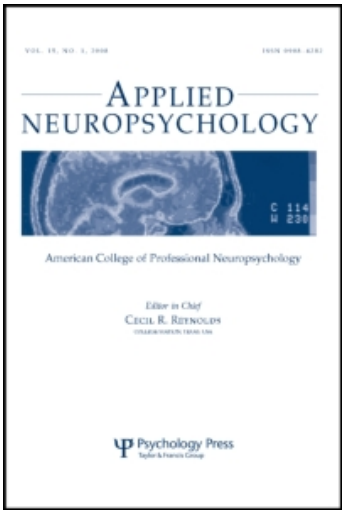

Applied Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775648089

The 'Obsolescence' of Assessment Procedures

Elbert W. Russell

Online publication date: 08 February 2010

To cite this Article Russell, Elbert W.(2010) 'The 'Obsolescence' of Assessment Procedures', Applied Neuropsychology, 17: 1, 60 — 67
To link to this Article: DOI: 10.1080/09084280903297917
URL: http://dx.doi.org/10.1080/09084280903297917

APPLIED NEUROPSYCHOLOGY, *17*: 60–67, 2010

ISSN: 0908-4282 print/1532-4826 online
DOI: 10.1080/09084280903297917

Psychology Press
Taylor & Francis Group

COMMENTARY

The 'Obsolescence' of Assessment Procedures

Elbert W. Russell
Miami, Florida

The concept that obsolescence or being "out of date" makes a test or procedure invalid ("inaccurate," "inappropriate," "not useful," "creating wrong interpretations," etc.) has been widely accepted in psychology and neuropsychology. Such obsolescence, produced by publishing a new version of a test, has produced an extensive nullification of research effort (probably 10,000 Wechsler studies). The arguments, attempting to justify obsolescence, include the Flynn Effect, the creation of a new version of a test or simply time. However, the Flynn Effect appears to have plateaued. In psychometric theory, validated tests do not lose their validity due to the creation of newer versions. Time does not invalidate tests due to the improvement of neurological methodology, such as magnetic resonance imaging. This assumption is unscientific, unproven, and if true, would discredit all older neuropsychological and neurological knowledge. In science, no method, theory, or information, once validated, loses that validation merely due to time or the creation of another test or procedure. Once validated, a procedure is only disproved or replaced by means of new research.

Key words: assessment, Flynn Effect, obsolescence, validation, Wechsler tests

After only 11 years, a "new version" of the Wechsler Adult Intelligence Scale (WAIS), the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; Wechsler, Coalson, & Raiford, 2008), has been published. This is the fifth version of this scale within a period of 70 years. There have also been four versions of the Wechsler Intelligence Scales for Children (WISC) and four versions of the Wechsler Memory Scales (WMS).

In regard to the WAIS, a new version has appeared every 17.2 years (Table 2). These revisions are appearing evermore rapidly. With the exception of the time between the Wechsler-Bellevue (W-B; Wechsler, 1939) and the WAIS (Wechsler, 1955), which was 16 years, the time has reduced between each revision. It was 26 years between the WAIS and the WAIS-Revised (WAIS-R; Wechsler, 1981), 16 between the WAIS-R and the WAIS-Third Edition (WAIS-III; Wechsler, 1997), and finally only 11 between the WAIS-III and the WAIS-IV (Wechsler et al., 2008). The age differences for the versions of the WISC (Wechsler, 1949, 1974a, 1991, 2003) and WMS (Wechsler, 1939, 1974b, 1987, 1997) are comparable.

With each new version, the previous version was generally considered obsolete, such that psychologists eventually had to replace the old version. The appearance of the WAIS-IV brings the concept of sequential test obsolescence to the forefront. This concept of test obsolescence also applies to the WISC, the WMS, and increasingly to other tests such as the Peabody Picture Vocabulary Test, Fourth Edition (Dunn & Dunn, 2007). However, this discussion of the effect of the concept of sequential obsolescence will emphasize the WAIS, since it represents the most obvious example of serial test obsolescence.

Official Acceptance of 'Test Obsolescence'

During the recent history of neuropsychological assessment, the concept that tests become obsolete

Address correspondence to Elbert W. Russell, 6091 SW 79th Street, Miami, Florida. E-mail: ewr@bellsouth.net

with time or when revised has gained general acceptance by psychologists. The requirement that tests remain "up to date" has been incorporated into both the *Ethical Principles of Psychologists and Code of Conduct* of the American Psychological Association (2002) and the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], APA, & the National Council on Measurement in Education, 1999). Both contain warnings that tests should be regularly updated or the test is in effect invalid. The *Standards* (1999) state:

> Standard 4.18.
>
> If a publisher provides norms for use in test score interpretation, then so long as the test remains in print, it is the publisher's responsibility to assure that the test is re-normed with sufficient frequency to permit continued accurate and appropriate score interpretations.
>
> *Comment*: Test publishers should assure that up-to-date norms are readily available, but it remains the test user's responsibility to avoid inappropriate use of norms that are out of date and to strive to assure accurate and appropriate test interpretations. (p. 59)

To the same affect the APA *Ethical Principles* (2002) state:

> *9.08 Obsolete Tests and Outdated Test Results.*
>
> (a) Psychologists do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose.
>
> (b) Psychologists do not base such decisions or recommendations on tests and measures that are obsolete and not useful for the current purpose. (p. 1072)

EFFECT ON NEUROPSYCHOLOGY

Although these rules may seem reasonable, the concept of obsolescence has important problematic implications for both the research and clinical practice of psychologists and neuropsychologists.

Research Studies

These effects are especially consequential for the foundation of neuropsychological assessment—research. The major disadvantage of the concept that obsolescence is created by producing new versions of a test is that this concept commonly nullifies all of the research performed on the earlier versions of a test. Any research results derived from a previous version are considered to be "out of date" and consequently they do not apply to the newer version. While they may in fact apply, there is no direct research evidence that supports the application.

For instance, if the WAIS-III has been demonstrated to be related to brain damage in a certain manner, there is no assurance that this finding applies to the WAIS-IV. As a more complex example, it was known using the W-B Intelligence Scale since 1955 (Reitan, 1955b) that left hemisphere brain damage affects verbal tests more than performance tests for right-handed people. In a 1984 study, it was demonstrated that this effect is not equilateral (Russell, 1984, pp. 81–90). Lateralized damage affects performance tests more than it does verbal tests. This study found that while there must be 20 intelligence quotient (IQ) points between the verbal IQ (VIQ) and the performance IQ (PIQ) to indicate lateralized rather than diffuse damage, the 20 points must be shifted toward the PIQ. That is, the 20 points must be added to the VIQ or subtracted from the PIQ to accurately lateralize the damage. Thus, if the PIQ is 80, a VIQ below 100 indicates possible *left* hemisphere damage. Apparently, this lack of equilateral effects is produced by the greater proportion of fluid intelligence in performance tests than in verbal tests (Barron & Russell, 1992; Russell, 1979, 1980).

While this shift in VIQ and PIQ scores probably is true for later versions of the WAIS, no research studies have demonstrated this shift. In a forensic case that employs a newer version of the WAIS, if such lateralization of scores is submitted as evidence of brain damage, an attorney could point out that there is no assurance this effect still holds.

Another problem is that the continued revision of a test, such as the Wechsler tests, makes certain longitudinal research studies problematic. If the research project continues to use the original versions of a test that has been revised, the research could be attacked as being out of date. The original tests may be difficult to obtain. Finally, file research—in which older material derived from files is compared to more recent material—is almost impossible to accomplish when the tests have been changed. One reason for the discovery of the Flynn Effect (Flynn, 1999) is that the Progressive Matrices (Raven, 1936) on which much of the research was based were not changed every decade or so.

Wechsler Scale Research

To demonstrate the effect of this concept of obsolescence on research, a statistical study was conducted to examine the cost to psychology produced by the loss of research results due to obsolescence. Each time a new version of a Wechsler scale has appeared, all of the previous research utilizing the older version of the tests was basically invalidated since it was "out of date." While in the theoretical area, findings for the W-B may continue to have some relevance (Loring & Larrabee,

2006; Russell, 2008), in forensics, only the findings from research derived from the most recent version of the test would be considered reliable. For findings from a previous version, the opposing attorneys can simply indicate that according to APA standards and ethics these findings were derived from an obsolete test and so are no longer reliable.

To document this loss of test research caused by continually changing the major Wechsler scales, the number of studies derived from each version of the Wechsler test is calculated in Table 1. The table presents the number of studies produced for each version of the WAIS, the WISC, and the WMS, prior to the most recent version. These data were derived from APA's PsycINFO, which provides the number of times a scale was listed under several categories. The category "key words" was adopted since this term is ordinarily applied only when a test is used in a study. Since different key words may be used for the same test such as "WMS" or "Wechsler Memory Scale," the footnotes indicate which key words were used.

TABLE 1
Studies Completed with Wechsler Scales, Compared to Two Other Batteries, Using PsycINFO Key Words

Tests		*Brain*[a]	*HRB*[b]	*RPM*[c]
W-B[d]	593	38		
WAIS[e]	2,607	200		
WAIS-R[f]	814	56		
WAIS-III[g]	231	7		
WAIS Total	4,245	245		
WISC[h]	3,149	54		
WISC-R[i]	1,324	11		
WISC-III[j]	428	1		
WISC Total	4,901	66		
WMS[k]	535	55		
WMS-R[l]	146	13		
WMS-III[m]	73	3		
WMS Total	754	71		
Grand Total	9,900	439	383	726

Notes. In combining key words used in PsycINFO, "or" means either word may be included while "and" means both are required.

[a]Key words included: "and brain damage" after each key word.

[b]Key words included: Reitan or Halstead Reitan, or HRB.

[c]Key words included: Progressive Matrices or Raven's Progressive Matrices.

[d]Key words used: W-B or Wechsler-Bellevue.

[e]Key words used: WAIS or Wechsler Adult Intelligence Scale.

[f]Key words used: WAIS-R or Wechsler Adult Intelligence Scale-Revised.

[g]Key words used: WAIS-III or Wechsler Adult Intelligence Scale-III.

[h]Key words used: WISC or Wechsler Intelligence Scale for Children.

[i]Key words used: WISC-R or Wechsler Intelligence Scale for Children-Revised.

[j]Key words used: WISC-III or Wechsler Intelligence Scale for Children-III.

[k]Key words used: WMS or Wechsler Memory Scale.

[l]Key words used: WMS-R or Wechsler Memory Scale-Revised.

[m]Key words used: WMS-III or Wechsler Memory Scale-III.

An examination of this table indicates that over the 69 years, from the W-B to the WAIS-IV, the WISC-IV, and the WMS-IV, 9,900 Wechsler studies derived from older versions of these tests were conducted. The number of items under the "key word" listing indicates a minimal inclusion since studies were undoubtedly missed. This means that there may well have been 10,000 Wechsler studies nullified by the production of new versions. Thus, almost 143 studies per year have been assigned to oblivion. Of course, the number of hours devoted to each study is unknown, but certainly each study represents a large amount of work.

The column using the code, "and brain damage," indicates the minimal number of these studies that were neuropsychological studies. However, there are so many alternative key words to "brain damage," such as "head trauma," "brain injury," "neuropsychology," etc., that these figures probably represent less than half the neuropsychological studies that have been eliminated. Thus, this habit of psychology and neuropsychology, of nullifying its past research by making previous tests invalid, indicates an enormous waste of assessment and research resources.

By contrast, the research completed with the Halstead-Reitan Battery (HRB), designed around 1955 (Reitan, 1955a), produced 383 studies for the key words of "Reitan or Halstead-Reitan or HRB." The Raven's Progressive Matrices (RPM), created in 1936, produced 725 items for the key word "Progressive Matrices." The irony is that *all* of these studies are still considered valid, since these tests have not been changed and no research has falsified them. Thus, the research expended on these tests has not been wasted.

Clinical Practice

The requirement for tests to be up to date is also an obstacle to neuropsychology practice, at least for interpretations that use these "obsolete" test batteries. That is, a new version of a test will largely nullify the experience as well as research derived from the previous versions of the test. Many of the interactions established between the various subtests and indexes, both within the batteries and between them and other tests, become inapplicable. As such, part of the experience and research basis of a neuropsychologist's skill will have been lost with each new version of a test.

In this situation, neuropsychologists in hospitals and private practice often ignore, at least for a time, the new versions. However, in forensic practice, obsolescence is a major argument that attorneys can use to attack previous versions of the Wechsler and other assessment batteries, since "obsolete" tests are "outlawed" by APA

ethics and standards. This, of course, forces forensic psychologists to continually buy the latest version of the test.

The various effects of new versions of intelligence tests on clinical practice are discussed in a recent paper on the Flynn Effect (Kanaya, Scullin, & Ceci, 2003; Russell, 2007). For instance, schools use intelligence tests to determine special placement, and courts may determine whether to impose capital punishment based on the IQ cutting score for mental retardation. These may be changed by a new version of a test.

Hindrance to Neuropsychological Development

Finally, this concept of obsolescence appears to have produced a form of damage to neuropsychology that is not immediately obvious. This is a hindrance to innovation and the further development of scientifically accurate neuropsychology assessment methods. As long as the same tests are rapidly revised, there will probably be a reduced attempt to study the implications of these existing tests. This supposition is supported by the finding that the number of studies devoted to each version of the WAIS has declined from 100 per year, for the WAIS, to 50 for the WAIS-R, to 21 for the WAIS-III.

JUSTIFICATIONS FOR AN OBSOLESCENCE DESIGNATION

Considering the potential harm that the rapid creation of new versions of tests appears to be doing to neuropsychology, the arguments used to justify the creation of new versions of various scales and for designating any test as obsolete need to be examined.

The Flynn Effect

A major argument is derived from the Flynn Effect, which was discovered in the late 1980s (Flynn, 1999). The effect was that the average intelligence of various populations in the Western world was increasing fairly rapidly. Flynn calculated an increase in IQ scores of about .3 points per year (Flynn). This was consistent with the difference between the WAIS and WAIS-R, which was .29 (Russell, 2007)—almost .3. Thus, this effect certainly existed since this discovery was based on a relatively large number of good studies.

In 1981, when the WAIS-R (Wechsler, 1981) was published, this effect was not recognized in the test manual. Rather, this new version of the WAIS was justified in terms of content improvements in the individual subtests (Wechsler, 1981, pp. 11–15, 36–48). Although, when the WAIS-R was re-normed, the differences between the scale scores and IQs of the two versions were noted (pp. 47–48). These differences were neither emphasized nor used to justify creating a new version of the scale.

However, by 1997 after the Flynn Effect had been accepted, the WAIS-III manual (Wechsler, 1997) stated that: ''Because there is a real phenomena of IQ-score inflation over time, norms for a test of intellectual functioning should be updated regularly . . . Regardless of the reasons for these changes in test performance, periodic updating of the norms is essential . . . '' (pp. 8–9).

This justification, based on the Flynn Effect, was so thoroughly accepted that both the APA *Ethical Principles* (2003) and the *Standards* (AERA et al., 1999) contained warnings that tests should be updated regularly or they would essentially be invalid. Although stating that using obsolete tests is unethical, the APA *Ethical Principles* (APA, 2002) makes no attempt to define obsolescence except to state that obsolescence meant the tests ''are outdated for the current purpose'' (9.08, a & b).

The *Standards* (AERA et al., 1999) are more detailed. They mention criteria for obsolescence in two places. First, Standard 4.18 states '' . . . that the test is re-normed with sufficient frequency to permit continued accurate and appropriate score interpretations . . . (p. 59)''.

This criterion is made more specific in Standard 11.16, which states:

> Test users should verify periodically that their interpretations of test data continue to be appropriate given any significant changes in their population of test takers, their modes of test administration, and their purposes in testing.
>
> *Comment*: Over time, a gradual change in the demographic characteristics of an examinee population may significantly affect the inferences drawn from the group averages. (p. 117)

With this comment, Standard 11.16 directly implies that the Flynn Effect is a criterion for obsolescence. No other general population effect has been reported.

However, in the last few years, it is becoming evident that the influence of the Flynn Effect may be attenuating. A recent commentary (Russell, 2007) describes how studies compiled from several countries indicated that the Flynn Effect appears to have plateaued for these advanced Western countries. The few studies that have directly examined the Flynn Effect since Russell's study have all found evidence of this plateauing effect (Beaujean, 2007; Shayar, Ginsburg, & Coe, 2007; Teasdale & Owen, 2008).

In regard to the WAIS, data from the WAIS, WAIS-R, and the WAIS-III presented evidence that the rate of increase in the Wechsler IQ was decreasing. and the increase would have either already ceased or would cease within the next 14 years (Russell, 2007).

TABLE 2
Wechsler Adult Intelligence Test Studies

Scales	*Years*	*Years Between Versions*	*FSIQ Increase*	*% IQ Increase per Year*
W-B[a]	1939–1955	16		NA
WAIS	1955–1981	26	7.5	0.29
WAIS-R	1981–1997	16	2.9	0.18
WAIS-III	1997–2008	11	2.9	0.26
WAIS-IV	2008			NA
Total	1939–2008	69 (17.2)[b]		

Notes. Data was derived from the test manuals.
[a]Wechsler-Bellevue.
[b]Mean number of years between the WAIS versions.

These studies seriously undermine the argument that the Flynn Effect still requires the re-norming of a test every decade.

In contrast, the present WAIS-IV manual (Wechsler et al., 2008) presented a comparison of the WAIS-IV and WAIS-III in which the reduction in the IQ levels due to an increase in intelligence was 2.9 for 11 years, or .26% per year (Wechsler et al., pp. 75–77). This appears to be relatively consistent with the initial prediction of the Flynn Effect in the 1980s, which was a .3% increase per year.

However, as indicated in Table 2, the rate of Full Scale IQ (FSIQ) increase for the WAIS had dropped to .18% per year in the 16 years between the WAIS-R (1981) and the WAIS-III (1997). This decrease was demonstrated in a paper on the Flynn Effect (Russell, 2007, p. 265). Consequently, it is somewhat surprising that after falling for 16 years, the rate of improvement would suddenly increase to .26% per year between the WAIS-III (1997) and WAIS-IV (2008). The authors of the WAIS-IV do not discuss or try to explain this phenomenon, although they present a list of all the influences that might affect the Wechsler norming (p. 75).

Effect of Screening Subjects on the Wechsler Tests

In this regard, it should be noted that the WAIS-IV volunteer subjects were vigorously screened (Wechsler et al., 2008, pp. 30–32). Such screening of the volunteers for the WAIS-IV undoubtedly affected the population representation of the sample. For instance, after screening his volunteers for any evidence of possible brain damage or a psychotic condition, Dodrill (1987, pp. 2–3) found that he had eliminated approximately one-third of his original group.

The only reported statistical study of the rate of subject loss due to the selection procedure was derived from a study using older normal subjects (Fastenau, 1998). In this study, 22.5% (over one-fifth) of the randomly selected patients were rejected because of possible neurological problems. Thus, while most norming studies using "volunteers" do not record their rejection rate, these examples demonstrate the rate is evidently quite significant.

In support of the possibility that selective bias has produced abnormally high volunteer norms are the findings of a recent study (Russell, 2005). When all of the neuropsychological norming studies of the HRB using volunteers and providing IQ levels in the literature were compared to the norms derived from the WAIS and WAIS-R, the volunteer norms were, without exception, almost a standard deviation above the WAIS and WAIS-R average. All of these volunteer studies had removed those subjects who had any possible type of brain damage or psychiatric condition. Since there was no indication from their manuals that the WAIS and WAIS-R had systematically excluded various types of subjects, their samples evidently did represent an unbiased sample of the general population. Thus, as a consequence, the volunteer norms for the neuropsychological tests examined in this study (Russell, 2005) did not represent an unbiased sample of the U.S. population, but rather they represented an above-average group.

The volunteer norms of the WAIS-III (Wechsler, 1997, p. 22) and the WAIS-IV (Wechsler, et al., 2008, p. 31) used a large number of criteria to exclude subjects from their samples. As such, it is quite possible that these highly selected norms represent an above-average sample rather than an unbiased random sample of the country's population. Consequently, the WAIS-IV and the WAIS-III samples are evidently no longer appropriate for any populations that contain the type of subjects that were eliminated. This includes the general population of the United States. For populations containing these subjects the WAIS-R, norms are evidently still appropriate.

Obsolescence Due to a New Version

Another argument for assigning obsolescence of a test is the assumption in psychology that a new version of any test renders, ipso facto, the older version obsolete. There is, of course, no scientific evidence for this, any more than the construction of any test in itself renders any other test obsolete.

The transformation and re-norming of the W-B (Wechsler, 1939) was justified since this created the WAIS (Wechsler, 1955) that incorporated needed improvements in the test design and obtained a representative set of norms. After this, the WAIS was not changed for 26 years.

If the indexes of a new version are as highly correlated with the old version as the test-retest correlation of the new test, then the two tests or subtests are psychometrically interchangeable. This is a standard

psychometric concept that applies to any two tests, not simply versions of a particular test. Research findings concerning the previous version psychometrically apply to the new version. For most assessment purposes, the old version remains equivalent to the new version, so there is no need to change scales. As such, the validated research performed on the old version remains valid for the new version.

On the other hand, substantial changes to the original test, such as adding several new subtests to a scale or battery or discontinuing subtests, changes the overall battery so that there is no assurance it is measuring the same patterns found to exist using the earlier battery. In this regard, evidently, there are no studies that directly demonstrate that the WAIS-IV is any more accurate or valid than the WAIS-R or WAIS-III or that they are even measuring the same group of mental components.

Improved Neurological Diagnosis

Another major argument for the obsolescence of a method, norms, or information is that neurological diagnostic methods have improved, particularly the establishment of magnetic resonance imaging (MRI), to such an extent that preceding theories or norms are no longer valid. Apparently, critics who use this argument believe that before the MRI, neurological diagnostic tools were too crude to produce an accurate diagnosis. This argument implies that previously no valid neuropsychological research including creating norms existed. Since no proof that the older norms are invalid is ever forthcoming, this criticism is predicated on faith, not on science.

In addition, this argument has an even more serious implication. If true, it would consign more than half a century of neuropsychological work to oblivion. More critically, it also means that three centuries of neurological research would be largely worthless. As such, all of the neurological findings in the "before the MRI era," such as the work of Broca, Wernicke, William James, Luria, Head, Hebb, Teuber, Werner (1956) etc., would be worthless.

In contrast, neuropsychologists who did not work with neurologists before the development of scanning techniques continually fail to take into consideration that along with many accurate neurological methods, such as neurosurgery, angiograms, and autopsies, the most accurate basis of diagnosis was as fully applicable then as today. This method is "time." Almost all progressive lesions eventually manifest themselves and can ultimately be accurately diagnosed, even at autopsy.

Finally, scanning techniques, at present, are still unable to diagnose many conditions, such as Alzheimer's disease, certain vascular pathologies, or closed-head trauma, unless it involves a hemorrhage (Bigler, 1991).

SCIENCE AND OBSOLESCENCE

On examination, it is apparent that none of these arguments is viable. The criticism that a test is "out of date" almost always means that the critic is committing a fallacy, which permeates neuropsychology. The fallacy is that methods or concepts become invalid simply due to the passage of time.

However, science does not work that way. Authentic science is not a fad that changes with every new general accepted notion. In science, time alone is not a factor that invalidates validated methods, theories, or information. That is, once validated, methods and their products only become obsolete or "out of date" by further research studies that disconfirm the method and its products. Thus, a pronouncement that criticizes a procedure or concept as being inadequate due to age must be referenced with the relevant studies that demonstrate the inadequacy. In science, just as validation studies are initially required to confirm a method or concept, disconfirming studies are necessary to label the method as invalid due to obsolescence. As an example, in medicine, recent years have brought enormous improvements in methodology, such as the MRI; however, the stethoscope, which was invented in 1816, is 192 years old. ("The Stethescope," 1998–2006) The version that was developed around 1851, 150 years ago, is still in common use. It is certainly not obsolete in spite of its age.

The 1985 Standards

In this regard, it is interesting to compare the wording of the 1985 *Standards for Educational and Psychological Testing* (AERA, APA, & National Council on Measurements in Education, 1985), concerned with re-norming, to the wording provided above in the 1999 *Standards* (AERA et al., 1999). It should be noted that in the 1985 *Standards,* there is no comment concerning the obligation of test users to only use up-to-date tests. The 1985 wording is as follows:

> Standard 3.18.
>
> A test should be amended or revised when new research data, significant changes in the domain represented, or new conditions of test use and interpretation make the test inappropriate for its intended uses. An apparently old test that remains useful need not be withdrawn or revised simply because of the passage of time. But it is the responsibility of test developers and test publishers to monitor changing conditions and to amend, revise, or withdraw the test as indicated. (p. 29)

This wording is consistent with the scientific concept of the effect of age on scientific methods, theories, and information. In contrast the recent psychological standards and ethics may support the authenticity of fads.

Some Implications for Neuropsychology and Psychology

This scientific concept that the designation of obsolescence requires disconfirming research studies has several implications for neuropsychology and psychology. First, neuropsychology should recognize that the research, including norming, based on a validated test or battery remains valid until it has been demonstrated to be invalid. Regardless of how long ago a validated test was published, its results are still sound unless research has demonstrated its lack of validity.

An older version of a test that essentially measures the same brain function often has advantages over a new version, especially in regard to the body of relevant research that has accumulated for the older version. Until the superiority of the new version has been demonstrated by relevant research, the older version retains its validity and accuracy. Thus, for most assessments and especially for forensic purposes, it may be more advantageous to retain the older version than adopt a new less researched version.

As an illustration, although the HRB is more than 50 years old, it is not obsolete since no studies have ever demonstrated its lack of validity or inadequacy. Time and again, the argument of obsolescence has been used to attack the HRB (Bigler, 2007; Lezak, Howieson, & Loring, 2004, pp. 670–676; see also Russell, 1998). However, the validity of the HRB has been demonstrated throughout its history (Russell, 1995). Since 1995, other studies have continued to repeat this validation, such as the recent study by Horwitz, Lynch, McCaffrey, and Fisher (2008). At present, there is no other battery methodology in neuropsychology that is so well validated. In contrast, none of the nonstandardized methods for using a battery of tests have been validated.

In addition, even in regard to changing norms, a new version of a test need not be constructed in order to update the norms. New norms may be obtained using the original test and published without creating a new version of a test.

Another implication is that there are reasons for using alternative intelligence tests in neuropsychology that are sufficiently compelling to designate the Wechsler tests as inadequate. For instance, the Kaufman tests such as the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2004), are designed to manage the problem of confounding lateralization and fluidity. The Wechsler tests lack a distinction between lateralized and fluid functions. As such, the scale scores confound the two forms of intelligence (Barron & Russell, 1992; Russell, 1979, 1980).

Secondly, many if not most of the major alternative intelligence scales combined the adult and children's versions in such a manner that an adult with brain damage will simply receive a score in what is normally the children's range. These tests include the Kaufman Brief Intelligence Test, Second Edition (Kaufman & Kaufman, 2004), Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003), and the Stanford-Binet Intelligence Scales, Fifth Edition (Roid, 2003), along with others.

The Wechsler tests separate the child's versions from the adult versions, thus making the transition from the child's version to the adult version difficult and unreliable. Combining these tests would overcome a major neuropsychological problem for the WAIS, which has been recognized for many years.

In addition, the scales for the WAIS do not have a sufficiently extended lower range to adequately measure a person's functioning when they have fairly severe brain damage. In order to partially deal with this problem, the WAIS-III and WAIS-IV inverted the customary direction of the subtest scales so that the top of the scale represents the worst score. This change reverses the psychometric convention for scale direction. It is often confusing and is always clumsy. In spite of this odd reversal, the adult test is still not sensitive to severe brain damage.

CONSEQUENCES OF THE OBSOLESCENCE DESIGNATION

As this discussion has demonstrated, the problem with retaining the presently accepted concept that age alone or revision of a test can assign the term ''obsolescence'' to a procedure or concept is that this is unscientific and constitutes a considerable impediment to the fields of neuropsychology and psychology. As illustrated by the effects discussed above, it is evident that psychology's accepted definition of obsolescence has annihilated the usefulness of thousands of research studies and the associated experience and skills of clinicians. As such, this definition appears to have been a significant obstacle to the development of neuropsychology. If the tests used in neuropsychology are continually being designated as obsolete and consequently abandoned, neuropsychology will have difficulty creating a substantial body of scientific knowledge based on assessment procedures.

As presently employed, the concept of obsolescence or being ''out of date'' needs to change to be compatible with scientific assessment, methodology, and theory construction. This requires a re-examination of cognitive

measurement practices along with the ethical codes and standards related to obsolescence that are expressed in the APA manuals.

REFERENCES

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct (5th ed.). *American Psychologist*, *47*, 1060–1073.

Barron, J. H., & Russell, E. W. (1992). Fluidity theory and the neuropsychological impairment in alcoholism. *Archives of Clinical Neuropsychology*, *7*, 175–188.

Beaujean, A. A. (2007). Using item response theory to assess the Lynn-Flynn effect. *Dissertation Abstracts International*, *68*(3-B), 1978.

Bigler, E. D. (1991). Neuropsychological assessment, neuroimaging, and clinical neuropsychology: A synthesis. *Archives of Clinical Neuropsychology*, *6*, 113–132.

Bigler, E. D. (2007). A motion to exclude and the 'fixed' versus 'flexible' battery in 'forensic' neuropsychology: Challenges to the practice of clinical neuropsychology. *Archives of Clinical Neuropsychology*, *22*, 45–51.

Dodrill, C. B. (1987). *What constitutes normal performance in clinical neuropsychology?* Paper presented at the 97th Annual Convention of the American Psychological Association, Atlanta, Georgia.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson Clinical Assessments.

Fastenau, P. S. (1998). Validity of regression-based norms: An empirical test of the comprehensive norms with older adults. *Journal of Clinical and Experimental Neuropsychology*, *20*, 906–916.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5–20.

Horwitz, J. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2008). Screening for neuropsychological impairment using Reitan and Wolfson's preliminary neuropsychological test battery. *Archives of Clinical Neuropsychology*, *23*(4), 393–398.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn Effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist*, *58*(10), 778–790.

Kaufman, A., & Kaufman, N. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Los Angeles: Western Psychological Services.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford.

Loring, D. W., & Larrabee, G. J. (2006). Sensitivity of the Halstead and Wechsler Test Batteries to brain damage: Evidence from Reitan's original validation sample. *The Clinical Neuropsychologist*, *20*(2), 221–229.

Raven, J. C. (1936). Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. Unpublished master's thesis, University of London.

Reitan, R. M. (1955a). An investigation of the validity of Halstead's measures of biological intelligence. *Archives of Neurology and Psychiatry*, *73*, 28–35.

Reitan, R. M. (1955b). Certain differential effects of left and right cerebral lesions in human adults. *Journal of Comparative and Physiological Psychology*, *48*, 474–477.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales* (RAS). Los Angeles: Western Psychological Services.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside Publishing.

Russell, E. W. (1979). Three patterns of brain damage on the WAIS. *Journal of Clinical Psychology*, *35*, 611–620.

Russell, E. W. (1980). Fluid and crystallized intelligence: Effects of diffuse brain damage on the WAIS. *Perceptual and Motor Skills*, *51*, 121–122.

Russell, E. W. (1984). Theory and developments of pattern analysis methods related to the Halstead-Reitan battery. In P. E. Logue & J. M. Shear (Eds.), *Clinical neuropsychology: A multidisciplinary approach* (pp. 50–98). Springfield, IL: Charles C. Thomas.

Russell, E. W. (1995). The accuracy of automated and clinical detection of brain damage and lateralization in neuropsychology. *Neuropsychology Review*, *5*, 1–68.

Russell, E. W. (1998). In defense of the Halstead-Reitan Battery: A critique of Lezak's review. *Archives of Clinical Neuropsychology*, *13*, 365–381.

Russell, E. W. (2005). Norming subjects for the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology*, *20*(4), 479–484.

Russell, E. W. (2007). The Flynn Effect revisited. *Applied Neuropsychology*, *14*(4), 262–266.

Russell, E. W. (2008). Commentary on Loring and Larrabee's reanalysis of Reitan's original studies. *The Clinical Neuropsychologist*, *22*(3), 519–523.

Shayar, M., Ginsburg, D., & Coe, R. (2007). 30 years on a large anti-Flynn Effect? The Piagetian test volume & heaviness norms space, 1975–2003. *British Journal of educational psychology*, *77*(1), 25–41.

Teasdale, T. W., & Owen, D. R. (2008). Secular declines and cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, *36*(3), 121–126.

The stethoscope. (1998–2006). *Medical Antiques Online* [Electronic version]. Retrieved January 2009.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale manual*. New York: Psychological Corporation.

Wechsler, D. (1974a). *Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.

Wechsler, D. (1974b). *Wechsler Memory Scale*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *WAIS-III, WMS-III Technical manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Third Edition manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). *WAIS-IV technical and interpretive manual*. San Antonio, TX: Pearson.

Werner, H. (1956). Microgenesis and aphasia. *Journal of Abnormal and Social Psychology*, *52*, 347–353.