analyses, as executed by Golden and his associates, however, are confirmatory in that there is an attempt to relate the obtained factor structures to theory. When factor analysis is used in this manner, with dichotomous or trichotomous variables, there is the possibility that the obtained factor structure may have been greatly distorted because "the correlations between variables can be artificially limited in size or they can be grossly inflated, depending on the situation encountered" (Comrey, 1978, p. 651). Further, strictly speaking, it is not possible to meaningfully express dichotomous variables in a factor analytic model (Kim & Mueller, 1978).

The other major methodological problem with the factor analyses is that they are performed on the combined data generated from 90 neurologically intact subjects, 90 psychiatric subjects, and 90 brain-damaged subjects, with no consideration of the likelihood that the factor structures could differ in each of the three populations that were represented (Spiers, 1982). That this argument has substance is suggested by McKay and

Golden's (1981b) re-examination of the factor structure of the
Receptive Speech scale. McKay and Golden (1981b) employed 105
neurologically intact subjects, 94 psychiatric subjects, and 218
brain-damaged subjects. Two-hundred and seven of these sub-
jects were subjects that were used in the initial study (Golden,
Purisch, et al., 1980). While there is no indication as to why
only 77% of the initial subjects were used, one major difference
between the studies is that there were proportionately more
brain-damaged subjects used in McKay and Golden's (1981b) re-
examination than in the initial study (Golden, Purisch, et al.,
1980). McKay and Golden's (1981b) resulting factor structure
(7 factors retained) was markedly different from the factor
structure (2 factors retained) reported by Golden, Purisch, et
al. (1980). An inference that can be made is that since one
difference between studies is the greater proportion of brain-
damaged subjects in the McKay and Golden (1981b) sample, the
factor structure for the brain-damaged subjects may differ from
that of the neurologically intact and psychiatric subjects.
Another plausible and equally testable hypothesis is that the
difference in results between studies is due, not to the differ-
ences in sub-sample structure, but to the increased stability of
the factor structure that arises only because the factor analysis
in the McKay and Golden (1981b) study is based on a large number
of subjects.

Since McKay and Golden (1981b) state that their results
suggest that the earlier data (Golden, Purisch et al., 1980)
were "spurious", it is surprising that there has been no attempt

to re-examine the factor structures of the other summary scales because, even though they may appear consistant to what was predicted by theory, they could be equally "spurious". Before factor analyses are performed on the combined data from samples that may represent different populations, it behooves the researcher to demonstrate that the factor structures are equivalent across the samples.

Similar methodological problems are present in studies that have demonstrated intercorrelations among the items of the battery (Golden & Berg, 1980a, 1980b, 1980c, 1980d, 1981a, 1981b). The major problem with these studies, however, is the sheer quantity of data presented. These investigations have reported the correlation of each item of a particular scale with each of the remaining items of the battery. For example, each of the 13 items on the Writing scale was correlated with each of the remaining 256 items on the battery (Golden & Berg, 1980a). While there are only 36,046 nonredundant bivariate correlations possible, given 269 items, the procedure used by Golden and Berg can potentially generate 63,936 bivariate correlations using the 11 basic summary scales (excluding Left Hemisphere, Right Hemisphere, and Pathognomonic), of which results based on 26,534 correlations have been reported. The methodological, conceptual, and interpretative problems in dealing with such a large number of correlations stagger the imagination.

Although Golden and Berg (1980a, 1980b, 1980c, 1980d, 1981a, 1981b) have attempted to interpret only those correlations that were highly significant, their procedure is suspect because the

criteria for including correlations for interpretation changes from one study to the next. In the initial report investigating the correlations between Writing scale items and the remainder of the items on the battery (Golden & Berg, 1980a), only correlations that were above .40 were interpreted while, in the other reports, a criteria of .35 was applied (Golden & Berg, 1980b, 1980c, 1980d, 1981a, 1981b). Inspection of the data reveals that if the .40 criterion was used consistently throughout the studies, approximately 50% (Visual Scale) to 80% (Rhythm scale) of the "interpretable" correlations using the .35 criterion would have been lost. Thus, while it is obvious that the change in criteria was employed to prevent a major loss of data, such a procedure can not be endorsed because it represents a situation where decision rules are manipulated to create results that are consistent with the investigators expectations.

Given the concerns raised here regarding the factor analytic and item intercorrelation reports, these studies can not be used to speak to the construct validity of the summary scales of the LNNB. A factor analysis of the items is possible if the items are scored with many categories or if the summary scale scores are used as the data points, although different issues would be addressed by each approach. In this light, it seems unfortunate that Golden (1980; Golden, Hammeke, & Purisch, 1980) chose to forego a five point scoring system in favor of one which was dichotomous and trichotomous.

Correlation With Other Instruments

The correlation of a new instrument with instruments well established in current use provides evidence as to how well the construct(s) supposedly measured by the new instrument is assessed (Anastasi, 1976). When the correlation between the new instrument and those currently in use is high, evidence accumulates suggesting that the new instrument has construct validity (insofar as the currently used instruments have construct validity). Such information has great clinical relevance. If it can be established that the new instrument is highly correlated with instruments currently in use (i.e., measures the same skills), not only would it be possible for the clinician to feel confident that he or she is assessing motor skills, for example, with the Motor scale, but also, it would be possible to make a choice of which instrument to use based on factors such as administration time required, ease of administration, equipment needs, and the potential of the data generated to be used in making diagnostic and treatment decisions.

Golden, Kane, Sweet, Moses, Cardellino, Templeton, Vicente, and Graber (1981) have examined the relationship between the 14 LNNB summary scales and 14 of the major variables of the HRNTB. The 14 HRNTB scores that were used derived from the Category Test, Tactile Performance Test (Time, Memory, and Location), Rhythm Test, Speech Sounds Perception Test, Trail Making Test (A and B), Aphasia Screening Test, Sensory-Perceptual Test, Finger Tapping Test (Dominant and Non-Dominant), Wechsler Adult Intelligence Scale (WAIS) Verbal IQ, and WAIS Performance IQ.

Using a multiple correlation procedure, it was found that the 14

LNNB summary scales were highly correlated with each of the 14

HRNTB variables. When each of the HRNTB variables were used as

dependent measures with the LNNB summary scales as predictors,

the multiple correlations ranged from .71 (dependent measure,

Finger Tapping-Dominant) to .96 (dependent measure, WAIS Performance

IQ), with a mean multiple correlation of .85. When each of the

LNNB summary scales were used as the dependent measures with the

HRNTB variables as predictors, the multiple correlations ranged

from .77 (dependent measure, LNNB Rhythm) to .94 (dependent measure,

Visual), with a mean multiple correlation of .86. Although a

canonical analysis (Kerlinger & Pedhazur, 1973) would have been

a more appropriate technique to use with this data, the high

degree of shared variance (approximately 73%) between the LNNB

and the HRNTB tends to suggest that the two batteries overlap

considerably in the basic set of skills they each assess.

The high degree of shared variance between the two batteries

has been indirectly confirmed by the Golden, Kane, et al. (1981)

demonstration that essentially identical hit rates have been

achieved by each battery in discriminating between the same

groups of control and neurological subjects. Use of a discrim-

inant analysis on the 14 LNNB summary scales was able to achieve

a hit rate of 87% in the neurological group (42/48) and a hit

rate of 88% in the control group (53/60), while a discriminant

analysis on the 14 HRNTB variables achieved a 90% hit rate in the

neurological group (43/48) and a 84% hit rate in the control

group (50/60). Comparable hit rates for the two batteries in

detecting the presence of brain damage (each approximately 80%)

were also obtained in a blind, expert, clinical interpretation of protocals from a mixed psychiatric and brain-damaged sample (Kane, Sweet, Golden, Parsons, & Moses, 1981). A similar high degree of concordance between the expert clinical interpretation of the HRNTB and Luria's procedures has been reported by Diament and Hijmen (1981). This latter study is noteworthy because it was Christensen's (1975) adaption of Luria's neuropsychological investigation that was employed, not the LNNB.

While the Golden, Kane, et al. (1981) study indicates that the LNNB assesses roughly the same general set of functions as the HRNTB, there is a need to examine how well the individual LNNB summary scales assess the constructs that are thought to underlie the scales. This is a question of whether a scale such as Memory really does assess what are currently thought to be memory functions. Despite the suggestion that content analysis (e.g., Spiers, 1981) indicates that the summary scales do not assess the functions for which they are named, the final resolution of this issue rests primarily on data derived from empirical work.

Although the data indicate that the correlations between WAIS IQs and the Intellectual Processes scale are high (WAIS Verbal IQ, - .86, Performance IQ, - .76, Full Scale IQ, - .86, Prifitera & Ryan, 1981; WAIS Verbal IQ, -.84, Performance IQ, -.74, Full Scale IQ, - .84, McKay, Golden, Moses, Fishburne, & Wisiniewski, 1981), the specificity of this is lost when it is realized that the other LNNB summary scales also correlate highly with WAIS IQs. Excluding the correlations with the Intellectual

Processes scale, correlations with the WAIS Verbal IQ range from
- .47 (Tactile) to - .81 (Arithmetic) with a mean of - .67,
correlations with the WAIS Performance IQ range from - .56
(Tactile) to - .71 (Memory) with a mean of - .65, and correla-
tions with the WAIS Full Scale IQ range from - .53 (Tactile) to
- .80 (Arithmetic) with a mean of - .70 (McKay, Golden et al.,
1981). WAIS IQs were also found to be the strongest predictors,
in the set of HRNTB variables, of all the LNNB summary scales
with exceptions being the Tactile and Right Hemisphere scales
where they were the second and third strongest predictors, res-
pectively (Golden, Kane, et al., 1981). Thus, while the Intelle-
ctual Processes scale measures psychometric intelligence as
defined by the WAIS, to a lesser extent, so too do the other
summary scales. This is not surprising when one considers that
since intelligence tests measure adaptive functioning (Matarazzo,
1976), and that adaptive functioning is mediated by the cerebral
cortex, it would be expected that there would be a considerable
proportion of shared variance between tests designed to measure
intelligence and those designed to measure neuropsychological
functioning.

When the WAIS IQ scores are excluded from the correlations
between the HRNTB and the LNNB, the Aphasia Screening Test
emerges as one of the strongest predictors of the LNNB summary
scales (Golden, Kane, et al., 1981). Since the Aphasia Screening
Test assesses various aspects of receptive and expressive lang-
uage skills (Reitan & Davison, 1974), the correlations with the
summary scales indicate that language skills are involved in

the performance of all facets of the skills that the LNNB assesses. This partially supports the contention that scales presumably not intended to assess language functions (e.g., Tactile) are contaminated by the requirement of the use of language skills (Crosson & Warren, 1982; Delis & Kaplan, 1982; Spiers, 1981, Note 1). Given this, it is possible that the construct measured by a particular summary scale differs depending on the type of patient group assessed (Crosson & Warren, 1982). For example, in aphasic patients, unless the administration and scoring criteria are greatly changed to account for receptive and expressive difficulties, scales such as the Tactile or Visual scales may only assess language deficits, not tactile and visual skills as is suggested by the scale labels. In patients without language disorders, however, these scales may well assess tactile and visual functions.

Additional research, using patient groups that differ in their presentation of neuropsychological symptom patterns, is needed before it would be possible to suggest the nature of the constructs that underlie the summary scales on the LNNB. Although the data suggest that the LNNB and the HRNTB overlap considerably in the skills that they assess, and that the LNNB measures much of what is considered to be psychometric intelligence, little is known regarding the construct validity of the LNNB summary scales (the exception being the Intellectual Processes scale, see above). One important avenue for future research would be to examine how well the individual summary scales correlate with other instruments which are thought to

assess the same underlying construct. A valuable component of this approach would be to use the methodology as suggested by Campbell and Fiske (1959) in which it is necessary not only to demonstrate that summary scales correlate well with other instruments that tap the same construct (convergent validation), but also that the summary scales do not correlate well with instruments that tap other constructs (discriminant validation).

CONCLUSIONS

In attempting to standardize Luria's (1966, 1980) procedures, Golden and his team have developed an instrument that falls far short of the goal of combining the qualification of the symptom and the standardized, quantitative approaches to neuropsychological assessment. The only similarity that the LNNB has to the neuropsychological investigation procedures Luria (1966, 1980) described is that there is, most likely, some degree of overlap in the type of tasks patients are asked to perform. This similarity, however, is more apparent than real because Luria's procedure for investigation was not tied to any set of items, but to a particular methodology that was flexible, yet systematic, in an attempt to provide a thorough evaluation of each patient's strengths and weaknesses. In Luria's framework, procedures for each individual patient are carefully chosen so that the fundamental defect underlying a symptom can be isolated.

In contrast, the LNNB embodies a methodology in which a fixed set of items is presented fairly rigidly to every patient, and hence, the relevance of many of the items for a particular patient may be minimal. Although Golden (1981a; Golden, Ariel,

Moses, et al., 1982) states that the interpretation of the LNNB involves combining quantitative data (summary scales, lateralization scales, localization scales) and qualitative impressions (performance on individual items, patterns across items), it is far from clear how the battery could be used to systematically decompose a symptom into it's constituent parts as is required by the qualification of the symptom approach (Spiers, 1982). The standardized administration of a fixed set of items to every patient seems incompatible with the flexible and individualized investigation of the qualification of the symptom approach.

Since, in this analysis, the LNNB does not approximate the qualification of the symptom approach to neuropsychological assessment, what can be said concerning the utility of the instrument as a standardized, quantitative battery? In the main, although there is a need for replication to demonstrate that some of the effects are not artifacts due to statistical and methodological inadequacies, the literature suggests that the battery may have promise in detecting the presence of brain damage, and in localizing focal lesions to one of the quadrants of either hemisphere. Particularly suggestive are the data that demonstrate that the use of objective clinical rules (e.g., highest score on the localization scales) can localize brain damage (Golden, Moses, Fishburne, et al., 1981), and that the battery can be used to predict cerebral ventricular enlargement in chronic schizophrenics (e.g., Golden, MacInnes, et al., 1982).

Although there are data to indicate that the LNNB and the HRNTB assess much the same neuropsychological skills

(Golden, Kane, et al., 1981), and are virtually identical in diagnostic power in discriminating neurological subjects from control and psychiatric subjects (Golden, Kane, et al., 1981; Kane et al., 1981), it is premature to endorse the use of the LNNB for clinical practise. The research base that unambiguously demonstrates the effectiveness of the battery is not large enough to justify placing confidence in the clinical use of the instrument. The caution extends to the use of the current set of norms (Golden, Hammeke, & Purisch, 1980) for which there are major questions regarding the limitations that should be imposed upon their general use.

An important clinical question for which there is no clear answer is whether the LNNB, as a standardized instrument, is able to provide information for the neuropsychologist that has direct treatment implications. Although the suggestion that the battery "does not adequately or comprehensively assess any major neuropsychological function" (Spiers, 1981, p. 337) seriously calls into question the ability of the battery to describe and measure the behavioral manifestations of brain damage, it is clear that only empirical tests will provide resolution to this issue. As of yet, the only tests to appear have demonstrated fairly convincingly that the LNNB does provide an assessment of neuropsychological functions comparable to the HRNTB (Golden, Kane, et al., 1981) and WAIS (McKay, Golden, et al., 1981). Additional research will be needed to examine how adequately or comprehensively the battery assesses the neuropsychological functions that the summary scales are named after.

This form of research is crucial because the neuropsychologist's role is such that he or she must be prepared to provide not just some form of prediction regarding the nature and location of cerebral lesions, but also, precise statements about the behavioral consequences of lesions for litigation and rehabilitation counselling (Parsons, 1970; Parsons & Prigatano, 1978; Satz & Fletcher, 1981; Wedding & Gudeman, 1980; Bigler & Steinman, 1981).

The clinical utility of the LNNB does not depend upon either the publisher's (University of Nebraska Press, 1979; Western Psychological Services, 1981) and test developer's claims (Golden, Hammeke, & Purisch, 1980), or conceptual and methodological critiques (Adams, 1980a, 1980b; Crosson & Warren, 1982; Spiers, 1981, 1982), but upon carefully planned and well executed research. The clinical utility of the LNNB can only be determined on the basis of the findings of future research. Until such research is completed, the promotion and marketing of the LNNB as an instrument providing comprehensive neuropsychological evaluation, "developed and standardized under rigorous test construction and validation procedures" (Western Psychological Services, 1981, p. 43) is, at best, misleading.

Reference Notes

1. Spiers, P.A. <u>Analysis of Scales, 2, 4, 7, 8 and 9 of the Luria-Nebraska Battery</u>. Unpublished Manuscript, 1981.

2. Hammeke, T.A. <u>Validity of the Luria Neuropsychological Investigation in the diagnosis of cerebral function</u>. Doctoral dissertation, University of South Dakota, 1978.

3. Wilkening, G.N., Golden, C.J., MacInnes, W.D., Plaistead, J.R., & Herman, B. <u>The Luria-Nebraska Neuropsychological Battery Children's Revision: A preliminary report</u>. Unpublished manuscript, University of Nebraska Medical Center, 1981.

4. Golden, C.J., Moses, J., & Graber, B. <u>Relationship of ventricular size to neuropsychological functioning</u>. Paper presented at the International Neuropsychological Society Meeting, San Francisco, 1980.

References

Adams, K. M. An end of innocence for behavioral neurology?
Adams replys. Journal of Consulting and Clinical Psychology,
1980, 48(4), 522-524. (a)

Adams, K. M. In search of Luria's battery: A false start.
Journal of Consulting and Clinical Psychology, 1980, 48(4),
511-516. (b)

American Psychological Association. Standards for educational &
psychological tests and manuals. Washington, D.C.: American
Psychological Association, 1974.

Anastasi, A. Psychological testing (4th ed.). New York:
Maxmillian, 1976.

Bigler, E. D., & Steinman, D. R. Neuropsychology and computer-
ized axial tomography: Further comments. Professional
Psychology, 1981, 12(2), 195-197.

Butcher, T. N., & Tellegin, A. Common methodological problems
in MMPI research. Journal of Consulting and Clinical Psychology,
1978, 46(4), 620-628.

Butters, N. & Cermak, L. S. Alcoholic Korsakoffs syndrome.
New York: Academic Press, 1980.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant
validation by the multitrait - multimethod matrix.
Psychological Bulletin, 1959, 56, 81-105.

Chmielewski, C., & Golden, C. J. Alcoholism and brain-damage:
An investigation using the Luria-Nebraska Neuropsychological
Battery. International Journal of Neuroscience, 1980, 10
(2/3), 99-105.

Christensen, A. L. Luria's neuropsychological investigation.
New York: Spectrum, 1975.

Comrey, A. C. Common methodological problems in factor analytic
studies. Journal of Consulting and Clinical Psychology, 1978,
46(4), 648-659.

Craig, P. L. Neuropsychological assessment in public psychiatric
hospitals: The current state of the practise. Clinical
Neuropsychology, 1979, 1(4), 1-7.

Cronbach, L. J. Essentials of psychological testing (3rd ed.).
New York: Harper and Row, 1970.

Crosson, B. & Warren, R. L. Use of Luria-Nebraska Neuropsycho-
logical Battery in aphasia. A conceptual critique. Journal
of Consulting and Clinical Psychology, 1982, 50(1), 22-31.

Davison, L. A. Introduction. In R. M. Reitan & L. A. Davison
(Eds.), Clinical neuropsychology. New York: Wiley, 1974.

Delis, D. C., & Kaplan, E. The assessment of aphasia with the
Luria-Nebraska Neuropsychological Battery: A case critique.
Journal of Consulting and Clinical Psychology, 1982, 50(1),
32-39.

Diament, J. J., & Hijmen, R. Comparison of test results between
two neuropsychological test batteries. Journal of Clinical
Psychology, 1981, 37(2), 355-359.

Duffala, D. Validity of the Luria-South Dakota Neuropsychological
Battery for brain-injured persons. Dissertation Abstracts
International, 1978-79, 39, 3506B.

Filskov, S. B. & Goldstein, S. G. Diagnostic validity of the Hal-
stead-Reitan Neuropsychological Battery. Journal of Consulting
and Clinical Psychology, 1974, 42(3), 382-388.

Fletcher, J. M., Rice, W. J., & Ray, R. M. Linear discriminant function analysis in neuropsychological research: Some uses and abuses. Cortex, 1978, 14, 564-577.

Galin, D. Hemispheric specialization: Implications to psychiatry. In R. G. Grenell & S. Gobay (Eds.), Biological foundations of psychiatry, VI. New York: Raven Press, 1976.

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood, N. J.: Prentice-Hall, 1970.

Golden, C. J. Clinical interpretation of objective psychological tests. Grune & Stratton, Inc., New York, 1979.

Golden, C. J. In reply to Adams "In search of Luria's battery: a false start". Journal of Consulting and Clinical Psychology. 1980, 48(4), 517-521.

Golden, C. J. The Luria-Nebraska Neuropsychological Battery: Theory and research. In P. McReynolds (Ed.), Advances in Psychological Assessment, (Vol. 5). San Francisco: Jossey-Bass Publisher, 1981. (a)

Golden, C. J. A standardized version of Luria's neuropsychological tests. In T. Boll & S. Filskov (Eds.), Handbook of clinical neuropsychology. New York: Wiley, 1981. (b)

Golden, C. J., Ariel, R. N., McKay, S. E., Wilkening, G. N., Wolf, B. A., & MacInnes, W. D. The Luria Nebraska Neuropsychological Battery: Theoretical orientation and comment. Journal of Consulting and Clinical Psychology, 1982, 50(2), 291-300.

Golden, C. J., Ariel, R. N., Moses, J. A., Wilkening, G. N., McKay, S. E., MacInnes, W. D. Analytic techniques in the interpretation of the Luria-Nebraska Neuropsychological Battery. Journal of Consulting and Clinical Psychology, 1982, 50(1), 40-48.

Golden, C. J., & Berg, R. A. Interpretation of the Luria-Nebraska Neuropsychological Battery by item intercorelation: The Writing Scale. Clinical Neuropsychology, 1980, 2(1), 8-12. (a)

Golden, C. J. & Berg, R. A. Interpretation of the Luria-Nebraska Neuropsychological Battery by item intercorelation: Items 1-24 of the motor scale. Clinical Neuropsychology, 1980, 2(2), 66-71. (b)

Golden, C. J. & Berg, R. A. Interpretation of the Luria-Nebraska Neuropsychological Battery by item intercorelation: Items 25-51 of the Motor scale. Clinical Neuropsychology, 1980, 2(3), 105-108. (c)

Golden, C. J. & Berg, R. A. Interpretation of the Luria-Nebraska Neuropsychological Battery by item intercorelation: The Rhythm scale. Clinical Neuropsychology, 1980, 2(4), 153-156. (d)

Golden, C. J. & Berg, R. A. Interpretation of the Luria-Nebraska Neuropsychological Battery by item intercorelation: The Tactile scale. Clinical Neuropsychology, 1981, 3(1), 25-29. (a)

Golden, C. J. & Berg, R. A. Interpretation of the Luria-
Nebraska Neuropsychological Battery by the item intercorelation:
The Visual scale. Clinical Neuropsychology, 1981, 3(2), 22-26.

Golden, C. J., Berg, R. A. & Graber, B. Test-retest reli-
ability of the Luria-Nebraska Neuropsychological Battery.
Cited in Golden, C. J., Hammeke, T. A., & Purisch, A. D.
The Luria-Nebraska Neuropsychological Battery Manual. Los
Angeles: Western Psychological Services, 1980.

Golden, C. J., Fross, K.H. & Graber, B. Split-half and item-
scale consistency of the Luria-Nebraska Neuropsychological
Battery. Journal of Consulting and Clinical Psychology, 1981,
49(2), 304-305.

Golden, C. J., Graber, B., Moses, J. A., & Zatz, L. M.
Differentiation of chronic schizophrenics with and without
ventricular enlargement by the Luria-Nebraska Neuropsycho-
logical Battery. International Journal of Neuroscience, 1980,
11(2), 131-138.

Golden, C. J., Hammeke, T. Osmon, D., Sweet, J., Purisch, A.,
& Graber, B. Factor analysis of the Luria-Nebraska Neuro-
psychological Battery: IV. Intelligence and Pathognomomic
scales. International Journal of Neuroscience, 1981, 13(2-3)
87-92.

Golden, C. J., Hammeke, T. A., & Purisch, A. D. Diagnostic
validity of a standardized neuropsychological battery
derived from Luria's neuropsychological tests. Journal of
Consulting and Clinical Psychology, 1978, 46, 1258-1265.

Golden, C. J., Hammeke, T. A., & Purisch, A. D. The Luria-
Nebraska Neuropsychological Battery Manual. Los Angeles:
Western Psychological Services, 1980.

Golden, C. J., Kane, R., Sweet, J., Moses, J. A., Cardellino,
J. P., Templeton, R., Vicente, P., & Graber, B. Relation-
ship of Halstead-Reitan Neuropsychological Battery to the
Luria-Nebraska Neuropsychological Battery. Journal of
Consulting and Clinical Psychology, 1981, 49(3), 410-417.

Golden, C. J., & Kuperman, S. K. Graduate training in clinical
neuropsychology. Professional Psychology, 1980, 11, 55-63.

Golden, C. J., MacInnes, W. D., Ariel, R. N., Ruedrich, S. L,,
Chu, C. C., Coffman, J. A., Graber, B. & Bloch, S. Cross-
validation of the ability of the Luria-Nebraska Neuropsycho-
logical Battery to differentiate chronic schizophrenics with
and without ventricular enlargement. Journal of Consulting
and Clinical Psychology, 1982, 50(1), 87-95.

Golden, C. J., Moses, J. A., Fishburne, F. J., Engum, E., Lewis,
E. P., Wisniewski, A. M., Conley, F. K., Berg, R. A., & Graber,
B. Cross validation of the Luria-Nebraska Neuropsychological
Battery for the presence, lateralization, and location of brain
damage. Journal of Consulting and Clinical Psychology, 1981,
49(4), 491-507.

Golden, C. J., Moses, J. A., Graber, B., & Berg, R. Objective
clinical rules for interpreting the Luria-Nebraska Neuro-
psychological Battery: Derivation, effectiveness and valida-
tion. Journal of Consulting and Clinical Psychology, 1981,
49(4), 616-618.

Golden, C. J., Moses, J. A., Zelazowski, R., Graber, B., Zatz, L. M., Horvath, T. B., & Berger, P. A. Cerebral ventricular size and neuropsychological impairment in young chronic schizophrenics. Archives of General Psychiatry, 1980, 37, 619 -623.

Golden, C. J., Osmon, D., Sweet, J., Graber, B., Purisch, A., & Hammeke, T. Factor analyses of the Luria-Nebraska Neuropsychological Battery III: Writing, Arithmetic, Memory, Left and Right. International Journal of Neuroscience, 1980, 11(4), 309-315.

Golden, C. J., Purisch, A. D., & Hammeke, T. A. The Luria-Nebraska Neuropsychological Battery: A manual for clinical and experimental uses. Lincoln, University of Nebraska Press, 1979.

Golden, C. J., Purisch, H., Sweet, J., Graber, B., Osmon, D., & Hammeke, T. Factor analysis of the Luria-Nebraska Neuropsychological Battery II: Visual, Receptive, Expressive and Reading scales. International Journal of Neuroscience, 1980, 11(3), 227-236.

Golden, C. J., Strider, F. D., Strider, M. A., Moore, G. F., & Gust, W. F. Neuropsychological effects of acute syphilitic involvement of the central nervous system: A case report. Clinical Neuropsychology, 1979, 1(3), 24-28.

Golden, C. J., Sweet, J., Hammeke, T., Purisch, A., Graber, B., & Osmon, D. Factor analysis of the Luria-Nebraska Neuropsychological Battery I: Motor, Rhythm, and Tactile scales. International Journal of Neuroscience, 1980, 11(2), 91-99.

Goldstein, G., & Shelley, C. H.   Univariate vs. multivariate
analysis in neuropsychological test assessment of lateralized
brain damage.   Cortex, 1973, 9, 205-216.

Goodglass, H., & Blumstein, S.   Psycholinguistics and aphasia.
Baltimore: Johns Hopkins University Press, 1973.

Goodglass, H., & Kaplan, E.   Assessment of cognitive deficits
in the brain-injured patient.   In M. S. Gazzaniga   (Ed.),
Handbook of behavioral neurology; Volume 2:   Neuropsychology.
New York: Plenum Press, 1979.

Hammeke, T. A., Golden, C. J., & Purisch, A. D.   A standardized
short and comprehensive neuropsychological test battery based
on the Luria-Nebraska Evaluation.   International Journal of
Neuroscience,   1978, 8(3), 135-141.

Hartlage, L. C., & Telzrow, C. F.   The practise of clinical
neuropsychology in the U. S.   Clinical neuropsychology, 1980,
2(4), 200-202.

Hays, W. L.   Statistics for the social sciences (2nd ed.).
New York: Holt, Rinehart, & Winston, 1973.

Heaton, R. K., Baade, L. E., & Johnson, K. L.   Neuropsychological
test results associated with psychiatric disorders in adults.
Psychological Bulletin, 1978, 85(1), 141-162.

Heilman, K. M. Apraxia.   In K. M. Heilman & E. Valenstein (Eds.),
Clinical neuropsychology. New York: Oxford, 1979.

Heilman, K. M., & Valenstein, E.   Clinical neuropsychology.
New York: Oxford, 1979.

Hellige, J. B., Cox, P. J., Litvac, L. Information processing
in the cerebral hemispheres: Selective hemispheric activa-
tion and capacity limitations. Journal of Experimental
Psychology: General, 1979, 108, 251-279.

Heven, V. W. Recent validity studies of the Halstead-Reitan
approach to neuropsychological assessment: A critical review.
Clinical Neuropsychology, 1980, 2(2), 49-61.

Kane, R. L., Sweet, J. J., Golden, C. J., Parsons, D. A., & Moses,
J. A. Comparative diagnostic accuracy of the Halstead-
Reitan and standardized Luria-Nebraska Neuropsychological
Battery's in a mixed psychiatric and brain-damaged population.
Journal of Consulting and Clinical Psychology, 1981, 49(3),
484-485.

Kerlinger, F. N., & Pedhazur, C. J. Multiple regression in
behavioral research. New York: Holt, Rinehart, & Winston,
1973.

Kim, J. D. & Mueller, C. W. Factor analysis: Statistical
methods and practical issues. Beverly Hills, C. A.: Sage
Publications, 1978.

Klove, H. Validity studies in adult clinical neuropsychology.
In R. M. Reitan & L. A. Davison (Eds.), Clinical neuro-
psychology, New York: Wiley, 1974.

Leli, D. A. & Filskov, S. B. Actuarial detection and descrip-
tion of brain impairment with the W-B Form I. Journal of
Clinical Psychology, 1981, 37(3), 615-622.

Levin, H. S., Grossman, R. G., Rose, J. E., & Teasdale, G. Long term neuropsychological implications of closed head injury. Journal of Neurosurgery, 1979, 50, 412-422.

Lewis, G. P., Golden, C. J., Moses, J. A., Osmon, D. C., Purisch, A. P., & Hammeke, T. A. Localization of cerebral dysfunction with a standardized version of the Luria Neuro-psychological Battery. Journal of Consulting and Clinical Psychology, 1979, 47(6), 1003-1019.

Lewis, G., Golden, C. J., Purisch, A. P., & Hammeke, T. A. The effects of chronicity of disorder and length of hospital-ization on Luria Neuropsychological Battery in a schizo-phrenic population. Clinical Neuropsychology, 1979, 1(4), 13-17.

Lezak, M. D. Neuropsychological assessment. New York: Oxford, 1976.

Luria, A. R. Neuropsychology in the local diagnosis of brain damage. Cortex, 1964, 1(1), 3-18. (Reprinted in Clinical Neuropsychology, 1980, 2(1), 1-7.)

Luria, A. R. Higher cortical functions in man. New York: Basic Book, 1966.

Luria, A. R. The functional organization of the brain. Scientific American, 1970, 222 (March), 66-78.

Luria, A. R. The working brain. New York: Basic Books, 1973.

Luria, A. R. Neuropsychology and the study of higher cortical functions. In A. L. Christensen, Luria's neuropsychological investigation. New York: Spectrum, 1975.

Luria, A. R. Higher cortical function in man (2nd ed.). New York: Basic Books, 1980.

Luria, A. R., & Majovski, L. W. Basic approaches used in American and Soviet neuropsychology. American Psychologist, 1977, 32, 959-968.

Malloy, R. R. & Webster, J. S. Detecting mild brain impairment using the Luria-Nebraska Neuropsychological Battery. Journal of Consulting & Clinical Psychology, 1981, 49(5), 768-770.

Marvel, G. A., Golden, C. J., Hammeke, T., Purisch, A., & Osmon, D. Relationship of age and education to performance on a standardized version of Luria's Neuropsychological Tests in different patient populations. International Journal of Neuroscience, 1979, 9(2), 63-70.

Matarazzo, J. D. Wechsler's measurement and appraisal of adult intelligence. Baltimore: Williams & Wilkins, 1976.

Matarazzo, J. D., Matarazzo, R. G., Weins, A. N., & Gallo, A. R. Retest reliability of the Halstead impairment index in a normal, a schizophrenic, and two samples of organic patients. Journal of Clinical Psychology, 1976, 32(2), 338-349.

McKay, S. E. & Golden, C. J. Empirical derivation of experimental scales for localizing brain lesions using the Luria-Nebraska Neuropsychological Battery. Clinical Neuropsychology, 1979, 1(4), 19-23. (a)

McKay, S. E., & Golden, C. J. Empirical derivation of neuropsychological scales for the lateralization of brain damage using the Luria-Nebraska Neuropsychological Battery. Clinical Neuropsychology, 1979, 1(2), 1-5. (b)

McKay, S. E., & Golden, C. J. The assessment of neuropsychological skills using scales derived from factor analysis of Luria Nebraska Neuropsychological Battery. _International Journal of Neuroscience_, 1981, _14_, 189-204. (a)

McKay, S. E., & Golden, C. J. Re-examination of the factor structure of the receptive scale of the Luria-Nebraska Neuropsychological Battery. _International Journal of Neuroscience_, 1981, _14_, 183-188. (b)

McKay, S. E., Golden, C. J., Moses, J. A., Fishburne, F., & Wisniewski, A. Correlation of the Luria-Nebraska Neuropsychological Battery with WAIS. _Journal of Consulting and Clinical Psychology_, 1981, _49_(6), 940-946.

Moscovitch, M., & Klein, C. Material specific perceptual interference for visual words and faces: Implications for models of capacity limitations, attention and laterality. _Journal of Experimental Psychology: Human Perception and Performance_, 1980, _6_, 590-604.

Moses, J. A., & Golden, C. J. Cross-validation of the discriminative effectiveness of the standardized Luria Neuropsychological Battery. _International Journal of Neuroscience_, 1979, _9_(3), 149-155.

Moses, J. A., & Golden, C. J. Discrimination between schizophrenic and brain-damaged patients with the Luria-Nebraska Neuropsychological Test Battery. _International Journal of Neuroscience_, 1980, _10_ (2/3), 121-128.

Newlin, D. B., & Tramontana, M. G. Neuropsychological findings in a hyperactive adolescent with subcortical brain pathology. Clinical Neuropsychology, 1980, 2(4), 178-183.

Osmon, D. E., Golden, C. J., Purisch, H. D., Hammeke, T. A., & Blume, H. G. The use of a standardized battery of Luria's tests in the diagnosis of lateralized cerebral dysfunction. International Journal of Neuroscience, 1979, 9(1), 1-9.

Parsons, O. A. Clinical neuropsychology. In C. D. Spielberger (Ed.) Current topics in clinical and community psychology (Volume 2). New York: Academic Press, 1970.

Parsons, O. A., & Prigatano, G. P. Methodological considerations in clinical neuropsychological research. Journal of Consulting and Clinical Psychology, 1978, 46(4), 608-619.

Prifitera, A., & Ryan, J. J. Validity of the Luria-Nebraska Intellectual Processes scale as a measure of adult intelligence. Journal of Consulting and Clinical Psychology, 1981, 49(5), 755-756.

Purisch, A. P., Golden, C. J., & Hammeke, T. A. Discrimination of schizophrenic and brain injured patients by a standardized version of the Luria Neuropsychological tests. Journal of Consulting and Clinical Psychology, 1978, 46(6), 1266-1273.

Reitan, R. M. Methodological problems in clinical neuropsychology. In R. M. Reitan & L. A. Davison (Eds.), Clinical neuropsychology. New York: John Wiley, 1974.

Reitan, R. M. Assessment of brain-behavior relationships. In P. McReynolds (Ed.) Advances in psychological assessment. San Fransisco: Jossey-Bass, 1975.

Reitan, R. M. Neuropsychology: The vulgarization Luria always wanted. _Contemporary Psychology_, 1978, _21_(10), 737-738. (a)

Reitan, R. M. Neurological and physiological bases of psychopathology. _Annual Review of Psychology_, 1976, _27_, 189-216. (b)

Reitan, R. M., & Davison, L. A. _Clinical neuropsychology_. New York: John Wiley, 1974.

Russell, E. W. Theoretical bases for the Luria-Nebraska and the Halstead-Reitan batteries. Paper presented at the American Psychological Association, Montreal, 1980. Cited in Crosson, B., & Warren, R. L. Use of the Luria-Nebraska Neuropsychological Battery in aphasia: A conceptual critique. _Journal of Consulting and Clinical Psychology_, 1982, _50_(1), 22-31.

Satz, P., & Fletcher, J. M. Emergent trends in neuropsychology: An overview. _Journal of Consulting and Clinical Psychology_, 1981, _49_(6), 851-865.

Sergent, J., & Bindra, D. Differential hemispheric processing of faces: Methodological considerations and reinterpretation. _Psychological Bulletin_, 1981, _88_(3), 541-554.

Spiers, P. A. Have they come to praise Luria or bury him?: The Luria-Nebraska Battery controversy. _Journal of Consulting and Clinical Psychology_, 1981, _49_(3), 331-341.

Spiers, P. A. The Luria-Nebraska Neuropsychological Battery revisited: A theory in practise or just practicing? _Journal of Consulting and Clinical Psychology_, 1982, _50_(2), 301-306.

Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), _Handbook of experimental psychology_. New York: Wiley, 1951.

Synek, V., & Reuben, J. R. The ventricular-brain ratio using planimetric measurement of EMI scans. _British Journal of Radiology_, 1975, _49_, 233-237.

Thorndike, R. L., & Hagen, E. P. _Measurement and evaluation in psychology and education_ (4th ed.). New York: Wiley, 1977.

Tucker, D. M. Lateral brain function, emotion, and conceptualization. _Psychological Bulletin_, 1981, _89_(1), 19-46.

University of Nebraska Press. (Advertisement for the Luria-Nebraska Neuropsychological Battery.) _Clinical Neuropsychology_, 1979, _1_, 5.

Wedding, D., & Gudeman, H. Implications of computerized axial tomography for clinical neuropsychology. _Professional Psychology_, 1980, _11_, 31-35.

Weinberger, D. R., Torrey, E. F., Neophytides, A. N., & Wyatt, R. J. Lateral cerebral ventricular enlargement in chronic schizophrenia. _Archives of General Psychiatry_, 1979, _36_, 735-739.

Western Psychological Services. Advertisement. _APA Monitor_, 1981, October, 43.

Wheeler, L., Burke, C. J., & Reitan, R. M. An application of discriminant functions to the problem of predicting brain damage using behavioral variables. _Perceptual and Motor Skills_, 1963, _16_, 417-440.

Wheeler, L., & Reitan, R. M. Discriminant functions applied to the problem of predicting cerebral damage from behavior tests: A cross validation study. Preceptual and Motor Skills, 1963 16, 681-701.

Zelazowski, R., Golden, C. J., Graber, B., Blose, I. L., Bloch, S., Moses, J. A., Zatz, L. M., Stahl, S. M., Osmon, D. C., & Pfefferbaum, A. Relationship of cerebral ventricular size to alcoholics performance on the Luria Nebraska Neuro-psychological Battery. Journal of Studies in Alcohol, 1981, 42(9), 749-756.