In Defense of the Halstead Reitan Battery: A Critique of Lezak’s Review

Elbert W. Russell

Private Practice, Miami, FL

Since many neuropsychologists utilize the Halstead Reitan Battery (HRB), it is important to answer a critical review contained in Lezak’s (1995) book. Lezak understands hypothesis testing but not HRB pattern analysis or the use of fixed batteries. A fixed battery provides a constant background, which reveals individual characteristics of the patient. Issues discussed include, types of batteries, legitimate review methods, and Reitan’s method. This writing demonstrates that almost all of Lezak’s criticisms of either the HRB or the Halstead Russell Neuropsychological Evaluation System (HRNES) are either incomplete, misleading or erroneous. Her critique of the Reitan method involved a confusion of terms. In spite of attempts to discredit the HRB not a single sound study questioning the validity of the HRB was presented, while many studies have demonstrated its validity. The fallacy of nonrefutation asserts that it is a fallacy to condemn methods that have been validated while recommending procedures that have not been validated. Lezak questions Reitan’s method, the HRB and the HRNES, which have been thoroughly validated, while presenting no validating studies of her own recommended methods. © 1998 National Academy of Neuropsychology, Published by Elsevier Science Ltd

In a section of her recent book Neuropsychological Assessment, Lezak (1995) reviews the Halstead Reitan Battery (HRB) and some of its derivative methods. Lezak’s book is widely referenced and consequently, quite influential. However, the review often misrepresents the rationale and validity of the HRB as it exists today. Since a fairly high proportion of neuropsychologists utilize, at least portions of, the HRB (Butler, Retzlaff, & Vanderploeg, 1991; Lees-Haley, Smith, Williams, & Dunn, 1996), it is important to present a rebuttal to Lezak’s review.

As part of the rebuttal some of the rationale for the HRB will be presented. This is necessary since much of Lezak’s criticism is derived from a misunderstanding of the methods utilized by the HRB. These methods are often not well understood by many neuropsychologists, especially those using the approach that Lezak advocates.

In addition, several of the methods derived from the HRB will also be discussed, due to Lezak’s strong criticisms. These include Reitan’s Neuropsychological Deficit Scale (NDS; Reitan, 1987), the Comprehensive Norms for an Expanded Halstead-Reitan Battery (CNE-HRB; Heaton, Grant, & Matthews, 1991) and the Halstead Russell Neuropsychology Evaluation System (HRNES; Russell & Starkey, 1993).

Address correspondence to: Elbert W. Russell, 6262 Sunset Drive, Suite PH228, South Miami, FL 33143.
BACKGROUND DISCUSSION

Lezak’s book is excellent in discussing the hypothesis testing method and reviewing the research on single tests from that perspective. However, Reitan’s method is basically not a hypothesis testing method, but a pattern analysis method. This pattern analysis method is generally misunderstood by Lezak and many other neuropsychologists. Since, many of Lezak’s criticisms are derived from this misunderstanding, a discussion of her review must involve an explanation of the method used by Reitan.

Pattern Analysis Versus Hypothesis Testing

The major underlining division in neuropsychology assessment methods appears to be between what is called “hypothesis testing” and “pattern analysis” approaches. A large proportion of the other differences between approaches, methods, and types of assessment is derived from this distinction. These differences, such as that between flexible and fixed batteries, have been argued over the years. While this division will be briefly presented here, a more thorough explanation of it has been presented elsewhere (Russell, 1994, 1997).

Hypothesis Testing

In the hypothesis testing approach the neuropsychologist selects and utilizes tests in order to answer a particular question framed as a hypothesis (Bauer, 1994; Lezak, 1995, pp. 111–113, 125; Russell, 1994). The neuropsychologist begins with a question; then a hypothesis is constructed to answer the question. A test is then selected in order to test the hypothesis. The score that the patient obtains on a particular test, or the way in which the patient obtained the score, answers the hypothesis. Thus, the hypothesis testing method is basically a one-question one-test method, although the examiner might use additional tests to support the first test. The procedure of asking a series of questions, many exploratory, may be called “multistage,” “serial,” or “ongoing” hypothesis testing. Such neuropsychological batteries become lengthy when there are a number of questions to be answered.

The hypothesis may not simply refer to externally derived questions but may be used as a method for clarifying the nature of deficits within the test battery. An example might be determining the reason for a slowed performance on Block Design. This is a type of exploratory questioning or hypothesis testing. It assumes that the neuropsychologist knows the relationships between tests and brain functions.

The hypothesis testing process almost always requires a flexible battery. Various tests are selected according to their relation to a series of questions or hypotheses and not according to their relation to each other. As such, this method employs a group of tests that have no intrinsic relationship to each other. They are only related to the various hypotheses. Consequently, in the pure hypothesis testing approach, one does not begin with a group of tests that are unrelated to any hypothesis.

However, there is a modified form of this method that many neuropsychologists recommend. This method advocates using a fixed or relatively fixed “core” (Bauer, 1994) or “basic battery” (Lezak, 1995, pp. 121–123). In this method, the neuropsychologist is looking for impaired test scores to provide an indication as to the condition of the patient, which will then provide hypotheses to test. This method is a rudimentary form of pattern analysis, with a partially fixed battery.
Pattern Analysis

The pattern analysis method is primarily concerned with the relationships between tests rather than with the individual scores or level of functioning on particular tests. This method compares tests with each other in order to discover a pattern that reveals information about a cognitive condition (Russell, 1994, 1997). Probably Reitan’s most important contribution to neuropsychology will be his development of the pattern analysis method on an inferential basis. The method and value of pattern analysis has been discussed in several places (Bauer, 1994; Russell, 1984, 1986, 1994).

Neuropsychologists using a pattern analysis method may, of course, use the hypotheses testing method, but it is secondary to that of analyzing patterns. Pattern analysis requires certain design constituents for its application (Russell, 1984, 1986, 1994). Many neuropsychologists apparently are not aware of these requirements or how they are related to pattern analysis.

LEZAK’S UNDERSTANDING OF THE PATTERN APPROACH

Although Lezak’s book correctly presents some of the basic concepts related to pattern analysis (Lezak, 1995, pp. 165–166, 168), Lezak appears to lack an understanding of the implications and requirements of that method for designing test batteries.

The method that Lezak calls pattern analysis does not correctly describe Reitan’s method. She identifies pattern analysis with test scatter, either intratest or intertest (Lezak, 1995, pp. 166–168). While strong intratest scatter may indicate an organic problem, this is generally not a form of pattern analysis. The scatter may involve no comparisons between individual tests, only a comparison of several tests, in general, to the person’s overall level of cognitive ability.

In the pattern analysis method employed in Reitan’s work, comparisons to a general level of normal functioning are not the central procedure. The tests are compared to each other, often without regard as to whether any of the tests are in the normal range. For instance, in a case in which all of the scores are abnormal, if the right hemisphere scores are consistently more impaired than the left hemisphere scores, then the person is considered to have right hemisphere damage. Neuropsychologists who are experienced with a set battery know that a large proportion of the scores of patients who have even moderate brain damage are impaired to some extent. That is, the patterns occur on a background of generalized impairment.

The Fixed Battery

In order to utilize pattern analysis, it is necessary to establish a fixed battery. This has been designated as a set of tests or a set battery (Russell, 1995, 1997). The HRB is a set of tests designed for pattern analysis, not an uncoordinated group of tests. Patterns between tests cannot be observed when tests are continually changed. One cannot observe a relationship between tests A and B when sometimes only A is included in the battery and sometimes only B is included. Consequently, in a flexible battery, no stable basis for comparisons exist. In addition, the formal set battery requires coordinated norms and a consistent form of scale scores. This insures equivalent test scores, which produce a constant background (Russell, 1994).

Without mentioning the HRB directly, Lezak disparages fixed batteries in several ways. She calls fixed batteries “Ready-Made Batteries” (Lezak, 1995, pp. 123–125), which presumably do not require any expertise to assemble. An examiner using a fixed battery has “a
general lack of knowledge about how to...'' conduct neuropsychological testing. As such an
examiner is ‘‘naive’’ or ‘‘inexperienced’’ (Lezak, 1995, pp. 123–125). While these comments
would be relevant to batteries using a hypotheses testing approach, they demonstrate an igno-
rance of the nature of the pattern analysis method, in that pattern analysis requires a set
battery.

The competent use of a fixed or set battery to perform pattern analysis actually requires
more expertise than the use of a flexible battery. In addition to utilizing the same neurological
and assessment knowledge that a hypothesis testing method utilizes, pattern analysis requires
more extensive knowledge of psychometrics and particularly knowledge of how tests interact
with each other.

**Coordinated Norms**

One of the great advantages of a fixed or set battery is that it utilizes coordinated norms.
All of the tests are either normed on the same sample or statistical methods are used to equate
the measures. In such a battery the scores may be directly compared in order to observe
patterns. Lezak appears to understand and appreciate this advantage. She states that The
Wechsler batteries are used for pattern analysis ‘‘...because of their obvious statistical
comparability’’ (Lezak, 1995, p. 168). In addition the CNEHRB uses a ‘‘single standard
score system’’ (p. 168).

She also recognizes that uncoordinated norms derived from various tests normed on diver-
genent populations present problems. In regard to norms from different populations, Lezak
writes, ‘‘although this situation results in less than perfect comparability between the different
tests ... these ‘mixed bag’ norms generally serve their purpose’’ (Lezak, 1995, p. 157). She
presents no research to support this claim.

In Lezak’s presidential address for the International Neuropsychological Society (Lezak,
1988), after criticizing the concept of an IQ, she indicates that ‘‘...subtest profile is an
invaluable aid in conceptualizing the multiplicity of mental abilities and cognitive functions
...’’ (p. 360). She ends this speech by saying ‘‘...the development of a well-standardized,
neuropsychologically sound, and practically relevant set of mental abilities tests—lies before
us as the next major challenge for neuropsychological assessment’’ (p. 360). That statement
was made in 1988. Now there are four batteries that have meet this challenge. These are the
HRNES (Russell & Starkey, 1993), the CNEHRB (Heaton et al., 1991), the NDS (Reitan,
1987) and the Luria-Nebraska Neuropsychological Battery (LNNB) (Golden, Purisch, Ham-
meke, 1985). Of these, three are derived from the Halstead Reitan battery. Profiles can be
derived for the HRNES, CNEHRB, and LNNB.

**The Individual Approach**

One of Lezak’s rules (Lezak, 1995, p. 110), and one that she believes a fixed battery
violates, is that a battery should be individualized or tailored to a patient’s needs, abilities
and limitations. Contrary to her understanding, a fixed battery is designed to examine the
individual person. As Lezak states ‘‘The concept of behavioral deficit presupposes some
ideal, normal, or prior level of functioning against which the patient’s performance may be
measured’’ (Lezak, 1995, p. 98). That is, the attributes that make a person different from
other people are due to variations from a normal brain.

This implies that the way to assess a patient’s individuality is to compare a person’s
responses against a standard fixed background, which represents a normal brain. This is what
an individual test does. The test is fixed in that it cannot be changed each time that it is
administered and it is normed and validated to represent a standard background. If one
changes the construction of a test each time it is given, then there is no basis for distinguishing
the person’s individual responses. The need for this constancy is the basis for the requirements
of reliability and validity in regard to an individual test (American Psychological Association,
1985).

The fixed battery extends this concept of a constant standard to the entire battery. As such,
a fixed battery represents the normal brain. The battery provides a constant background
against which individual differences become evident (Russell, 1994). A pure flexible battery
changes the tests each time a person is examined; so that the person’s overall individual
pattern of responses can not be adequately examined. Thus, the fixed battery provides a
consistent “prior level of functioning against which the patient’s performance may be mea-
sured” (Lezak, 1996, p. 98).

If a person has major problems, such as the loss of an arm or a lack of ability in speaking
English, the fixed battery must be modified, just as a flexible battery must be modified. These
problems do not affect the fixed battery to any greater extent than a flexible one. In fact,
scoring and prescribed changes in the testing procedure are incorporated into the HRB. These
modifications handle most of the more common problems due to sensory and motor impair-
ments.

In the usual neuropsychological examination situation, the hypothesis testing approach
using a flexible battery is not tailored to the whole person but to the referral questions. In
most cases, the pure hypothesis testing method does not even attempt to examine the whole
individual person. Rather it attempts to answer a series of specific referral questions (Russell,
1984).

In addition, a battery that is developed during the pressure of an examination will almost
inevitably lack tests for some aspects of the person. This danger can be countered if the
examiner has a group of tests in mind prior to the examination, which covers all areas of
brain functioning; but then this is an informal fixed battery. In addition, most neuropsychol-
gists who use a fixed battery are not opposed to adding a few tests or modifying the battery
to some extent as the situation and the patient’s condition dictates. Nevertheless, the major
portion of the battery is “fixed” so that the individuality of the patient can be assessed.

**FALLACY OF NON-REFUTATION**

Since this is a discussion of a critical review, it is important to examine a type of fallacy¹
that critiques often make. The type of fallacy, which often occurs in neuropsychology reviews
and arguments, will be called the *fallacy of non-refutation*. The fallacy is committed when
a writing implies that if a test or battery has not been shown to be invalid then it is valid.
This is so obviously false that its theoretical base need not be discussed.

However, in a subtle form, this fallacy is rampant in the critical writings of neuropsychol-
ogy. The fallacy occurs in this manner: A critic presents research studies which show that
a test, battery, method, or approach is not impressively valid and concludes or implies, on
the grounds of the research studies that have been performed, that one should not use this
test or battery. At another point in the writing the same critic advocates the use of a different
test, battery, method, or approach for the same purpose. No validation or reliability studies
have ever been performed on this alternative procedure. The critic states or implies that this
alternative procedure is highly valid and should be used. The implied basis for this belief is
that no one can point to a study that shows the procedure to be invalid.

¹The term *fallacy* is commonly used in philosophy to indicate a general type of erroneous thinking rather than
a specific error. This usage of the term fallacy appeared appropriate to the intentions of this paper.
Application to Lezak’s Work

Lezak’s book is full of such arguments, especially in the review of the HRB and its derivative tests and batteries. The review seemingly demonstrates the faults of the HRB, implying that one should not use this battery. Elsewhere, in the same book, she advocates using her battery. There is not one study in the entire book concerned with the validation of the method proposed in Lezak’s book. Thus, every review statement utilizing research results, which implies that the HRB or the HRNES is too invalid to be used clinically, commits this fallacy. Both of these methods have been validated as batteries.

Battery Validity

The fallacy of non-refutation implies validation. The contention in the present paper is that, since batteries are used to assess human performance as well as individual tests, the batteries should be validated as rigorously as are single tests. If a “battery” is, in fact, only a group of unrelated tests then this requirement does not apply and each test must be evaluated on its own. However, any method or battery that claims to have merit as a unified entity, rather than as a group of unrelated tests, should be statistically treated as a unit. This is the manner in which the WAIS-R battery has been treated. The rules related to norming, reliability, and validity must be applied to such units as well as to individual tests (Russell, 1995). In her critiques Lezak jumps from extreme concern with reliability, validity, and norming for individual tests to an almost complete disregard for these psychometric requirements when groups of tests are concerned; even though she believes that use of test batteries for assessment is necessary.

In failing to examine battery validity she overlooks the rule for determining the validity of a battery when a battery has not been validated as a unit. If the battery has not been validated as a unit then its known validity is only equivalent to the validity of the most accurate single test that is utilized in the battery. Consequently, in order to determine the most valid answer to any question or hypothesis one need administer only that test. Administering other tests does not increase the known validity and it may decrease it. The examiner may presume that administering more tests increases validity, but when there is no evidence for this presumption, the only validity that can be presented to a court in litigation is the validity of the individual tests.

The validation of batteries is complex, but it may be accomplished through several methods. The most basic method is the use of clinical judgement. Studies using clinical judgement have demonstrated that, in fact, clinicians using a particular battery or method can make valid assessments (Russell, 1995). In addition, objective methods may be used. The preliminary and most common objective method is to use an index. Indices include the Halstead Index, the HRNES AIS, the Wechsler IQ scores and many other indices. Indices of lateralization (Adams, 1986; Russell, 1984; Russell, Neuringer, & Goldstein, 1970; Russell & Starkey, 1993) begin to refine the construct validation of neuropsychological batteries. Other actuarial methods, such as discriminant function, may also make a contribution to validation.

LEZAK’S REVIEW OF THE HRB AND DERIVATIVE BATTERIES

Halstead Reitan Battery

In her book, Lezak (1995) devotes a number of pages to reviewing the HRB and some derivative batteries. However, she fails to apply any understanding of pattern analysis to her review of the HRB. In her review, no mention of pattern analysis is made (pp. 709–715). Elsewhere she states that “By and large, the use of pattern analysis has been confined to
tests in the Wechsler batteries . . .” (p. 168). However, the primary method used by Reitan is that of an inferential pattern analysis (Reitan & Wolfson, 1986). In this regard, without mentioning the HRB, Lezak states that “Probably the most common approach to the psychological evaluation of organic brain disorders is through comparison of the test score levels. . . . in other words, through analysis of interest score scatter . . .” (p. 167). To support this statement she references four neuropsychologists, all of whom use the HRB.

**Halstead’s Original Work**

In her review of the HRB, an extraordinary amount of space was devoted to discussing Halstead’s original work (Halstead, 1947). Lezak painstakingly examines the norming of this original work, apparently in order to show that the “norms” were inadequate, at least for the Halstead Index. In this regard she is correct. They were inadequate by present day standards. Unfortunately for the review, Halstead’s original work is about as relevant to the present-day HRB as saying that jet liners are deeply flawed because the Wright brothers first plane only flew 120 feet. This emphasis on the origin of the HRB is probably due to an inability to understand that a test battery can be developed over time. Obviously, flexible batteries cannot develop.

**Reitan’s Method**

The crux of Lezak’s critique was an evaluation of Reitan’s method. This consisted of an attempt to demonstrate that what she calls the “clinical” method is superior to Reitan’s method (Lezak, 1995, pp. 710–711). The means that she used was an attempt to use HRB studies to discredit the HRB method. While somewhat difficult to follow, her argument, evidently, follows these steps.

First, she asserts that Reitan uses an “actuarial” method for interpretation. She states “A distinctive feature of Reitan’s handling of the examination data of the Halstead-Reitan Battery has been reliance on test scores for predicting the nature and the site of the lesion as well as its presence” (Lezak, 1995, p. 710). “Predictions about the site of the lesion and its nature (diffuse or focal, static or changing) are based on statistically identified relationships between test scores . . .” (Lezak, 1995, p. 710). She calls such usage an “actuarial approach” (p. 710).

Secondly, she affirms that several HRB computerized interpretative programs have been developed, which use this actuarial method. “This actuarial approach has encouraged development of computerized interpretations of Halstead-Reitan test protocols . . .” (Lezak, 1995, p. 710).

Third, she quotes studies that used the HRB, which maintain that the “actuarial” method used in the computerized programs is not as accurate as the “clinical” method. Here she quotes G. Goldstein and several studies which found that the clinical method was superior to the actuarial method. From this reasoning Lezak implies that the “clinical” method is more accurate than Reitan’s “actuarial” method. Lezak, of course, espouses the clinical method.

When this argument is examined closely, it is apparent that the argument involved two gross definition shifts. First, Lezak labels Reitan’s method an “actuarial” method, when it is, in fact, a clinical method, in that clinical inference, based on test scores, is used to make interpretations. Her statement that “Predictions about the site of the lesion and its nature (diffuse or focal, static or changing) are based on statistically identified relationships between test scores . . .” (Lezak, 1995, p. 710) is absolutely wrong. Such a statement indicates her lack of understanding of Reitan’s work over the years. Reitan has only used statistics to
reverify the HRB clinical lore, which was derived from clinical experience. He never used statistics to create an actuarial system.

The interpretative computer programs based on the HRB do use “actuarial” or “quasi-actuarial” (Heaton et al., 1981; Russell, 1995) methods but these actuarial programs should not be confused with the “clinical” method that Reitan utilizes. The two methods are quite distinct. Reitan’s clinical method uses clinical inference based on test scores, while an algorithm makes the interpretative decisions in the actuarial computer programs.

When Anthony, Heaton, and Lehman (1980), Heaton, Grant, Anthony, and Lehman (1981), and the other HRB studies used the term clinical they were referring to Reitan’s inferential method, which Lezak had relabeled in her review as an “actuarial” method. Thus, Lezak’s misapplies the term actuarial to Reitan’s inferential method.

Lezak’s second shift occurs when she implies that the “clinical method” that Goldstein and Shelly, 1982, Anthony et al. (1980), Heaton et al. (1981), and other HRB studies used was the same as the clinical method that she advocates in her book. Lezak’s method includes the patient’s history, medical records, and qualitative analysis of tests as well as test scores, whereas the Reitan clinical method, used in these studies, employed only test scores. In fact, none of these quotes or studies, to which Lezak refers, use the term clinical in Lezak’s sense. They all used Reitan’s inferential method. (In practice, as opposed to studies, HRB neuropsychologists use patient’s history, etc.) Thus, Lezak’s attempt to discredit the HRB method is groundless because she does not understand the way that the terms actuarial and clinical are used in regard to these studies.

In addition, Lezak could not have compared studies of her method with those of Reitan’s method, since no comparison studies have ever been made between the HRB and her method. In fact, no comparison studies have ever been made between the HRB and any other neuropsychological approach, except for the LNNB (Golden et al., 1991). (Both methods were approximately equally accurate in assessing the existence of brain damage). Lezak’s clinical method has never been validated, much less compared to Reitan’s method. Since no comparative studies were cited, this part of Lezak’s review has no relevance to the comparative validity of the HRB.

In addition, out of the vast HRB literature Lezak picked out the studies’ and quotes, usually taken out of context, that were the most damaging to the HRB. A thorough review of the validity of the HRB using both clinical judgement and actuarial programs has recently been completed by Russell (1995). This review found that a large proportion of Lezak’s statements in her book, concerning the accuracy of the HRB were either wrong, inaccurate or distorted. An example of such distortion is Lezak’s use of a study by Klesges, Fisher, Pheley, Boschee, and Vasey (1984).

Review of Klesges et al. (1984). In regard to validity, Lezak described only one study that reported any low validity rate for the HRB. This was the Klesges et al. (1984) study. It was used to show that “… one study found many of the HRB tests to be relatively weak discriminators” (Klesges et al., 1984). The Klesges et al. study found a low correspondence between the diagnosis of brain damage based on a computerized axial tomography (CAT) scan examination (and one other unspecified neurological test) and a discriminant function analysis of the HRB. The results were completely at variance with all of the other 26 studies that have been done (Russell, 1995).

A close investigation demonstrated some very questionable aspects of this study. The diagnosis of only 94 of the 141 “brain-damaged” cases was provided. Thus, the diagnosis of 47 cases was unspecified. Although the average age of the subjects in this study was 35 for the “normals” and 37.5 for the brain-damaged subjects, the classification of “head
trauma” was not included among the diagnostic categories. In any general hospital, head trauma patients compose a large proportion of neurological patients in this age range.

These unusual results are completely accounted for when one realizes that, along with several other diagnostic categories, the CAT scan will miss head trauma (when there is no hematoma or skull fracture) especially during the time that the data for this study was collected. Thus, the CAT scan would have found no evidence of brain damage for a major proportion of patients who had actually sustained head trauma. This is common knowledge for neurologists and some neuropsychologists (Barth, Gideon, Sciara, Hulsey, & Anchor, 1986; Bigler, 1991). Since these patients were relatively young and were obtained from a general hospital, it is apparent that many of the patients in this study had head trauma. Table 4 in the Klesges et al. (1984, p. 32) study, indicates that only 54% of the subjects that were called normal by the CAT scan were designated as normal by the HRB. The HRB was correctly identifying the head trauma cases that the CAT scan missed.

Consequently, the Klesges et al. (1984) study was completely defective in regard to the overall accuracy of the HRB and should never have been referenced. The neurological criterion for a neuropsychology validation study should never rely on any single neurological test, even the MRI scans, since every neurological procedure will miss certain kinds of cases. Rather, the neurological criterion for brain damage should be the diagnosis of a qualified neurologist who uses the history and all relevant tests.

Validity of the HRB

What then can be stated about the validity of the HRB method? Contrary to Lezak’s review, the validity of the HRB is well established and high. The first demonstration of the validity or effectiveness of the Halstead battery as a clinical instrument was performed by Reitan in 1955 (Reitan, 1955). Again, in a paper published almost 30 years ago, Reitan (1964) produced the most thorough demonstration of the clinical accuracy of the HRB that has been accomplished. In fact, no other study since then has so well demonstrated the clinical validity of any battery or approach. These two papers were completely neglected in Lezak’s review. Nor was Reitan’s 1962 Annual Review paper referenced. This paper laid out the method of validating a test battery, as well as discussing the validity of the HRB.

Subsequently, there have been a large number of studies validating the HRB and the computer programs based on it. Recently, the use of clinical judgement with the HRB has been reviewed by Garb and Schramke (1996). All of the validation studies concerned with the HRB have also been exhaustively reviewed by Russell (1995). Russell’s review found that the HRB has been validated for the existence of brain damage 39 times: 24 times using indices, 7 times using discriminant function and 8 times using clinical judgement. Since often several of these methods were used in the same study, there have been 26 individual validation studies done on the ability of the HRB to discriminate brain-damaged subjects from normal subjects.

Consequently, the review was able to decisively establish the accuracy of the HRB. The accuracy of indices varied from 58% to 92%, depending on the conditions, but overall an index was about 80% correct. Clinical judgement in predicting brain damage varied from 72% to 92% but when the quality of the studies was taken into consideration, the accuracy was about 85%. The accuracy of discriminant function was essentially the same as the accuracy of clinical judgement. Thus the accuracy of the actuarial programs, including indices, while not quite as accurate as the clinical judgement of experts, was acceptable for clinical purposes.

The accuracy of most nonexpert neuropsychologists can be estimated from a paper by Garb (1989). This paper identified the experience level of the participants in a study by Wedding (1983). The Wedding study, which concerned the validity of the HRB, compared
the accuracy of 14 judges, the Neuropsychological Key (Russell et al., 1970), which is a computer program, and a discriminant function program. The accuracy was determined through the relative ability of these methods to separate right hemisphere, left hemisphere, and diffusely brain-damaged subjects from normals and schizophrenics. The judges included 1 expert and 13 nonexperts with varying degrees of experience.

The expert was somewhat more accurate than the actuarial Neuropsychological Key, 63% to 60%. However, the average of the nonexperts was 54%, definitely below that of the Key. (The discriminant function had the same accuracy as the expert). The Neuropsychological Key was more accurate than all but two of the 13 nonexperts. Thus, this study presents data indicating that at least one neuropsychology actuarial program is more accurate than nonexpert clinical judges using test scores. This finding contradicts the conclusion by Anthony et al. (1980) and echoed by Lezak, that the Neuropsychological Key’s accuracy was “. . . unacceptably low for clinical purposes” (Lezak, 1995, p. 710).

From this review of the HRB studies it is evident that Lezak’s review of the HRB validity studies distorted the data derived from HRB studies and failed to compare the HRB to any other method. There is not one sound research study that has demonstrated the HRB to have low validity (Russell, 1995). As such, Lezak’s review concerning the validity of the HRB must be considered fallacious.

Except for the Luria Nebraska Neuropsychological Battery (LNNB), there are no validation studies of any assessment methods using batteries, other than the HRB. This includes flexible batteries, Benton’s battery, the Boston Process approach and Lezak’s individualized approach. Although these methods are probably fairly accurate, their accuracy is not known and, in particular, it is not known whether their accuracy is as great as that of the HRB. They probably are not as accurate, since more work has been devoted to improving the accuracy of the HRB than to any other battery. In addition, undoubtedly the validity of a flexible battery is highly variable and depends on the expertise of the psychologist. Lezak’s recommendation of the use of her own battery, while discrediting the HRB, is an obvious example of the fallacy of non-refutation.

Factor structure. Lezak also questions the factor structure of the HRB. She states: “Recent studies of the HRB factor structure produced differing results” (Lezak, 1995, p. 712). In spite of the many studies that have been performed on the HRB, the results of only three studies were presented. Examination of these studies found that they did not produce differing results. Even Lezak states: “Replicating this study [2 of the 3 studies referenced] . . . gave the same set of factors” (p. 712). The results of the third study are obscure in Lezak’s review but they were in no way “differing.” From the many factor studies of the HRB (Casey, 1991), it is evident that the HRB provides a consistent set of factors across studies. Except for the LNNB, where are the factor analyses of other batteries, especially Lezak’s battery?

Reliability. Lezak also questions the reliability of the HRB Halstead Index and, by implication, the whole battery. “The reliability of the Impairment Index summary score appears to be questionable” (Lezak, 1995, p. 711). This statement was based almost entirely on one study (Matarazzo, Wiens, Matarazzo, & Goldstein, 1974), in which the test-retest reliability was reported as being $r = .08$. Russell (1992), in a study conveniently not mentioned by Lezak, reanalyzed the data and demonstrated that this low correlation was due to an artifact, that of a restricted range. When the range was corrected the correlation became .88. Other

---

2 I invite any neuropsychologists who know of a validity or reliability study of any battery, method or approach other than those related to the HRB or LNNB to inform me of the studies. Such studies will be referenced in any related future publication that I write.
Defence of the Halstead Reitan Battery

reliability studies using other populations have shown the reliability to be acceptably high (Garb & Schramke, 1996; Russell & Starkey, 1993, pp. 5–6). Only one reliability study has been done on a battery other than the HRB and it was found to have acceptable interrater reliability (Brown, Del Dotto, Fiske, Taylor, & Breslau, 1993).

Presumed disadvantages. Lezak states that some of the disadvantages of the HRB are that it ‘. . . is unwieldy, takes relatively long to administer, and is not suitable for the thorough examination of patients with sensory and motor handicaps . . . ’ (Lezak, 1995, p. 713). In fact, the HRB including the WAIS-R can be completed in about 6 hours by a technician. Lezak’s battery takes ‘two three-hour examination sessions’ (Lezak, 1995, p. 122). At times, the first session may take as long as 6 hours (Lezak, 1995, p. 121). Half of Lezak’s battery is given by the neuropsychologist. As such, her battery requires considerably more time from the neuropsychologist than the HRB. Thus, one could say of Lezak’s battery that it ‘. . . is unwieldy and takes relatively long to administer.’

In regard to patients with sensory and motor handicaps, measurement of these deficits is a forte of the HRB. The measurement procedure related to these deficits is, for the most part, handled by established adjustments in the battery and is accounted for in the scoring system. As such, the scoring makes use of such sensory and motor problems to aid in the interpretation. This usage of the HRB might not be known to a person who has never been trained in the use of the battery. As with every battery, some problems are too severe to be handled without modifying the battery. In this case, there are often established methods that are familiar to a neuropsychologist trained in the use of the HRB.

BATTERIES DERIVED FROM THE HALSTEAD-REITAN BATTERY

There have been a number of variations on the HRB, some of which Lezak reviews. In particular, a major development related to the HRB has been the creation of three major computerized scoring systems using new sets of norms. These are the NDS (1987), the CNEHRB (Heaton et al., 1991) and the HRNES (Russell & Starkey, 1993). In regard to these significant developments, Lezak’s book is completely inadequate. As far as her book is concerned, Reitan’s NDS does not exist: it was not mentioned. In addition, one would not know that the CNEHRB uses a computer scoring program or that the HRNES is primarily a computer scoring program.

In a footnote (Lezak, 1995, p. 15) Lezak states that the subject of computer programs will not be dealt with in her book, except in a few cases where computer use is related to the material she is covering. Although, a computer program has been produced for the CNEHRB to aid in scoring, the program was not central to the norms. Consequently, a review of this HRB method falls within the purview of Lezak’s book and so the CNEHRB should have been reviewed. However, without providing an explanation, Lezak does not review it. The CNEHRB was recently reviewed (Russell, 1997) and found to have the same problems as the HRNES. Nevertheless, both provide a better scoring procedure for the HRB than any other norms. Since Lezak does not review either the NDS or the CNEHRB her review of these batteries cannot be discussed.

HALSTEAD RUSSELL NEUROPSYCHOLOGICAL EVALUATION SYSTEM

Lezak’s review devotes a fair amount of space in her book to reviewing the HRNES (Lezak, 1995, pp. 714–715). This is unexpected in that the HRNES, which is entirely a
computer scoring program, falls outside the purview of Lezak’s book (p. 15). Nevertheless, since she reviews it, her review will be examined.

Other Reviews of the HRNES

There have been several reviews of the HRNES that were generally favorable (Lynch, 1995; Mahurin, 1995; Retzlaff, 1995). These reviews were accurate in their description of the program. While mainly supportive, they had some criticisms, which were generally correct. Some of the criticisms were answered in a paper by Russell (1997) and some remain problems. However, the other three existing scoring programs also had these and/or other problems (Russell, 1997). In the only study so far published of the HRNES, which used airline pilots, Kay, Morris, and Starbuck (1993) found the corrections for educational level to be adequate for both the HRNES and CNEHRB. In a recent article Heaton, Matthews, Grant, and Avitable (1996) demonstrated that age and education corrections do improve the accuracy of the HRB, especially at the extremes of age and education.

Lezak’s Review of the HRNES

Although Lezak’s review of the HRNES severely criticizes the battery, almost every statement which she makes, that was not purely descriptive, was either incomplete, misleading, or erroneous. Even the most basic description of the HRNES was inadequate in that the review did not mention that the HRNES is a computerized scoring program. This is like reviewing a “Windows” program without telling the reader that the review concerns a computer program. There are also various specific errors and misrepresentations that will be examined in some detail, since they concern important but erroneous information concerning the system.

WAIS-R comprehension. Lezak states that the HRNES will not accept the WAIS or WAIS-R Comprehension subtest (Lezak, 1995, p. 714). This is incorrect. During norming, the N for Comprehension was not as large as for the other subtests since it was not routinely used in Russell’s laboratory. It was found to be redundant for neurological patients. The manual (Russell & Starkey, 1993) was not clear on this point. However, if a person had run the WAIS-R portion of the program they would have immediately perceived that Comprehension was included. Consequently, it is quite evident that Lezak only examined the program manual and reviewed this computer program for a professional book without even perusing the actual program.

WAIS and WAIS-R. The review states that the difference between the WAIS and WAIS-R was not equated (Lezak, 1995, p. 714, 1, 22). Again the manual was not clear in regard to this operation. The computer program allows one to use either test. The HRNES transforms the WAIS scores into WAIS-R equivalent scores by adding the difference between the tests to the WAIS-R norms. This method was demonstrated to provide an acceptable equivalence (Russell, 1992).

Examination time. Lezak’s review estimated that the time required for the battery was 10 hours for normal subjects (Lezak, 1995, p. 714). In our laboratory, it takes about 6 hours to administer this examination. Patients are regularly scheduled for 8 hours, during which time they receive the whole battery, the MMPI-2 and a half-hour interview. This time is about the same as that which Lezak’s battery requires (see Lezak, 1995, pp. 121–123).
Coordinated norms. The review questions whether the norms are “coordinated” since the $N$ varies for some of the tests (Lezak, 1995, p. 714, l. 38). As the manual states, the reference scale norming procedure was designed, among other things, to provide equivalent norms for tests that were not part of the original norming group (Russell & Starkey, 1993, p. 33). This problem does not seem to bother the reviewer when it occurs in the CNEHRB, even though the CNEHRB has no method of producing equivalent norms when the $N$ varies.

Norming. Lezak’s review questions the norming procedures used in the HRNES, since the subjects came from the Veterans Administration system and the normal control or comparison patients were medical patients (Lezak, 1995 p. 714). These procedures are thoroughly discussed in the HRNES manual (Russell & Starkey, 1993) and in a more recent publication (Russell, 1997). With the exception that about nine tenths of the veterans are men, they represent a cross section of the country, since a majority of them were drafted. The draft, which supplied most of the veterans in the United States, provided the most representative group of subjects used in any norming program. During the time that the draft was in existence, all men in the United States between the ages of 21 and 45 were drafted. Only those with physical disabilities, mental retardation, psychosis, or those who had occupations essential to the war effort were eliminated. Otherwise they were selected from the general population on a random basis. Thus, these veterans provide the most unbiased sample of the American male population in existence. The representativeness of this selection is supported by the mean IQ of the normal veteran subjects in the HRNES sample. This was a WAIS-R, FSIQ of 102. In regard to females, studies have found that only motor tests demonstrate a significant gender difference (Heaton et al., 1986; Russell, 1997).

Lezak censures the HRNES for using medical patients as normal controls. This issue is hotly debated in neuropsychology but in some studies medical patients have been found to better represent the population, that is to be assessed clinically, than norms using selected normal subjects (Russell, 1988). In addition, Lezak does not mention that the CNEHRB manual avoids stating whether patients with negative neurological examinations were used for norming, raising the suspicion that a portion of their norming subjects were also patients with negative neurological examinations.

The HRNES manual and Russell’s (1997) recent paper discuss the problems that every other attempt at norming the HRB has had. There is no problem with the norming of the HRNES that is not found in all other sets of norms.

After studying the two batteries Russell (1997) concluded that, the norming for both the HRNES and the CNEHRB is more adequately accomplished than any other norms for the HRB tests. Since only the HRB derivative batteries and the LNNB have been normed as batteries, the HRNES and CNEHRB, along with perhaps the LNNB, represent the most acceptably developed norms, that are available.

Lateralization tables. Lezak’s review questions the accuracy of the lateralization tables in the manual since the MRI was not used as criterion. True, as with the norming for almost all of the tests reviewed in her book, most of the subjects were tested before the MRI was developed (Lezak, 1995, p. 714). However, is the reviewer so unsophisticated in neurology as to think that neurologists were unable to accurately lateralize any focal damage, using autopsies, surgery and CAT scans, prior to the advent of the MRI? In the HRNES tables, lateralized cases were not used unless there was unequivocal evidence that the patient had lateralized damage. In this regard, the HRNES manual has some of the most accurate and extensive neuropsychology tables in existence. These indicate the ability of various HRB and related tests to detect both lateralization and the existence of brain damage.
Location charts. Lezak’s review labels the two charts of proposed test localization that are in the HRNES manual, as “phrenology maps” (Lezak, 1995, p. 714). This critique is unnecessary, unwarranted, and unprofessional. It is unnecessary since the validity of the program itself had no direct relation to the diagrams. They could have been left out of the manual without any effect on the understanding or use of the program. They were included because some neuropsychologists had found them useful. In the caveats provided in the HRNES manual, their character and the cautions in their use were carefully explained.

The criticism is unwarranted since the best neurologists and neuropsychologists use functional diagrams of the brain (Geschwind, 1979; Goodglass, 1980). Lezak’s book utilizes two such “phrenology maps” (Lezak, 1995, pp. 57, 70). One might say the charts in Lezak’s book are different since they show functions and not tests. However, phrenology used the term faculties, which was simply another name for functions. Thus, the maps of brain functions in Lezak’s book are more similar to phrenology faculty maps than are the maps in the HRNES manual.

Lezak’s review provides no rationale for criticizing these location maps but substitutes scorn for rationality. Nowhere is a reasoned basis for the scorn presented. Had a rational explanation been attempted, the justification for the review’s ridicule of these neuropsychological test location maps would have been seen to be weak.

Presumably, the problem with the charts in the HRNES manual was that tests rather than functions were utilized. A rational criticism might go like this: It is obvious that functions, not tests are related to various areas of the brain. The brain, after all, does not contain tests. The areas to which tests are related are too poorly known to locate them on a brain map, even when cautions are stated. Further, neuropsychology does not know enough about tests to determine which functions a test represents since a test generally represents several functions. Consequently, “. . . The examiner looks for . . . evidence of impairment on tests involving function or skills that are associated neuroanatomically, in their cognitive expression . . .” (Lezak, 1995, p. 167). This argument, that Lezak might have presented, appears to be reasonable. Reports are written in terms of functions. As such, Lezak might assume that knowledgeable, professional neuropsychologists describe the functioning of the brain not test results (Adams, 1986). The difficulty with this reasoning is that research relates tests to areas of the brain not functions. In support of this examine Lezak’s (1995) book. Almost every test reviewed in that book will have studies relating the test to brain areas. The relation of functions to the brain is only secondarily derived from these studies utilizing tests. In this regard, there has been almost no research designed to determine what specific functions are measured by particular tests. In support of this again, examine Lezak’s (1995) book. There are almost no research studies demonstrating the functions that particular tests represent. Most of the ideas that people have concerning the functions that a test is measuring are “intuitive” or derived from lore, not research. Consequently, we really do not know what functions particular tests represent. This is the missing link in the argument that functions and not tests can be represented in brain maps. Some neuropsychologists have the temerity to believe that when there is no research relating tests to functions, they, in their infinite wisdom, can disentangle the relationships between functions and tests so as to determine which “functions or skills are associated . . . neuroanatomically, in their cognitive expression . . .” (Lezak, 1995, p. 167).

This is an example of the fallacy of non-refutation. No research has demonstrated that these intuitive ideas are wrong, thus, if the ideas have not been proved to be wrong, then they must be right. In fact, our knowledge concerning the location relationship of tests to the brain, especially HRB tests, is better established than that related to functions. Thus, there is more justification for relating tests to areas of the brain than locating cognitive functions to areas. Consequently, the review was unprofessional, since lacking a legitimate rational, it stooped to name calling rather than being a reasoned critique.
Reliability. The review states that the entire HRNES battery was not subjected to a reliability study (Lezak, 1995, p. 714). This is true. The reliability indications were derived from previous studies (Garb & Schramke, 1996; Russell & Starkey, 1993, pp. 35–36). Incidentally, there have been no reliability studies of other HRB batteries and certainly not of Lezak’s method and battery.

Validation. The review questioned the validation studies of the HRNES (Lezak, 1995, p. 714). Such studies, in fact, were extensive (Russell & Starkey, 1993, pp. 35–41). In addition, the validity of the HRNES was supported by the validity studies of the HRB (Russell, 1995). Lezak’s attack appears rather hypocritical when her book reports no validation studies of either her method or her battery.

To justify her questioning of the validity of the HRNES, Lezak’s review states that as an example of the inaccurate “. . . level of cognitive functioning of the population contributing to these scores . . .” (Lezak, 1995, p. 714), the cutting point between normality and brain damage for the Category Test was “10 points higher” than the cutting point of the one that Halstead had derived, which was 51. The HRNES cutting point was given in the manual as 61 (Russell & Starkey, 1994 p. 37, Table 6).

Lezak had previously criticized Halstead’s original norms for being based on a sample whose mean was only 28.3 years old (Lezak, 1995, p. 711). She then failed to notice that the mean age for the HRNES sample was 47.1 (p. 32). When corrected to the age of Halstead’s sample the Category score cutting point was approximately 48, which was almost the same as Halstead’s original score and slightly more stringent. (Prorating was necessary to produce a score at the 28 year age level.) This is rather good agreement for norms that were gathered almost 50 years apart.

Interpretation. The review states that the HRNES has a “rather naively programmed set of recommended interpretations” (Lezak, 1995, p. 714). What this statement is referring to is rather obscure. There was no intention in the HRNES manual of providing an interpretative system (p. 19). The manual does provide some aids for interpretation. Any thorough test or battery will have a similar section, for instance the WAIS-R manual pages 26–50 covers similar types of material (Wechsler, 1981).

Any attempt at a thorough discussion of neuropsychological interpretation using the HRNES would require a full book in itself. The HRNES is a scoring system with some statistical aids that can be used to help interpretation. These include the AIS index and a lateralization index; both of which have been validated (Russell & Starkey, 1993, Russell, 1997). In addition, to provide a statistical basis for interpretation, the manual does present probably the most extensive analyses of the lateralization and diagnostic ability of various neuropsychological tests that has been published.

From this examination of Lezak’s review, it is evident that the eloquent peroration advice of her review is based on a critique, in which almost every item is either wrong, misleading or pejorative, as such it lacks credibility.

CONCLUSION

From this extensive examination of Lezak’s review of the HRB in the most recent edition of her book (Lezak, 1995), it is evident that she has a limited understanding of Reitan’s method. Most of her criticisms would be appropriate to a hypothesis testing approach but not to a pattern analysis approach. Consequently, much of the critique was misplaced. She
failed to understand the reason for a fixed battery in pattern analysis, such as its providing a constant background to reveal the individual characteristics and diagnoses of the patient.

Even then the review was plagued by inaccuracies. Lezak’s attempt to question the validity of the HRB confused the meaning of the terms clinical and actuarial as used in regard to the HRB studies. In addition, Lezak was unable to present a single sound study that questioned the validity of the HRB. The only study that appeared to show a low validity for the HRB was fatally flawed. To the contrary, the validity of the HRB has been thoroughly established and it is quite appropriate for clinical purposes. When scrutinized almost no criticism in Lezak’s review of either the HRB and HRNES was confirmed.

Ultimately, Lezak’s review falters on the fallacy of non-refutation, which asserts that it is a fallacy to condemn methods that have been validated, while recommending alternate methods that have not been validated. Lezak questions the reliability, validity, norming and factor structure of the HRB without presenting a single study concerning any of these aspects of the battery that she advocates. In this regard, some of Lezak’s criticisms of the HRB or the HRNES might have had some cogency if she had presented any evidence that the method and battery, which she advocates, had a single study supporting its norming base, its reliability or its validity.

REFERENCES


