The Wechsler Adult Intelligence Scale-Revised in Clinical Neuropsychology Practice

Robert G Knight
University of Otago

Intellectual assessment of clients with suspected or known brain damage using the WAIS-R is common practice in neuropsychology. However, such assessments are being increasingly subjected to critical review, particularly in courtroom settings. This means that psychologists may be challenged about the validity of their testing and interpretation strategies, often with well-informed reference to the published literature. This necessitates that clinicians who use the WAIS-R and present formal reports on their findings are well-versed in the credentials of the test.

In this article, a broad range of studies that bear on the usefulness of the WAIS-R in neuropsychological practice are reviewed. Neuropsychologists are well advised to be cautious in the interpretation of WAIS-R profiles and to refer to base rate data before expressing opinions about abnormal performance. It is also important to ensure that WAIS-R findings are well-integrated with a client's neurological and life history data. Although the courtroom is not the best environment in which to determine the testing strategies, the prospect of such scrutiny may at least stimulate psychologists to maintain a healthy skepticism about their tests and measures.

A central part of any neuropsychological assessment is the evaluation of the client's ability to reason, think, and judge, mental processes grouped together under the rubric of intelligence. No measures of intelligence have been more successful in finding widespread use in clinical practice than those developed by David Wechsler. The Bellevue, which appeared in 1939, and the subsequent Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), with the supporting monograph *Measurement of Adult Intelligence* (Wechsler, 1939, 1958), found immense favour with clinical practitioners. The WAIS-Revised (WAIS-R) continued the tradition of test construction that emphasised the collection of norms based on a representative national sample, careful standardisation, and well-established psychometric credentials. The significance of the WAIS and WAIS-R in adult neuropsychology is attested to by their prominence in reviews provided by both proponents of neuropsychological testing (Lezak, 1994; Crawford, Parker, & McKinley, 1992) and skeptics (Ziskin & Faust, 1988).

Although testing occupies a central place in the practice of psychology, this has not gone without challenge. Indeed there is evidence that the issues that face the psychological testing industry today, were debated by the Chinese over 2000 years ago (Bowman, 1989; Matarazzo, 1990). In the mid 1960's in the United States, attacks on testing as an unscientific invasion of privacy culminated in two Congressional inquiries, stimulated by reports that the MMPI was being misused as a means of selecting Peace Corps volunteers. The case against testing and the defence mounted by the American Psychological Association (APA), were reported in detail in the *American Psychologist* of 1965. In the subsequent 30 years, concern about the improper use of testing in the public education system and in personnel selection in the US has resulted in legislative regulation of testing in some States. An important outcome of attacks on testing has been to focus on the accountability of testers and to stimulate the development of functional or skill-based achievement tests, particularly in the education system.

Much of the criticism of testing in human resources or educational settings has been focused on the usefulness of testing in the process of decision making. Until recently, the use of tests in clinical practice with individual clients has largely escaped public scrutiny. This is primarily because, as Matarazzo (1990) has contended, the administration of tests in clinical practice has typically formed part of an assessment procedure designed to determine how psychological treatment might benefit the client. Decisions taken in the course of working with a client that arise from

formal and informal assessment are usually embedded in an ongoing relationship between clinician and client governed largely by professional guidelines and expectations. All this is changing. Increasingly, clinicians are testifying in the courtroom or in other judicial arenas as the custody of children is decided or claims for compensation addressed. The written opinions of clinical psychologists or neuropsychologists are increasingly subject to sophisticated examination by lawyers with access to the advice of experts on challenging the credibility of the tests and measures psychologists use. The consequences are graphically described by Matarazzo (1990, p. 1002):

> The effect has been that increasing numbers of us who practice clinical psychology and clinical neuropsychology (whose knowledge of the bases for the reliability and validity of the most frequently used psychological instruments, including the clinical interview, is usually dated) have had to return to the library in order to better prepare answers to the most searching questions we have been asked since the days we suffered through our doctoral or final oral examinations.

In their quest to undermine psychological testimony, lawyers have recourse to Ziskin and Faust's (1988) three volume work *Coping with psychiatric and psychological testimony*. One aim of these texts is to instruct lawyers involved in the examination of psychologists and neuropsychologists about selected research studies and professional opinions that describe the limitations of psychological assessment strategies. As critics have observed (e.g., Adams & Putnam, 1994; Matarazzo, 1990, 1991; Rogers, Bagby, & Perera, 1993), these texts are not a dispassionate, integrative, or scholarly review of the literature. As Faust, Ziskin, and Hiers (1991) candidly acknowledge, their reviews are selective and contain only material that might undermine the credibility of a psychologist's testimony:

> There is literature which is not contained in this book that is supportive of forensic neuropsychology, although for some topics this literature is slim to nonexistent (Faust et al., 1991, p.xv).

The specific aim of the present article is to examine the credentials of the WAIS-R as a tool for neuropsychological assessment. The objective is to survey the psychometric foundations of the WAIS-R and the way results from the test can be used in constructing evidence relevant to determining neuropsychological impairment. This review is motivated by the concerns of Matarazzo and others, who have expressed the need to revisit psychometric issues as a consequence of more informed scrutiny of neuropsychological assessment tools in the public domain. In doing so, it is important to be mindful that all measurements made in the physical and social sciences are prone to error, and are neither perfectly reliable nor completely valid under all circumstances; psychological tests are not exceptional in being subject to error. Refinements in measurements in physics, medicine, and chemistry, for example, have been ongoing for centuries. In any sphere of science, a measurement that can be made now is almost certain to be less accurate than one that will be made in the future. This consideration is unlikely to be much consolation in the witness box, but it does help focus on what is most important, that clinical neuropsychologists learn to appreciate the limitations of what they can achieve currently with the tools to hand, and work to construct better tools for the future. In this endeavour, the probing of skeptics has a part to play in ensuring that the highest standards of test construction and professional practice are observed. We find ourselves, however, in complete agreement with the Adams and Putnam's sentiments, when they conclude that:

> As incomplete and developing as scientific neuropsychology is (as with any science), in the balance, even the most critical of neuropsychologists find the prospect of hard work at the tasks of science preferable to nihilism for the benefit of legal advocates who may find pseudoscientific despair appealing in the adversarial context (Adams & Putnam, 1994, p.6).

The Wechsler Adult Intelligence Scale - Revised

The Wechsler Intelligence scales, with their multiple subtests (with scaled means of 10 and standard deviations of 3) and IQ scale scores (with a mean of 100 and a standard deviation of 15), are amongst the most widely used tests in clinical neuropsychology practice. Wechsler developed his scales to measure "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively" (Wechsler, 1939, p.3) with the external environment. His objective was to compile a battery of tasks that in combination would quantify general mental ability. He was influenced by Spearman's 2-factor theory of intelligence, acclaiming this model as one of the foremost discoveries in psychology (Wechsler, 1939). It is important to emphasise that the Wechsler scales were not constructed, nor the subtests chosen, with their use as neuropsychological tests in mind:

> ... so far as measuring intelligence is concerned, these specific tasks are only a means to an end. Their object is not to test a person's memory, judgement, or reasoning ability, but to measure something which it is hoped will emerge from the sum total of the subject's performance, namely, ...general intelligence. (Wechsler, 1939, p.6).

Nonetheless, Wechsler soon found interest in examining the pattern of subtest scores from his neuropsychiatric patients. Reporting on the test scores of various clinical samples, he observed that patients with organic brain damage had higher Verbal IQ (VIQ) than Performance IQ (PIQ) scores, with preserved performance on the Information and Comprehension subtests, and especially low scores on Digit Span, Block Design, and Digit Symbol (Wechsler, 1941). He also drew attention to the uneven pattern of deterioration in subtest scores as a function of age (Wechsler, 1944), characterising those that were impervious to aging effects as "Hold" tests, and those that declined as "Don't Hold". Generations of clinicians thereafter have found themselves irresistibly drawn to examine and speculate about patterns of subtest scores (Hunt, 1949; McFie, 1975).

At the time when Wechsler constructed the Bellevue, the primary work role of clinical psychologists was the testing of patients in large public psychiatric institutions. Only in comparatively recent times have psychologists been rou-

tinely involved in the assessment of persons with neurological disorders in general hospitals; widespread professional practice in clinical neuropsychology is a phenomenon of the last two decades. When psychologists began working with brain damaged clients, they began by using the tests that were most familiar to them in their psychiatric practice; thus the Wechsler scales found a place in their armamentarium of assessment tools. With time, the use of the WAIS-R as a neuropsychological test became increasingly important. There were several reasons for this. First, the test was relevant to the deficits that clinicians needed to calibrate. Although the validity and sensitivity of the WAIS-R can be challenged in some circumstances, a deterioration in cognitive functioning is a characteristic of many neurological conditions, and there is no doubt that the WAIS-R can detect this. A matched sample of dementing persons with Alzheimer's disease will invariably have a lower average IQ than a matched sample of healthy people (e.g., Randolph, Mohr, & Chase, 1993). Second, the structure of the scale encouraged comparison of scaled subtest scores as a means of establishing an individual client's preserved and deficient performance. The opportunity to engage in this kind of decomposition of test scores had considerable appeal for neuropsychologists concerned to quantify deterioration or deficit on rational, statistical grounds. Before moving on to consider the clinical applications of the scale that arose from this approach, it is perhaps instructive to first review the psychometric properties of the subtests and scale scores.

Psychometric bases of the WAIS-R

Norms

The WAIS-R was standardized on a sample of 1880 Americans aged from 16 to 74 during the period 1976-1980. Norms from a group of 130 elderly volunteers, aged between 75 and 79 years (n=60) and 80+ years (n=70), living in Kansas, Missouri, and Iowa have been prepared by Ryan, Paolo, and Brungardt (1990a) to supplement those provided by Wechsler (1981). The WAIS-R test items have been modified for use in other countries, including New Zealand (Petrie, Dibble, Long-Taylor, & Ruthe, 1986), the United Kingdom (Smith, 1988), and Australia (DeLemos, 1983). The WAIS-R has not been systematically renormed outside the United States, so reliability, difficulty levels, and validity must be assumed to be similar to those established in the United States. Thus, at the outset, it must be acknowledged that the use of the WAIS-R in New Zealand requires the user to accept the credentials of a test normed in the United States. The psychometric characteristics of the test in New Zealand are largely unknown, and may be especially problematic with some ethnic and cultural subgroups.

The introduction of the WAIS-R made the use of the WAIS obsolete. This has created problems for neuropsychologists retesting clients previously tested with the WAIS, since a discrepancy averaging 7 to 8 points between WAIS and WAIS-R IQ scores has been repeatedly demonstrated (e.g., Field & Sisley, 1986; Prifitera & Ryan, 1983; Urbina, Golden, & Ariel, 1982; Wechsler, 1981). Although tables for converting WAIS scores to WAIS-R IQs have been published (e.g., Russell, 1992), such procedures should be used cautiously. In addition, it can not be assumed that relationships between the WAIS and demographic variables or other neuropsychological test scores will generalise to the WAIS-R. For example, the National Adult Reading Test (NART) was constructed to predict WAIS IQ scores and so WAIS-based regression formulas from the NART may or may not predict WAIS-R IQs accurately. Similarly, Matarazzo and Herman (1984a) found that the correlation between total number of years of schooling completed and IQ for the WAIS-R standardization was lower (.62, .63) than similar correlations reported for the WAIS (which average .70).

Another issue, which Ziskin and Faust (1988) draw to the attention of their audience, is the obsolescence of norms. This is the subject of extensive research by Flynn (1984), who has shown a gradual increase in average IQ scores over time, suggesting that norms become obsolete. Furthermore, there is evidence that this process of score inflation is not consistent and is more noticeable on Performance than Verbal subtest scores (Crawford, 1992). This potentially raises difficulties for the interpretation of subtest differences as test norms begin to age.

Reliability

The validity of a test and its usefulness in detecting deficits is constrained in large part by its reliability. By any reasonable standard, the WAIS-R IQ scales are highly reliable. The average split-half reliabilities for VIQ (.97), PIQ (.93), and FSIQ (.97), estimated for the standardisation sample, are as high as for any other measure commonly used in neuropsychology, and far higher than most. The manual provides reliability coefficients for the subtests, which confirm that the individual subtest scores are highly reliable (median = .81), with the exception of Picture Arrangement and Object Assembly, which have more modest reliability estimates. The WAIS-R has also been shown to be reliable at the upper end of the age range (Ryan, Paolo, & Brungardt, 1992; Snow, Tierney, Zorzitto, Fisher, & Reid, 1989).

The reliability of the WAIS-R in clinical samples has been addressed in several studies. Boone (1992a) reported test-retest coefficients in excess of .80 on all subtests, except Object Assembly, for 100 psychiatric inpatients (50% schizophrenic, 29% affective disorders) tested twice, 2-8 weeks apart. Ryan, Georgemiller, Gleisser, and Randall (1985) reported high test-retest coefficients (median = .75) for 21 male patients with neurological or psychiatric disorders tested twice, an average of 38 weeks apart. Split-half reliabilities for a similar mixed sample of patients were reported by Ryan, Prifitera, and Larsen (1992), and these were comparable to the estimates provided by Wechsler (1981). Moore, Stambrook, Hawryluk, Peters, Gill, and Hymans (1990) found generally high subtest stability coefficients for a sample of 60 head-injured patients with a retest interval of 8.48 months (median = .77, range = .92 to .61). Given the substantial improvement in cognitive functioning seen in head trauma patients, retest coefficients may tend to underestimate test reliability for this sample. Piedmont,

Sokolove, and Fleming (1989a) calculated internal consistency reliability estimates for 229 psychiatric patients and found these parallelled those in the WAIS-R manual. The general conclusion from all these studies is that the WAIS-R can be used reliably with psychiatric and neurological clients. IQ scale scores have reliability estimates in the same range as those for the standardisation sample; individual subtest reliabilities are comparable in most cases to those computed for normal healthy individuals.

The magnitude of score errors

All measurements in science are error prone, and scores on the WAIS-R, although reliable, are no exception. The most common statistic for quantifying this error is the standard error of measurement (SE_m), which is normally distributed about the predicted true score of any individual and can be interpreted in a manner similar to the standard deviation (Knight,1983; Silverstein, 1984b). Thus, for a predicted true score of 110, and a SE_m of 5, it can be concluded that there is a 68% chance that the actual true score lies between 105 and 115. A range $\pm 2SE_m$ defines the 95% confidence interval; a range of $2.58SE_m$ defines the 99% confidence interval. Strictly speaking, an obtained score is not a predicted true score, because the WAIS-R is not perfectly reliable, and the predicted true score will tend to regress to the mean. For scores lying in the average range, however, this regression effect will be minimal and so treating the SE_m as the confidence interval around an obtained score is not greatly misleading.

The SE_m is a function of the reliability of a test. Not surprisingly, the least reliable subtests of the WAIS-R have the largest error terms. The SE_m statistics from the WAIS-R manual and elsewhere (e.g., Boone, 1992a; Ryan et al., 1992) show that the 95% confidence interval around a predicted true FSIQ score is about ± 5 points, in both normal and clinical samples, which is an entirely satisfactory finding. It is instructive to observe the magnitude of the error for the subtests. For the majority of subtests, the 95% confidence interval is about 2 to 3 IQ points. Clinicians interested in reporting the precise SE boundaries around obtained subtest scores are referred to Kramer (1990).

Intercorrelation of WAIS-R subtests

In Table 1 a summary of the intercorrelations between subtests, derived from the *WAIS-R Manual*, is presented. As a cross-validation of the manual data, results from Piedmont et al. (1989a), based on data from 229 psychiatric patients receiving in- or outpatient care, are also provided. The pattern of intercorrelations is generally consistent across the two studies and has some interesting features. Overall, the median subtest intercorrelation is between .45 and .50, suggesting that although there is considerable overlap between them, each subtest has some unique features. The magnitudes of these correlations are constrained by their reliability, and it is noticeable that the least reliable test, Object Assembly, has the lowest median correlation with all other subtests. Two subtests, Digit Span and Object Assembly, stand out as generally having low correlations with other subtests, although Object Assembly and Block Design are highly correlated with each other, as are Digit Span and Arithmetic. In the *WAIS-R Manual* corrected subtest - total scale score correlations are presented. These reveal that Digit Span (r = .57) is the verbal subtest with lowest correlation with the total verbal score; all other such correlations exceed .70. Correlations between Performance subtests and the total Performance score are uniformly lower, with Picture Arrangement (.56) and Digit Symbol (.52) having the lowest correlations. Of all the subtests, Digit Span (.58), Object Assembly (.57), and Digit Symbol (.57) have the lowest correlations with FSIQ.

Of note also is that the FSIQ estimates based on subtests that have low correlations with total scale scores (i.e., Digit Span, Picture Arrangement, Object Assembly, and Digit Symbol), are likely to be highly error prone predictions of IQ scale scores. This can be illustrated by predictions calculating the SE of the estimate (SE_{est}) for the WAIS-R subtest - total score correlations. Even subtests with high item-total correlations are still unacceptably error prone. For example, the test with the highest such correlation, Vocabulary, has an SE_{est} of ± 7.9; thus the magnitude of two standard errors is ± 16. The significance of this magnitude of error can be illustrated as follows. Suppose a person ob-

Table 1
Intercorrelations Between Subtests

	Median r All subtests		Range (Manual)[a]
	Manual[a]	Piedmont et al	
Information	.51	.59	.44 (Digit Symbol) .81 (Vocabulary)
Digit Span	.43	.52	.33 (Object Assembly) .56 (Arithmetic)
Vocabulary	.53	.58	.41 (Object Assembly) .81 (Information)
Arithmetic	.56	.61	.42 (Object Assembly) .63 (Vocabulary)
Similarities	.52	.55	.43 (Object Assembly) .72 (Vocabulary)
Picture Completion	.52	.49	.43 (Object Assembly) .55 (Vocabulary)
Picture Arrangement	.50	.49	.37 (Digit Span) .54 (Block Design)
Block Design	.49	.56	.43 (Digit Span) .63 (Object Assembly)
Object Assembly	.40	.43	.33 (Digit Span) .63 (Block Design)
Digit Symbol	.44	.48	.38 (Object Assembly) .47 (Vocabulary/ Block)

[a] Results based on the average correlation across 9 age ranges.

tained a Vocabulary score of 13, one standard deviation above the mean, and a clinician proposed that this person's IQ was also one standard deviation above the mean (i.e., 115) on the grounds that the Vocabulary score was considered to be the best estimate of the person's premorbid functioning, or because this score was the highest in a set of subtest scores. Ignoring for the moment any effect due to regression to the mean, it can be seen that the error involved in making this estimate is substantial (115±16 at the 95% confidence level). For Picture Completion, with a subtest-total score correlation of .67, the SE_{est} is 11.13, and for the worst case, Object Assembly, the SE_{est} is about ±12.43. These results suggest that the clinical strategy of selecting by inspection a high subtest score and concluding that this is somehow indicative of an overall high (premorbid) IQ score is a dangerous practice. Some subtests have low correlations with FSIQ and therefore provide particularly error-prone IQ estimates. Even correlations in excess of .70 have SE_{est} values of about 10 points. The intercorrelation data for the WAIS-R suggest prudence in basing clinical inferences about IQ levels on the results of any one subtest. The analysis of intersubtest scatter, which will be reported in following sections, reinforces this caveat.

Factor Analysis of the WAIS-R

The pattern of intercorrelations between subtests has been explored extensively using factor analysis. This work has had the general objective of attempting to uncover the structure of human intelligence. For example, factor analysis of the earlier Wechsler scales (Cohen, 1957; Matarazzo, 1972) was used to test the notion that intelligence is a single trait (Spearman's *g* factor), or alternatively that the structure of intellect was composed of two or more group factors.

Theories of intelligence that have emerged from factor analysis are based on the premise that a series of latent sources of variance, or factors, can be identified in a set of intercorrelations which explain the variation in a set of subtest scores. Such views of intelligence, which have been labelled *psychometric* or *differential* theories (Sternberg, 1985), differ in the number of factors that are regarded as necessary (from 1 to 150) and also in the geometric arrangement of these factors. Various structural arrangements have been proposed, for example, an unordered structure, where each factor is equally important, like Thurstone's primary mental abilities, or a hierarchical structure, as proposed by Burt (1940). Although these theories often appear radically different, typically they are mathematically equivalent. Different numbers of factors and factorial structure can be derived from the same set of intercorrelations simply by choosing different (but equally valid) methods of analysis. As Sternberg (1985, p. 8) observed: "From this point of view, the different theories say the same thing in different ways". Factor analyses of sets of tests of human abilities have provided a number of interesting alternative views of the structure of intelligence, but no way of choosing between them. This conclusion applies also to the Wechsler scales. The factor analysis of the WAIS-R has produced several alternative views of the structure underlying the performance of the test, which may serve to inform the clinical use of the scale. However, no simple answer has been provided to such fundamental questions as "How many factors best fit the WAIS-R?" or "How are the WAIS-R factors structurally related to each other?" (Leckliter, Matarazzo, & Silverstein, 1986).

If the intercorrelation of subtests is subjected to either principal components analysis (i.e., the communalities are set equal to unity), or principal axis factoring (i.e., the communalities are estimated by squared multiple correlations), then the first unrotated component/factor to emerge may be regarded as an estimate of Spearman's general factor, *g*. Several studies have reported loadings for a general factor (e.g., Atkinson & Cyr, 1984; Blaha & Wallbrown, 1982; Canavan, Dunn, & McMillan, 1986; Crawford, Allan, Stephen, Parker, & Besson, 1989; O'Grady, 1983; Silverstein, 1982a; Sturmey, Gatherer, Ghadiali, Hallett, & Searle, 1993). Overall, the general factor tends to account for between 50% and 60% of the total score variance and over 90% of the common variance (e.g., Blaha & Wallbrown, 1982; Silverstein, 1982a). These results support the interpretation of the FSIQ score as a general indication of the quality of an individual's overall intellectual capacity. The individual factor loadings reflect the pattern of intercorrelations between FSIQ and the subtests discussed in the preceding section. There is evidence that the Verbal subtests Information, Vocabulary, Comprehension, and Similarities most strongly correlate with FSIQ; the lowest loadings are for Digit Symbol and Digit Span. Despite this variability, all subtests load to at least a moderate extent on the general factor.

Alternatively, if principal axis factoring followed by orthogonal rotation to simple structure are employed, and a two-factor solution selected, then two factors, typically labelled Verbal-Comprehension (VC) and Perceptual-Organisation (PO), emerge (e.g., Athanasou, 1993; Atkinson & Cyr, 1984; Beck, Horwitz, Seidenberg, Parker, & Frank, 1985; Burgess, Flint & Adshead, 1992; Crawford et al., 1990; Gutkin, Reynolds, & Galvin, 1984; Parker, 1983; Ryan et al., 1990b; Ryan, Rosenberg, & DeWolfe, 1984; Siegert, Patten, Taylor, & McCormick, 1988). A similar pattern of results has been found for hierarchical analyses (e.g., Blaha & Wallbrown, 1984; Blaha & Mandes, 1993). The Verbal subtests (particularly Information, Vocabulary, Comprehension, and Similarities) load on the VC factor; Block Design, Object Assembly, Picture Completion, and Picture Arrangement load on the PO factor. The structure is not completely simple, however, with several subtests loading at a substantial level on both factors, for example, Arithmetic, Picture Arrangement, Picture Completion, and Digit Symbol.

Earlier factor analyses of the WAIS revealed that in addition to the VC and PO factors, a third factor, Memory/Freedom from Distraction (MF), typically emerged (Matarazzo, 1972). In several factor analytic investigations, the 3-factor solution, resulting from a principal axis factoring with orthogonal rotation to simple structure, has been judged the most satisfactory (Parker, 1983; Ryan & Schneider, 1986). In these analyses, Digit Span and Arithmetic load most highly on the third factor. In general, where 3-factor solutions are found, VC is represented by Vocabulary, In-

formation, Comprehension, and Similarities, PO by Object Assembly, Block Design, and Picture Completion, and MF by Digit Span and Arithmetic. Picture Arrangement and Digit Symbol are more variable in their loadings and show a weaker approximation to simple structure. For example, Digit Symbol has been found to load on the MF factor (e.g., Waller & Waldman, 1990), the PO factor (e.g., Ryan et al., 1990b), or on all three factors (e.g., Parker, 1983). It should be remembered that the 3-factor solution does not emerge from all factor analyses (e.g., Siegert et al., 1988; Sturmey et al., 1993) and that the MF factor is less robust than either PO and VC.

The factor analytic studies of the WAIS-R provide an excellent example of both the usefulness and limitations of this technique. Leckliter et al. (1986) in their comprehensive review of the factor analytic studies of the Wechsler scales concluded that the results provided "corroboration for a global general factor (*g*), a robust Verbal Comprehension factor, a Perceptual Organisational factor, and a weaker, but still seemingly robust Memory/Freedom from Distraction factor" (p. 340). Subsequent research has provided no reason to revise this conclusion.

The question of how many factors best fit WAIS-R subtest correlations can not readily be answered. The most parsimonious model is the single-factor structure; nevertheless, the 2-factor and 3-factor models tend to fit the data with equal success (e.g., O'Grady, 1983). On the whole, the solution preferred by most authors is the one most congruent with their clinical needs and theoretical perspective. For the clinical neuropsychologist, the multiple exercises in factor analysis provide some information about the relationships between subtests that can be of value in interpreting results (e.g., Lawson, Inglis, & Stroud, 1993; Zillmer, Waechtler, Harris, Khan, & Fowler, 1992). For example, the 3-factor solution is significant because it identifies as distinctive two subtests, Digit Span and Arithmetic, that appear to assess primary memory and concentration (Sherman, Strauss, Spellacy, & Hunter, 1995). It is useful to know that Arithmetic draws not only on educational skills, but also on memory and attentional factors. Similarly, the 2-factor outputs suggest that although there is a basis for Wechsler's division of subtests into Verbal and Performance, there is considerable overlap. This in turn implies that the direct evaluation of the magnitude of VIQ and PIQ differences is likely to be over simplistic. There is also reassurance from single factor outputs that the WAIS-R can be summed meaningfully to provide an overall index of general abilities.

A final caution about the limitations of factor analysis is in order. Although factor analysis can be used to determine the underlying dimensions of individual differences, it gives no clues about the process individuals use to solve the problems presented in each subtest. It is worth reiterating that the Wechsler scales were devised without concern for the cognitive skills needed to solve the specific problems posed by individual subtests. Consequently, in most circumstances, failure on a particular subtest or group of subtests is hard to ascribe convincingly to specific impairments.

Deficit measurement using the WAIS-R

An IQ score is no more than an estimate of a person's functioning on the occasion of testing. Usually, however, a clinical neuropsychologist's objective is to determine whether there has been any deterioration in a client's performance as a consequence, for example, of a traumatic insult to the brain or a degenerative disease process. This requires knowledge of a client's premorbid abilities. Sometimes the clinician has available some background demographic information that allows a comparison of past performance with current IQ. For example, knowing that a once eminent scholar now has an IQ of 95 allows the inference that a change of functioning of some magnitude has occurred. Regression procedures have been formalised that allow prediction of WAIS-R IQ's on the basis of demographic information (e.g., Barona, Chastain, & Reynolds, 1984; Crawford, Stewart, Cochrane, Foulds, Besson, & Parker,1989). The validity of these equations is confined to the countries and circumstances in which they were developed, and will be of questionable value elsewhere. Sometimes previous IQ scores are available from school, military, or personnel testing and can be directly contrasted with present findings.

Frequently, however, a clinician is required to infer premorbid intelligence from the patient's present level of functioning or current test scores. This process relies on the general assumption that cognitive abilities are highly and positively correlated, and that damage to the brain does not affect all abilities equally. Results from the Bellevue provided initial support for both these assumptions. The subtests were significantly intercorrelated and those measures, such as Vocabulary and Information that drew on knowledge consolidated in the past, were resistant to change. Thus, clinicians came to regard those subtests on which an individual obtained the highest scores as potential indicators of premorbid intelligence.

Before turning to an examination of studies that have considered intersubtest scatter, it is necessary to review some statistical issues relating to methods of evaluating score differences. Payne and Jones (1957) introduced the useful distinction between the reliability and abnormality of test score differences to clinicians using the Wechsler scales. A *reliable* difference between any two test scores is one that cannot be attributed to measurement error. Calculating the reliability of a difference between any two subtests requires a formula given by Kelley (1923) and first used with the WAIS by McNemar (1957). This formula and a similar method for computing the reliability of a difference between any one subtest and the average of all subtests are discussed by Silverstein (1991). Two test scores may be reliably different, but a more significant issue is whether or not the difference is abnormal. The *abnormality* of a test score difference refers to the probability with which such a difference was found in the standardisation sample. Tables of abnormal difference can be constructed from the actual distribution of the test scores from the standardisation sample or from estimates based on the correlation between the two tests and their respective reliabilities (Payne & Jones, 1957; Silverstein, 1982b, 1989).

Examining Subtest Scatter

It is common practice in clinical neuropsychological assessment to consider WAIS-R subtest scatter and to draw inferences about impaired or preserved functions based on the spread of these scores. This is illustrated by Table 2, which contains a list of selected articles offering a variety of methods for assessing intersubtest scatter. Urging caution in the interpretation of intersubtest scatter, Kaplan, Fein, Morris, and Delis (1991) observed that:

> For decades clinicians have used intersubtest scatter to make inferences about premorbid intellectual level and acquired brain dysfunction. For example, an individual's highest WAIS-R subtest score is often interpreted as representative of his or her premorbid intelligence; subtest scaled scores that are lower by 3 or more points are interpreted as indicative of cognitive impairment secondary to brain damage. Recent studies, however, have demonstrated the limitations of such a practice (p. 73).

This recommendation of caution is reinforced by the analyses of the WAIS-R standardisation data presented by Matarazzo, Daniel, Prifitera, and Herman (1988) and Matarazzo and Prifitera (1989). They found that the average difference between an individual's highest and lowest subtest scaled score in the IQ range 90 to 119 was about 7 IQ points. Furthermore, the average highest to lowest difference amongst the verbal subtests and performance subtests was about 5 points. About 70% of the sample had a highest-lowest subtest difference across the full test of 6 or more points, 18% of the sample had differences of 9 or more points. Since the WAIS-R standardization sample was carefully screened for individuals with known brain damage or other disorders that might have impaired cognitive ability, the conclusion that "Even when it is substantial such scatter is by itself *not* an indicator of brain dysfunction, in as much as it is a characteristic of the cognitive functioning of normal subjects" (Matarazzo, 1990, p. 1007) is inescapable.

Tables of the magnitude of reliable subtest discrepancies are available, beginning with Table 13 in the *WAIS-R Manual*, where the magnitude of reliable differences between subtests at the 15% confidence level are detailed. These differences range from 1.60 (Vocabulary and Information) to 3.01 (Object Assembly and Picture Arrangement), depending on the magnitude of the SE_{diff} between any two subtests. This table should be used with care. An alpha level of 15% can not be regarded as providing a stringent protection against Type I errors. Compounding this is the difficulty arising from implicitly testing so many subtest differences. The *a posteriori* examination of a WAIS-R profile allows 55 possible subtest comparisons and it will be intuitively self-evident that the probability of finding a spuriously significant difference is high. The formula for determining how likely it is that the null hypothesis will be rejected incorrectly is given by the following formula, where α is the significance level selected, k is the number of statistical tests made, and pr is the probability of one or more significant results, given the null hypothesis is true:

$$pr = 1 - (1-\alpha)^k$$

Substituting $\alpha = .15$ and $k = 55$ in the equation, it becomes obvious that this probability of spuriously rejecting the null hypothesis will be vanishingly close to 1 long before $k = 55$. Making the size of alpha smaller (.05 or .01) helps, but not a great deal. One conservative option is to apply a simple Bonferroni correction, that is, to divide an alpha value of .05 by 55, giving a probability level of .0009, which is more than 3 standard deviations from the mean. If this more conservative procedure is adopted, then the magnitude required for a statistically reliable difference between Vocabulary and Information, for example, increases from 1.60 to 3.66, and between Vocabulary and Digit Symbol from 2.29 to 5.25. The important lesson from this analysis is that a *post hoc* pattern analysis that involves simultaneous comparison of all subtest differences is the same as making 55 simultaneous statistical tests on the same data set. Using traditional cut off points for statistical tests in this situation, that is, setting alpha at .05 or .01, is almost certain to produce Type I errors.

The option of setting the significance level at the level

Table 2

Selected Articles Containing Tables for the Interpretation of WAIS-R Intersubtest Scatter

Study	Tables
Silverstein (1982b)	Reliability of differences between any one subtest and average Verbal, Performance, and Full Scale subtest scores.
Silverstein (1984a)	Abnormality of differences between any one subtest and average Verbal, Performance, and Full Scale subtest scores.
Kramer (1990)	Confidence intervals for WAIS-R subtest scores.
McLean et al. (1989)	Abnormality of highest minus lowest subtest scores and the number of subtests deviating from the person's mean subtest score by ± 3 points. Based on the scaled scores of the standardization sample.
McLean et al. (1990)	Abnormality of subtest variability index of Plake et al. (1981). Based on the scaled scores of the standardization sample.
Matarazzo & Prifitera (1989)	Table of percentage of cases at or above each magnitude of scatter across all 11 subtests for the standardization sample.
Ryan & Paolo (1992)	Abnormality of subtest scatter as used by McLean et al. (1989) for 130 elderly subjects.
Schinka et al. (1994)	Tables of percentages of cases in the standardization sample by highest subtest scaled score at or below the value for the lowest subtest score.
Crawford & Allan (1996)	Tables of the abnormality of subtest range and subtest deviations from the mean from 200 healthy UK subjects.

of significance determined by a simple Bonferroni correction (e.g., at the .0009 level computed above), is probably too stringent a procedure. In addition, it is unlikely that in practice clinicians really want to consider all 55 comparisons in a totally random fashion. Silverstein (1982b) and Knight and Godfrey (1984) concluded that a better procedure is to compare each individual subtest score with either the average of all subtest scores, or the average of the verbal or performance subtest scores. As Silverstein observed, this procedure was first suggested by Wechsler (1941), and the statistical rationale for making these comparisons provided by Davis (1959). The procedure requires the computation of the standard error of measurement of the difference between an average score and one of the scores making up this average. The nominal .05 or .01 level of significance is then corrected by dividing by *k*, the number of comparisons involved. This process allows the clinician to compute the probability that there is a reliable difference between a subtest and the average subtest scores. For the convenience of clinicians, tables of significance of WAIS-R subtests have been provided elsewhere (Knight & Godfrey, 1984; Silverstein, 1982b). In general, a reliable difference between any one Verbal test and the average Verbal subtest score is about 2 to 3.5 points at the .01 level. Reliable differences for Performance subtests range from 3.0 to 4.2 at the .01 level. When comparing one subtest with the average of all 11 others, the differences needed to reach significance at the .01 levels range from about 2.0 to 5.0. The magnitude depends on the reliability of individual subtest; thus for Vocabulary the difference required for significance is 2.2, whereas for Object Assembly the difference is 4.8.

Before leaving the topic of reliability of subtest differences, an issue raised by Krauskopf (1991) should be noted. Krauskopf observed that all the consideration given to interpreting intersubtest scatter focuses on avoiding Type I errors, whereas for clinical purposes it may be more important to avoid Type II errors (i.e., failing to reject a false null hypothesis, or a missing difference that is actually valid). Methods that involve applying Bonferroni corrections to alpha levels, substantially reduce the risk of Type I errors, but at the price of increasing the probability of Type II errors. Silverstein (1982b) suggested that when making *k* comparisons, alpha be set at .05/*k*; in contrast, Krauskopf suggested setting alpha at .15 for each comparison, as recommended by Wechsler (1981). Silverstein's recommendation results in a power level of about .28 (i.e., 72% of true differences of greater than 2 standard errors will be missed); Krauskopf's suggestion sets the family-wise Type I error rate at 80%. Silverstein (1993) has this to say about the dilemma this creates:

> What then is the poor clinician to do? I confess that I am not sure. I no longer stand by the solution I proposed in my 1982 article, but I also find Krauskopf's recommendation (it is not actually a solution) unpalatable. One possibility that suggests itself is to compromise: Make each comparison not at the .05/*k* level or the .15 level but somewhere in between, say at the .05 level. The clinician will continue to miss about half of the true differences of two standard errors and still make one or more errors in interpreting the protocols of about two out of five subjects. It may well be that this is the worst of both worlds, Krauskopf's and my own. The fact of the matter (alas) is that there is not a completely satisfactory solution (p.73).

For the clinical neuropsychologist, the optimal testing strategy may depend on the goal of assessment. Where the aim is to develop hypotheses, then a low Type I error rate is less desirable because the generation of possible hypotheses will be inhibited. On the other hand, if it is important to avoid spurious differences then a low Type I error rate is preferable. Such a conservative approach is likely to be more defensible in a legal context. This issue touches upon the importance of the cost of error, and the way in which a particular testing exercise fits into an overall assessment strategy. Clinicians may adopt a conservative or liberal approach with respect to Type I errors depending on the nature of the assessment and the cost of making mistakes. For example, if a test score indicates some probability of a brain lesion, then it may be best to take further action in a sequential assessment process. The cost of ignoring the deficit may be high. It may not be wise, however, to place much weight on a test result of low reliability, where this constitutes the only evidence of impairment. This is a matter of clinical judgement.

Silverstein (1993) goes on to suggest that perhaps the debate over the reliability of subtest score differences can be set aside and attention focused on the abnormality of such differences. He acknowledges that this "may be the coward's way out, but the multiple - comparisons problem that arises in the case of [reliability] is irrelevant in the case of [abnormality]" (p. 73). Tables of abnormal differences have been computed and are available to the clinician. For example, Silverstein (1984a) derived a formula that allows computation of the magnitude of difference between any one subtest and the average of a set of subtests, and applied this to the WAIS-R standardisation data. For this article, he prepared a useful table of the differences between each subtest score and the average subtest score obtained by selected percentages of the standardization sample. The size of the difference between any one subtest and the average of all 11 subtests obtained by 1% of the standardisation sample ranges from about 6 (for Object Assembly, Digit Symbol, and Digit Span) to 4 for Vocabulary. Similar information is available for the Verbal average and Performance average. Matarazzo and Herman (1985) have provided a similar table of the frequencies of highest-lowest subtest differences, based on the distribution of results from the 1880 individuals who made up the standardisation sample, which was described at the beginning of this section. An equivalent analysis has been presented by Ryan and Paolo (1992a) with the results from their old age sample and by Crawford and Allen (1996) for a healthy UK sample.

The analysis of intersubtest scatter has become increasingly detailed. McLean, Kaufman, and Reynolds (1989) computed two indices of subtest scatter, the highest-lowest subtest difference, and the number of subtests that deviated by 3 points from the person's average scaled subtest score on the Verbal, Performance, and Full Scale subtests. They then produced a cumulative distribution of the scaled score highest-lowest differences (similar to that provided by

Matarazzo et al., 1988) for the WAIS-R VIQ, PIQ, and FSIQ scores. In addition, they produced cumulative distributions of the number of subtest scores deviating by ±3 points from the average of all the subtests. In all, 11% of the sample had 4 or more subtests substantially deviating from the average.

McLean et al. (1989) also determined the size of the highest-lowest subtest score difference needed to exceed the magnitude of the differences achieved by 1, 2, 5, and 10% of the sample at different levels of FSIQ. This recognises the fact that the magnitude of a possible difference in subtest scatter is constrained by the general ability of the individual subject. This is a specific instance of a more general issue that has been drawn to the attention of neuropsychologists by Chapman and Chapman (1988). All other things being equal, the magnitude of possible differences between any two components of a total score is determined in part by the size of that total score. Another way to consider this is to note that ceiling and floor effects will operate to reduce subtest scatter. With the WAIS-R there is a noticeable floor effect, which means that severely impaired patients do not show as much evidence of scatter as they would if they scored overall in the median range. This is illustrated by McLean et al.'s results. The 1% frequency of occurrence of a highest-lowest subtest difference for the total 11 subtests was 10 for persons with an FSIQ 79 or less, compared to 13 for persons in the IQ range greater than 120. Similarly, persons with low IQ's had fewer subtest scores deviating from the average than persons with higher IQ's. Generally, overall IQ constrains the scatter of lower IQ persons but has less effect in the range above 80 FSIQ. McLean, Kaufman, and Reynolds (1990) prepared a further table of abnormal scatter based on an index of variability in subtest scores proposed by Plake, Reynolds, and Gutkin (1981).

Probably the best way of determining the abnormality of subtest scatter was advanced by Schinka, Vanderploeg, and Curtiss (1994). Their method allows comparison of an individual's test results with the relative scatter range of the standardisation sample. They defined relative scatter range as "the difference between the highest and lowest subtest score based on the value of the highest subtest score" (p, 364) and provided cumulative percentage tables, which can be used to find the percentage of people with each possible lowest subtest score for a particular highest subtest score. Thus, for a particular profile where the highest score is 14, the percentage of people who had a lowest score of 5 can be determined (in this case 20% of individuals in the standardisation sample with a highest score of 14 had a lowest score of 5). Most clinicians would be struck by a profile where the highest score was 17 and the lowest was 6, yet surprisingly, 22.5% of the standardisation sample with a score of 17 or more, also had a subtest score of 6.

It is possible that indices of abnormal subtest scatter might have some value as diagnostic markers of brain damage. For example, the tables provided by McLean et al. (1989) allow the determination of aberrant subtest scatter in terms of the number of an individual's subtests differing significantly from their average subtest score, an index of profile variability. However, Ryan, Paolo, and Smith (1992) found that abnormal intersubtest scatter measured in this way was not characteristic of brain-damaged clients. The amount of scatter observed in their sample of 216 male patients with a range of neurological conditions was equivalent to that seen in the standardisation sample. Other methods of computing the amount of scatter in a WAIS-R profile have been advanced, including the use of the Mahalanobis distance statistic (e.g., Burgess, 1991; Crawford & Allen,1994) but have not been established as valid markers of brain impairment. Using scatter as a means of screening for brain damage has not been shown to be useful in validation studies of brain damaged persons.

Conclusions

All available evidence converges on the conclusion that clinicians should offer interpretations made on the basis of the examination of intertest scatter with extreme care. The advice of Schinka et al. (1994) should be heeded: *Most clinicians underestimate the base rate for scatter and tables should be consulted to guard against over interpretation.*

Reliable differences between subtests are not necessarily indicative of any pathology. Evidence from statistical tables of abnormal differences and results from the standardisation sample reveal that large discrepancies occur with surprising frequency in people with no known or suspected pathology. It is therefore important that subtest scatter be considered only in the context of all available information about a client. Only where lines of evidence converge can clinicians be confident in their interpretation of significant subtest differences. For example, a client with an average subtest score of 12, and a Digit Span scaled score of 7, has a discrepancy on that subtest that is reliable (Silverstein, 1982b) and found in only about 2% of normal population (Silverstein, 1984a). This result becomes somewhat more salient, however, when it is known that the client is nonaphasic, but has undergone a neurosurgical resection involving the left parietal area. Primary memory is often affected by lesions in this area (Kolb & Milner, 1981), and the convergence of evidence strengthens the interpretation of the WAIS-R data as suggestive of acquired cognitive impairment.

Verbal-Performance IQ differences

There is a considerable literature on the significance of VIQ-PIQ differences (e.g., Crawford, 1992; Lezak, 1994), and opinions differ on the usefulness of this index in clinical practice. Proponents (e.g., Bornstein 1983; Bornstein & Matarazzo, 1982) draw attention to the fact that grouping subtests together enhances the reliability of scores, thereby making estimates of differences less prone to error. An inspection of the tests making up the verbal scale reveals that they are generally resistant to the effects of diffuse brain damage and aging, and rely more heavily on verbal-educational skills. The Performance subtests, which take more account of speed of problem solution and use more unfamiliar tasks, are generally more susceptible to the effects of aging and diffuse damage. Hence, clinicians have regarded a discrepancy in favour of VIQ as being indicative of right hemisphere damage or diffuse dementia, and one in favour

of PIQ as indicative of left hemisphere damage.

The problem with these interpretations is that the subtests grouped under the Verbal or Performance heading are not homogeneous with respect to the cognitive skills required or the putative localisation of function. This view has been asserted by Lezak (1994), who stated in no uncertain terms that PIQ and VIQ scores "are based on averages of quite dissimilar functions that have relatively low intercorrelations and bear no regular neuroanatomical relationship to one another" (p. 690). She observed that the subtests were assigned to one scale or the other primarily on the basis of Wechsler's intuitions and that there is considerable factor analytic and neuropsychological evidence to support the view that Wechsler's two-fold clustering of subtests is less than ideal.

The statistical interpretation of VIQ-PIQ discrepancies has been facilitated by the construction of tables similar to those designed to examine subtest scatter. In the WAIS-R Manual, Table 13 provides details of the discrepancies necessary for a *reliable* difference at the 0.15 and 0.05 levels. Naglieri (1982) subsequently calculated differences required at the 0.01 level. Significant differences determined by these tables establish that an observed score difference can not be attributed to measurement error. A 10-point discrepancy is reliable at the .05 level. A table for assessing the *abnormality* of a VIQ-PIQ difference was provided by Knight (1983). A discrepancy in the order of 23 points is found in 5% of the population, and about 27-29 points in 1% of the population. If the 1% and 5% levels are used as cutoff points, it can be seen that a substantial discrepancy is required before a VIQ-PIQ difference can be regarded as abnormal. This conclusion is supported by the results from the WAIS-R standardisation sample (Bornstein, Suga, & Prifitera, 1987; Matarazzo & Herman, 1984b, 1985; Matarazzo, Bornstein, McDermott, & Noonan, 1986; Silverstein, 1985a). Matarazzo and Herman (1985) constructed a table of the frequencies of the actual magnitudes of the differences in VIQ and PIQ for the 1880 subjects of the standardisation sample and found that the discrepancies were approximately normally distributed, ranging from -43 to +49, with a mean approaching 0.0, a standard deviation of 11, and a SE of .26. About 1% of the sample had VIQ-PIQ discrepancies greater than 29. The probability values of .15, .10, .05, and .01 occurred at discrepancies of 16, 19, 23, and 30 points, respectively. A difference of 16 points, which is *reliably* different, actually occurred in 15% of the healthy and unimpaired standardisation group. Paolo and Ryan (1992) and Ryan and Paolo (1992b) examined VIQ-PIQ differences in their sample of 130 normal volunteers over the age of 74, and found 32% had discrepancies greater than 12, and 7.7% had discrepancies greater than 19 points. Results from normative samples caution against the overinterpretation of seemingly impressive IQ scale discrepancies.

It should also be noted that these tables take no account of a person's overall intellectual level and the constraining effects this has on the potential magnitude of VIQ-PIQ differences. Floor effects in particular will drastically attenuate the possible size of any PIQ-VIQ difference, making this a difficult index to use as a general indicator of brain damage. A demented person with Alzheimer's disease and an FSIQ of 80, is unlikely to have a pathognomic VIQ-PIQ discrepancy, despite having obvious brain-related impairments. This constraint places a considerable limitation on the use of a VIQ-PIQ difference as an indicator of brain dysfunction.

There is some empirical evidence of the effects of brain injury on VIQ and PIQ scores. A review of 28 studies of clinical samples to which the Wechsler-Bellevue or WAIS had been administered, and VIQ and PIQ scale scores reported, was presented by Matarazzo and Bornstein (1982). They concluded that there was evidence that VIQ deficits were more commonly found in patients with left-hemisphere (LH) lesions, and PIQ deficits in patients with right-hemisphere (RH) lesions, but that these effects were stronger for male than female samples. Bornstein (1983) examined the WAIS-R results of 89 patients referred for neuropsychological assessment, 20 with LH damage, 24 with RH damage, and 45 with bilateral involvement. As predicted, LH damage resulted in PIQ being 4.9 points higher than VIQ, whereas RH damage resulted in VIQ being 10.33 points higher than PIQ. Bilateral damage resulted in VIQ being 7.73 points higher than PIQ. All differences were statistically significant. However, the magnitude of these average discrepancies was not large, reaching the 10-point difference needed to be reliable at the .05 level in only one instance. Bornstein concluded that the "present data reemphasize the point that VIQ-PIQ discrepancies (of whatever magnitude) *in isolation* are not effective indexes of cerebral dysfunction. Conversely, the lack of such differences does not necessarily argue against the presence of lateralized dysfunction" (p. 780).

There is some evidence that dementia results in a reduced PIQ score relative to VIQ. Randolph et al. (1993) reported average VIQ-PIQ discrepancies of around 15 for a group of Parkinsons's disease patients (FSIQ=109), 8 points for Alzheimer patients (FSIQ=82), and 7 points for Huntington's disease patients (FSIQ=87). The average age-scaled scores for the total sample on each of the Verbal subtests, with the exception of Arithmetic, were all higher than the scores on the five Performance tests.

Another method of combining subtests is to use factor loadings as the basis for aggregating subtests (Canavan & Beckman, 1993; Lawson & Inglis, 1983; Lawson, Inglis, & Stroud, 1983; Parker & Atkinson, 1995). Lawson and colleagues have related WAIS-R factor scores to lateralised deficits with some success. Atkinson (1991a) prepared tables for assessing the reliability and abnormality of score differences based on factor score differences, which are comparable to the more traditional VIQ-PIQ discrepancy tables. Vocabulary, Information, and Comprehension were combined to make the VC Factor score, Block Design and Object Assembly for PO, and Digit Span and Arithmetic for MF. These subtests were chosen because of their high and unique loadings on the factors they represent. Atkinson found the average internal reliability statistics for these three factors were .96, .86, and .89 respectively, suggesting also that the 7 subtest short form is likely to be highly reliable and correlate significantly with FSIQ.

Conclusions

Neuropsychologists have come to recognise that the summary IQ scale scores may disguise individual patterns of performance of significance in understanding a client's cognitive strengths and weaknesses. Unfortunately, since the individual subtests were not necessarily designed to measure specific cognitive processes, their construct validity is not particularly compelling and their individual reliabilities are in some cases only moderate. Thus, combining subtest scores to reduce error and thereby enhance the plausibility of interpretation, is on the face of it, an attractive strategy . The best validated index in clinical use employing a subtest combination strategy is the VIQ-PIQ discrepancy (Silverstein, 1985). There is some empirical evidence that large discrepancies are, on average, more common in impaired persons than in controls, nevertheless they also occur at a high rate in the standardization sample, which undermines the diagnostic utility of this index. In addition, in many disorders such as dementia, there is a global decline in functioning that depresses the magnitude of this, and most other such indices. For neuropsychologists who are concerned with Lezak's justifiable reservations about combining heterogeneous subtests to produce IQ scores, aggregating subtests on the basis of their factor loadings (e.g., Atkinson, 1991a) may appeal. This method at least circumvents criticisms of the lack of correlational association between the subtests combined to produce scale scores.

Profile Analysis

Traditionally there has been considerable interest in describing profiles of Wechsler subtest scores that are characteristic of various pathological groups, such as neurologically impaired clients, criminals, and psychiatric patients (e.g., Abrams, Redfield, & Taylor, 1981; DeWolfe & Ryan, 1984; Fuld, 1984; McDermott, Jones, Glutting, & Noonan, 1989; Piedmont, Sokolove, & Fleming, 1989b). The Fuld cholinergic profile, considered later in this section, is a recent example of this endeavour. McDermott et al. (1989) used cluster analysis to examine the data from the standardisation sample in an attempt to identify core WAIS-R profiles. They identified nine core profiles, defined primarily by overall ability level, incidentally providing further support for the notion of the test as a measure of *g*. Variations in the profile structure that deviated from this general finding involved VIQ/PIQ discrepancies in two cases and one other case where Digit Symbol was elevated relative to a range of low scores on the remaining subtests. The authors suggested that these results could be used to determine the uniqueness of an individual client's profile for use as a possible index of pathology. They proposed that this might be achieved in the clinical setting by comparing the individual profile with the most similar of the core profiles (as detailed in McDermott et al.'s article) and summing the squared differences between each scale score of the individual being assessed and the core profile mean, for all the subtests. If this sum exceeded 80, then the individual's profile could be interpreted as abnormal. Ryan and Bohac (1994) tested this procedure with a sample of 161 WAIS-R profiles from a series of brain damaged patients. They found that only 29 of their patients had abnormal profiles and concluded that this kind of core profile comparison was of little value in neuropsychological diagnosis.

Formal evaluation of another set of proposed diagnostic rules based on the WAIS-R has been equally unrewarding. Piedmont et al. (1989b) assessed the utility of Rapaport, Gill, and Schafer's (1968) diagnostic hypotheses for using subtest differences to determine psychiatric diagnoses. An example of such a hypothesis was that Information > Comprehension and Arithmetic > Digit Span were more common in psychosis than in other disorders. In a study of 422 psychiatric patients, Piedmont et al. found that these subtest combination strategies did not contribute significantly to diagnostic accuracy.

Cholinergic profile

A procedure for examining Wechsler subtest patterns in patients with dementia of the Alzheimer type (DAT) was proposed by Fuld (1984). She administered the anticholinergic drug methoscopalamine, known to induce memory and cognitive deficits in humans, to 20 students. Cholinergic depletion is known to be characteristic of DAT and she was interested in comparing the pattern of WAIS subtest scores in cholinergic treated subjects, DAT patients, and relevant controls. The elements of the cholinergic profile were defined as: A = [Information + Vocabulary] / 2, B = [Similarities + Digit Span] / 2; C = [Digit Symbol + Block Design] / 2; D = Object Assembly. The A, B, C and D terms were computed from age-corrected subtest scores and the profile defined as: $A > B > C \leq D$; $A > D$. Fuld found that 53% of her scopolamine group showed this profile compared to 18% of her controls. Further, 44% of the DAT patients versus 4% of the non DAT demented patients, showed the cholinergic profile.

A number of subsequent studies investigated the incidence of the Fuld profile in demented patients with or without DAT (e.g., Alexander, Prohovnik, Stern, & Mayeux, 1994; Brinkman & Braun, 1984; Gfeller & Rankin, 1992; Goldman, Axelrod, Giordani, Foster, & Berent, 1992; Logsdon, Teri, Williams, Vitiello, & Prinz, 1989). The results have been variable and in some cases have called into question the diagnostic sensitivity of this profile (e.g., Filley, Kobayashi, & Heaton, 1987; Logsdon et al., 1989). For example, Goldman et al. (1992) found the sensitivity of the Fuld cholinergic profile to be "unacceptably low", despite retesting DAT patients on three occasions a year apart. It appears that although such a profile may be seen more commonly in DAT patients than in controls, the incidence of true positives is too low to make it a useful diagnostic indicator of DAT. Gfeller and Ranklin (1991) concluded that: "...we believe factors including the heterogeneity of neuroanatomical structures affected by DAT, as well as the heterogeneity of premorbid cognitive abilities among individuals affected by the condition, present formidable obstacles in the search for a "cognitive marker" of DAT " (p. 634). They go on to state: "We believe that the utility of cognitive measures, such as the WAIS-R, lies in their ability to describe accurately the severity and extent of a patient's deficits, as well as the impact of such deficits on daily

activities" (p. 634).

Intrasubtest Scatter

Within each WAIS-R subtest, items are arranged in order of normative difficulty. Clinicians often observe variable performance on items within a subtest (e.g., success after multiple failures), and may view this as potentially important qualitative information, for instance, as a possible marker for brain impairment (e.g., Mittenberg, Hammeke, & Rao, 1989; Kaplan et al., 1991). Intrasubtest scatter has been calculated in several ways. Kaplan et al. (1991) proposed a scatter score that is the sum of the differences between each consecutive score, that is, a score of 1 followed by either a 0 or 2, is scored as 1. Thus a set of scores on the Information subtest items of 111, 0, 11, 00, 1, 000, 11, 00000 would give a scatter score of 7. In the manual for the *WAIS-R as a Neuropsychological Instrument*, Kaplan et al. have provided tables for assessing the abnormality of intrasubtest scatter for eight of the WAIS-R subtests (excluding Digit Span, Digit Symbol, and Picture Arrangement) based on data from the standardisation sample. These tables relate scatter scores to the total scaled scores on each subtest, thereby taking account of ceiling and floor effects.

The validity of intrasubtest scatter as an index of cognitive impairment has been considered in a handful of studies. Mittenberg et al. (1989) compared the scatter scores of 43 closed head injury patients with a group of normal controls using several different indices of scatter. They found some group differences but concluded that intrasubtest scatter was not sufficiently sensitive for routine clinical use. Mittenberg, Thompson, Schwartz, Ryan, and Levitt (1991) examined subtest scatter in 32 patients with Alzheimer's disease and 32 normal elderly controls. They found no striking evidence of increased scatter in their clinically demented patients and concluded that qualitative scatter analysis was unlikely to be useful in diagnosing dementia. In an uncontrolled study of 32 schizophrenic patients, scatter scores were found to be correlated with measures of attention and scores of the MF factor of the WAIS-R (Feinberg & McIlvried, 1991).

To date, although intrasubtest scatter is often conducted as part of a qualitative interpretation of a WAIS-R profile, there is little evidence to support the use of scatter scores as diagnostic aids. There are numerous ways of computing scatter, but the method used by Kaplan et al. (1991), because it is supported by well-constructed normative tables, is to be preferred.

National Adult Reading Test and the WAIS-R

The ability to recognise words is a cognitive skill that is particularly resistant to the effects of diffuse or progressive brain injury. For this reason, reading tests have often been employed to determine the premorbid level of functioning of brain damaged patients in order to establish the degree of deterioration that has occurred, particularly where progressive dementia is suspected. The rationale for comparing current functioning with reading test performance (which is assumed to be insensitive to brain damage) is the same as that supporting the use of demographic equations (e.g., Barona, Reynold, & Chastain, 1984; Sweet, Moberg, & Tovian, 1990) and the examination of intersubtest scatter.

The best known test of word-recognition for use in the estimation of premorbid IQ is the National Adult Reading Test (NART). Originally labelled the New Adult Reading Test (Nelson & O'Connell, 1978), the NART (Nelson, 1982) replaced the Schonell Graded Word Reading Test (previously used for making IQ estimates), which was found to be subject to ceiling effects. The NART requires the person tested to read aloud 50 words of irregular pronunciation (e.g., thyme, puerperal, etc.). The number of errors is used in a regression equation to predict IQ scale scores. This test was originally developed at the National Hospital in London and is used extensively in the UK (Crawford, 1992). Recently it has also has begun to find favour in the US (Blair & Spreen, 1989).

In support of the premise that word recognition is unaffected by brain damage, Nelson and O'Connell (1978) compared NART error scores in 120 non brain injured patients with those of 40 persons with bilateral cortical atrophy. They found a significant difference between these groups on the WAIS but not on the NART. This finding that word recognition is not substantially affected by brain lesions has been confirmed by others (Nebes, Martin, & Horn, 1984; O'Carroll & Gilleard, 1986; Sharpe & O'Carroll, 1991). Some cautions, however, are in order following the findings of Stebbins, Wilson, Gilley, Bernard, and Fox (1990). They investigated NART performance in groups of demented patients categorised in terms of dementia severity. They concluded that the accuracy of IQ prediction was least satisfactory when the level of dementia was severe and language was compromised. Under those circumstances, they concluded that NART IQs underestimate premorbid IQ.

The regression equations for predicting WAIS IQs from NART errors provided by Nelson (1982) have been refined and cross-validated by Crawford, Parker, Stewart, Besson, and DeLacey (1989). However, they became obsolete with the publication of the WAIS-R. Consequently there have been a number of studies undertaken to revise the procedures used and relate NART scores to WAIS-R IQ. In an Australian study, Willshire, Kinsella, and Prior (1991) calculated WAIS-R FSIQ prediction equations from NART error scores of 104 persons aged 20-69 and found that the NART predicted 46% of the variance in IQ scores. Sharpe and O'Carroll (1991) developed similar regression equations for a Canadian sample and found the NART errors accounted for 59% of the WAIS-R variance.

Ryan and Paolo (1992c) estimated WAIS-R IQs from NART scores in a sample of normal elderly. They found that their equations could be used to determine the presence of cognitive deterioration in a group of 20 demented patients. Ryan and Paolo (1992b) found correlations between NART errors and WAIS-R VIQ, PIQ, and FSIQ scores of -.78, -.56, and -.74 respectively. The corresponding SE_{est} values for the estimated WAIS-R IQ scores were 7.96, 11.56, and 8.81. These errors estimates are typical of NART IQ estimates and illustrate the degree of imprecision involved in making IQ estimates based on NART scores.

The original UK version of the NART has subsequently been revised by Blair and Spreen (1989). They prepared an amended word list and manual for use in North America (the NART-R). The validity of this version was assessed by Weins, Bryan, and Crossen (1993) using a sample of 302 healthy job applicants. They found a correlation of .46 between the WAIS-R FSIQ and the number of NART errors, and showed that NART IQ scores overestimated obtained WAIS-R IQs in low IQ groups and underestimated IQ scores amongst high IQ groups. They also found that 95% of their participants had less than a 15-point discrepancy between estimated and obtained IQ, and suggested that 15 be used as a cutoff for determining the presence of cognitive deterioration.

The NART has been increasingly used to establish baseline premorbid IQ, particularly in cases of mild or suspected dementia, in both research and clinical practice. It should be noted that the validity of the NART as a predictor of WAIS-R IQ is most firmly established in the United Kingdom and United States. Furthermore, scores on the NART are likely to be influenced by both educational opportunities and any damage to areas of the brain involved in language production. It is also important to note that the estimates are error prone and that discrepancies of 15 to 16 points for the FSIQ scores are to be expected as a result of measurement error alone. Another consideration is the magnitude of the constant in the regression equation, typically between 125 and 132. This value determines the upper limit that an estimated IQ score can assume and suggests that ceiling effects may occur for persons of superior intelligence.

Rest-Retest Changes on the WAIS-R

Neuropsychological tests are often administered as part of the baseline assessment of a patient's current status with a view to retesting on a later occasion and documenting changes. Thus a person with epilepsy might be tested before and after neurosurgery or a client with suspected dementia reassessed to see if any deterioration has occurred. Valid interpretation of any change in WAIS-R IQ over time presents difficulties because of the need to account for both practice effects and test unreliability.

A statistical solution to this problem was advanced by Lord and Novik (1968), who derived a form of the SE_m suitable for estimating the confidence interval for a score on a retest administration, given an obtained score on the first administration. The interpretation of this SE_m term was considered by Dudek (1979) and applied to the WAIS-R by Knight (1983). Lord and Novik (1968) called this term the *standard error of prediction* and it is defined as the confidence interval around the predicted true score within which two-thirds of those individuals with that predicted true score would lie. The predicted true score is not the obtained retest score but is a value regressed toward the mean, and can be calculated for any test, using the formula $T = M + r_{11} (X-M)$, where X is the obtained score, T is the estimated true score, M is the test mean, and r_{11} is the reliability of the test (Brophy, 1986). Values for the standard error of prediction of the WAIS-R are presented in Knight (1983). Because the WAIS-R FSIQ is highly reliable, the obtained and true scores will be nearly equivalent, thus the clinician can test the reliability of test-retest change using the obtained test score on the first administration and the appropriate SE value in Knight (1983) to obtain a preliminary indication of the significance of a change in FSIQ on a retest administration. It should be noted, however, that this procedure does not take account of the practice effects resulting from a repeat administration.

Another useful expedient that the clinician may adopt is to compare an individual's change in test scores with normative tables of changes. Matarazzo and Herman (1984b) provided such a table based on the results of the 119 individuals retested as part of the WAIS-R standardization programme. Ryan et al. (1992) have provided similar data for 61 healthy volunteers aged over 75. Examination of their data reveals that only about 7% of the retest standardisation group had a lower score on retest, and only 2 of 119 persons tested had a negative change score of 5 or more. Similarly, in the elderly sample a decline of -5 or more was recorded by only one of the 61 persons retested. These results establish that a deterioration of more than 5 points in FSIQ is abnormal. An increase in score of 15 FSIQ points was found in less than 5% of the standardization sample and a change score of 12 in about 10% of the sample. These difference scores provide preliminary cutoff points for consideration of an abnormally high increase in test scores. Note that both these values are substantially higher than the 8 points determined by the criterion of ± 2 SE that resulted from consideration of test score reliability.

The WAIS-R and the Wechsler Memory Scale-Revised (WAIS-R)

Scores on intelligence and memory tests have typically been found to correlate highly, thus a discrepancy between IQ and memory performance has often been used as an indicator of amnesic dysfunction or brain impairment (Lezak, 1994). The most widely used index of this kind employed in recent times has been the difference between the Wechsler Memory Scale (WMS) Memory Quotient (MQ) and the FSIQ from a Wechsler individual intelligence scale (Knight & Longmore, 1991) with discrepancies of 10 or 12 points being regarded as suggestive of memory impairment (Prigatano, 1978). The revision and restandardisation of both the WMS and the WAIS, however, have rendered the IQ-MQ index obsolete.

The WMS-R represents a considerable advance over the original version of the scale in that it allows the evaluation of a wide range of memory functions and the computation of several summary indices of memory performance: General Memory, Verbal Memory, Visual Memory, Attention/Concentration, and Delayed Recall. In a manner analogous to that used to evaluate VIQ and PIQ discrepancies, the statistical significance of IQ and Memory Index differences can be evaluated. The necessary calculations have been completed by Atkinson (1991b), who presented a table of the magnitude of reliable differences between WMS-R indexes and WAIS-R FSIQ, allowing the clinician to address such questions as: "Is a discrepancy between a

Delayed Recall score of 70 and an FSIQ of 95 attributable to measurement error?" The answer is no, because for Delayed Recall and FSIQ the difference required for significance is 15, and a difference of 25 is well beyond this cutoff point. Atkinson urged caution in the uncritical application of this (or any other) table of reliable differences. Estimates of significant difference scores depend on the comparability of the normative samples of the two measures (which is not complete with the WMS-R and WAIS-R), and the assumption, where the test is used outside the US, that estimates of reliability and correlation are applicable. In addition, such a table of reliable differences takes no account of possible regression effects. Regression occurs in any situation where the correlation between two measures is less than one and results in a trend towards the second score of any testee whose first score is deviant, being closer to the mean than the first score.

These problems can be countered by developing tables of abnormal test score differences. Bornstein, Chelune, and Prifitera (1989) examined the IQ of clinical patients referred for neuropsychological examination and 110 matched cases drawn from the WAIS-R standardisation sample. The normal sample had a mean FSIQ score of 103.5, and the clinical sample mean was 87.3. They found that discrepancies based on scales assessing immediate memory (Verbal and Visual Memory) did not distinguish between the two samples. The most useful discrepancy index was based on the difference between Delayed Memory and FSIQ. They found that a discrepancy of 15 points was found in only 10% of the control sample but in about 30% of the clinical sample. There are some limitations with this report, which the authors acknowledge. The control group comprised participants from the two age ranges (35-44 and 65-69) who happened to be administered both tests. The clinical sample comprised neurologically damaged patients who may or may not have had memory impairments. Furthermore, the average IQ of the clinical sample was 16 points lower than the controls, which raises the possibility of floor effects limiting the possible size of a discrepancy. It should also be remembered that an IQ-Memory deficit in test scores is likely to be characteristic primarily of persons with circumscribed memory deficits in the presence of normal intelligence. Such cases are unusual, and diffuse damage affecting a range of cognitive skills is more common. The authors conclude with the comment that "It is abundantly clear that simple IQ-Memory Index discrepancies cannot and should not be used in isolation to identify memory deficits" (p. 206).

Conclusions

The determination of deficit is a central issue for neuropsychologists using the WAIS-R. All of the research cited above adds up to a general expression of caution about the overinterpretation of WAIS-R results. This is particularly so in cases where the test is not being used in an exploratory manner but is being used directly in a decision making process, for example, to determine a claim for injury compensation.

Numerous methods of inferring deficit from direct or indirect evidence of premorbid function have been advanced. Direct evidence from school, occupational, or military testing records can be useful, although seldom available. Sometimes it is possible to estimate intellectual level from a person's history of occupational or academic achievements, but unfortunately, it is common for a client's background to provide only ambiguous clues about intellectual endowment. Often any such information is compromised by evidence of subsequent evidence of brain damage, such as concussion, alcoholism, or solvent exposure. It is as well to be cautious about asserting that particular achievements or lack of them are evidence of premorbid abilities. Ziskin and Faust (1988) review a case where an expert witness attempted to argue that a client, who had been a union shop steward, would have been of above average intelligence premorbidly, because he had obtained such an important position. Under cross-examination, the witness was forced to admit knowing nothing much about the work of a shop steward, having only a general knowledge of the industry in which the man had worked, and that there was no empirical evidence available about the usual intelligence level of shop stewards. The lesson from this interaction is obvious: Assertions about premorbid functioning that may at first blush appear plausible, can undermine the expert's credibility when subjected to detailed scrutiny. Regression formulas based on demographic characteristics provide a more systematic approach to estimating IQ, and have the advantage of being independent of any test results, but they are of little value outside the country in which they were developed. Thus UK or US formulas are of no value in New Zealand or Australia.

The examination of WAIS-R subtest scatter has become increasingly sophisticated and the test user has recourse to several tables describing the magnitude of subtest differences indicative of statistical significance or abnormality. This work relies on the relatively high degree of intercorrelation between the 11 subtests, and the consequent conclusion that the significant deviation of one or more subtests scores from the average subtest score is indicative of a specific cognitive dysfunction. The fact that few of the subtests actually measure specific cognitive factors that could deviate from the average, works against this. There is something of a paradox in expecting highly intercorrelated tests to reveal specific deficits. Of course, high correlations between cognitive tests do not mean that the same processes are necessary for their successful completion, but nonetheless any discussion of intersubtest scatter must acknowledge that the individual subtests were not chosen as measures of discrete neuropsychological functions. Wechsler built a test to measure *g* and the evidence from factor analytic and related investigations, shows that in this, he was largely successful. It should come as no surprise, therefore, that measures of scatter do not distinguish brain injured groups from normal groups. Marked intersubtest scatter is not a characteristic of brain injured persons. However, this does not preclude the possibility that a particular individual may display substantial scatter that is of clinical significance.

In examining a profile for scatter, the clinician should bear in mind the following:

(1) Be cautious about using personal decision rules when evaluating subtest differences. Clinicians often assert that a discrepancy of 3 or more points between two subtests is

clinically significant; there is no valid reason for informally using the subtest standard deviation as an index of subtest difference. (2) A statistically reliable difference between two test scores is not necessarily abnormal. (3) Scatter increases with high IQ and greater education, and is constrained in persons with lower than average IQ. (4) Greater scatter is to be expected in scaled scores than in age-scaled scores. Most tables of abnormal differences use scaled scores. (5) A premorbid IQ estimate based on the single highest subtest score in a profile is unlikely to be valid.

The general thrust of the literature is that using subtest scatter as evidence of deficit, without collateral evidence, is risky. On a more positive note, there are other ways in which the WAIS-R is employed to quantify impairments that may prove to be more successful. One is the use of factors derived from principal components analysis as a basis for combining subtest scores. In particular, the MF factor may provide a useful contrast with the other two factors in the assessment of acute brain dysfunctions (Sherman et al., 1995). Another method with promise is the use of discrepancies between IQ and memory scores to establish dysmnesia. Typically, IQ and memory test scores are highly correlated, however, there is considerable evidence from the study of amnesics that low memory scores can occur in the presence of high IQ. Using WAIS-R IQ scores as a baseline against which WMS-R index scores can be evaluated may be a valid means of determining memory impairment. Finally, WAIS-R retest comparisons can be informative. Evidence from retested samples indicates that for most people, WAIS-R IQ scores increase on a second testing. Where there is any decline this should be scrutinised carefully, and a substantial deterioration can provide solid evidence for impairment in cases of clients with suspected progressive disorders.

A Final Word

At the onset of this article, the challenges posed to expert testimony about the results from intelligence tests offered by Ziskin and Faust (1988) were introduced. The purpose of their review was to instruct lawyers about how to attack the credibility of neuropsychologists' testimony in the adversarial setting of a courtroom. Much of the specific information they provide can be countered or is now outdated. Nonetheless, clinicians who anticipate giving testimony about neuropsychological test scores might be advised to review their texts as a guide to matters likely to be traversed by a well-prepared lawyer during cross examination. The tactics they advocate for discrediting the WAIS-R can be used with any neuropsychological test. Some questions about general issues that they suggest putting to the expert witness concern the nature of intelligence, the appropriateness of the norms, and the validity of using the test with particular clients, such as the mentally retarded, criminals, and persons with psychosis. This involves the expert in framing answers to such questions as: "Isn't it true that there is no consensus amongst psychologists about what intelligence is and how to measure it?", "Has the WAIS-R been standardised for use in New Zealand?", or "Wide spread usage of the WAIS-R by clinicians does not constitute any evidence for its reliability or validity, does it?" For a transcript example of a highly qualified expert in WAIS-R interpretation undergoing the discomfort of a well-informed cross-examination, the reader is referred to Appendix A of Ziskin and Faust (1988).

Most clinicians who administer the WAIS-R have no expectation of being examined about their findings in court. Nor is the courtroom the ideal environment to debate and determine good clinical practice. As Adams and Putnam (1994) have observed: "Psychologists called to court are best advised to act as psychologists first and best, allowing the legal arena to evolve as it must and will. Psychologists and lawyers are professionals with differing goals, cultures, and rules and who operate in epistomologically diverse ways" (p. 6). The value of the scrutiny of sceptics is not, however, to be denied. It provides a stimulus for the neuropsychologists to clarify, develop, and refine methods of practice. In inhibits the possibility of the stagnation of a profession with no uncertainties and complete faith in the tools to hand. It reinforces the need for neuropsychologists to have a sound basis in the science of psychology, including knowledge of psychometric theory and the foundations of cognitive psychology. And, finally, it may also serve to bolster the need for specialist training in neuropsychology and the development of national guidelines for practitioner training.

References

Abrams, R., Redfield, J., & Taylor, M.A. (1981). Cognitive dysfunction in schizophrenia, affective disorder and organic brain disease. *British Journal of Psychiatry, 139*, 190-194.

Adams, K.M., & Putnam, S.H. (1994). Coping with professional skeptics: A reply. *Psychological Assessment, 6*, 5-7.

Adams, R.L., Smigielski, J., & Jenkins, R.L. (1984). Development of a Satz-Mogel short form of the WAIS-R. *Journal of Consulting and Clinical Psychology, 52*, 908.

Alexander, G.E., Prohovnik, I. Stern, Y., & Mayeux, R. (1994). WAIS-R subtest profile and cortical perfusion in Alzheimer's disease. *Brain and Cognition, 24*, 24-43.

Athanasou, J.A. (1993). Patterns of performance on the verbal and performance subtests of the Wechsler adult intelligence scale - revised: Some Australian data. *Journal of Clinical Psychology, 49*, 102-108.

Atkinson, L. (1991a). Some tables for statistically based interpretation of WAIS-R factor scores. *Psychological Assessment, 3*, 288-291.

Atkinson, L. (1991b). Concurrent use of the Wechsler Memory Scale - Revised and the WAIS-R. *British Journal of Clinical Psychology, 30*, 87-90.

Atkinson, L., & Cyr, J.J. (1984). Factor Analysis of the WAIS-R: Psychiatric and standardization samples. *Journal of Consulting and Clinical Psychology, 52*, 714-716.

Barona, A. Reynolds, C.R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52*, 885-887.

Beck, N.C., Horwitz, E., Seidenberg, M., Parker, G., & Frank, R. (1985). WAIS-R factor structure in psychiatric and general medical patients. *Journal of Consulting and Clinical Psychology, 53*, 402-405.

Blaha, G., & Wallbrown, F.H. (1982). Hierarchical factor structure of the Wechsler adult intelligence scale - revised. *Journal of consulting and clinical psychology, 50*, 652-660.

Blaha, J., & Mandes, E. (1993). The hierarchical factor structure of the WAIS-R for alcoholic adults. *Journal of Clinical Psychology, 49*, 740-745.

Blair, J.R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the

National Adult Reading Test. *Clinical Neuropsychologist, 3*, 129-136.

Boone, D.E. (1992a). Reliability of the WAIS-R with psychiatric inpatients. *Journal of Clinical Psychology, 48*, 72-76.

Bornstein, R.A. (1983). Verbal IQ-Performance IQ discrepancies on the Wechsler Adult Intelligence Scale - revised in patients with unilateral or bilateral cerebral dysfunction. *Journal of Consulting and Clinical Psychology, 51*, 779-780.

Bornstein, R.A., & Matarazzo, J.D. (1982). Wechsler VIQ versus PIQ differences in cerebral dysfunction: A literature review with emphasis on sex differences. *Journal of Clinical Neuropsychology, 4*, 319-334.

Bornstein, R.A., Chelune, G.J., & Prifitera, A. (1989). IQ-memory discrepancies in normal and clinical samples. *Psychological Assessment, 1*, 203-206.

Bornstein, R.A., Suga, L., & Prifitera, A. (1987). Incidence of Verbal IQ-Performance IQ discrepancies at various levels of education. *Journal of Clinical Psychology, 43*, 387-389.

Bowman, M.L. (1989). Testing individual differences in Ancient China. *American Psychologist, 44*, 576-578.

Brinkman, S.D., & Braun, P. (1984). Classification of dementia patients by a WAIS profile related to central cholinergic deficiencies. *Journal of Clinical Neuropsychology, 6*, 393-400.

Brophy, A.L. (1986). Confidence intervals for true scores and retest scores on clinical tests. *Journal of Clinical Psychology, 42*, 989-991.

Burgess, A. (1991). Profile analysis of the Wechsler intelligence scales: A new index of subtest scatter. *British Journal of Clinical Psychology, 30*, 257-263.

Burgess, A., Flint, J., & Adshead, H. (1992). Factor structure of the Wechsler Adult Intelligence Scale - Revised (WAIS-R): A clinical sample. *British Journal of Clinical Psychology, 31*, 336-338.

Burt, C. (1940). *The factors of the mind.* London: University of London Press.

Canavan, A.G., & Beckman, J. (1993). Deriving principal component IQ scores from the WAIS-R. *British Journal of Clinical Psychology, 32*, 81-86.

Canavan, A.G., Dunn, G., & McMillan, T.M. (1986). Principal components of the WAIS-R. *British Journal of Clinical Psychology, 25*, 81-85.

Chapman, L.J., & Chapman, J.P. (1988). Artifactual and genuine relationships of lateral difference scores to overall accuracy in studies of laterality. *Psychological Bulletin, 104*, 127-136.

Cohen, J. (1957). The factorial structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology, 21*, 283-290.

Crawford, J.R. (1992). Current and premorbid intelligence measures in neuropsychological practice. In J.R. Crawford, D.M. Parker, & W.W. McKinlay (Eds.), *A handbook of neuropsychological assessment.* Hove, England: Lawrence Erlbaum Associates.

Crawford, J.R., & Allan, K.M. (1994). The Mahalanobis Distance index of WAIS-R subtest scatter: Psychometric properties in a healthy UK sample. *British Journal of Clinical Psychology, 33*, 65-69.

Crawford, J.R., & Allan, K.M. (1996). WAIS-R subtest scatter: Base-rate data from a healthy UK sample. *British Journal of Clinical Psychology, 35*, 235-247.

Crawford, J.R., Allan, K.M., Stephen, D.W., Parker, D.M., & Besson, J.A.O. (1989). The Wechsler Adult Intelligence Scale - Revised (WAIS-R): Factor structure in a UK sample. *Personality and Individual Differences, 10*, 1209-1212.

Crawford, J.R., Jack, A.M., Morrison, F.M., Allan, K.M., & Nelson, H.E. (1990). The UK factor structure of the WAIS-R is robust and highly congruent with the USA standardisation sample. *Personality and Individual Differences, 11*, 643-644.

Crawford, J.R., Parker, D.M., Stewart, L.E., Besson, J.A.O., & DeLacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology, 28*, 267-273.

Crawford, J.R., Stewart, L.E., Cochrane, R.H.B., Foulds, J.A., Besson, J.A.O., & Parker, D.M. (1989). Estimating premorbid IQ from demographic variables: Regression equations derived from UK sample. *British Journal of Clinical Psychology, 28*, 275-278.

Davis, F.B. (1959). Interpretation of Differences among averages and individual test scores. *Journal of Education Psychology, 50*, 162-170.

DeLemos, M.M. (1983). Australian adaptation of the WAIS-R: Report on the selection of items for the Australian version of the Information Test. *ACER Bulletin for Psychologists, 33*, 17-21.

DeWolfe, A.S., & Ryan, J.J. (1984). Wechsler Performance IQ > Verbal IQ index in a forensic sample: A reconsideration. *Journal of Clinical Psychology, 40*, 291-194.

Dudek, F.J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin*, 86, 335-337.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31-55.

Faust, D., Ziskin, J., & Hiers, J.B. (1991). *Brain damage claims: Coping with neuropsychological evidence.* Los Angeles, CA: Law and Psychology Press.

Feinberg, J.R., & McIlvried, E.J. (1991). WAIS-R subtest scatter in a chronic schizophrenic population: Is it an attention problem? *Journal of Clinical Psychology, 47*, 327-335.

Field, G.E., & Sisley, R.C. (1986). IQ score differences between the WAIS and the WAIS-R: Confirmation with a New Zealand sample. *Journal of Clinical Psychology, 42*, 986-988.

Filley, C.M., Kobayashi, J., & Heaton, R.K. (1987). Wechsler intelligence scale profiles, the cholinergic system, and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 9*, 180-186.

Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932-1978. *Psychological Bulletin, 95*, 29-51.

Fuld, P.A. (1984). Test profile of cholinergic dysfunction and of Alzheimer - type dementia. *Journal of Clinical Neuropsychology, 6*, 380-392.

Gfeller, J.D., & Rankin, E.J. (1991). The WAIS-R profile as a cognitive marker of Alzheimer's disease: A misguided venture? Journal of Clinical and Experimental Neuropsychology, 13, 629-636.

Goldman, R.S., Axelrod, B.N., Giordani, B.J. & Foster, N. (1992). Longitudinal sensitivity of the Fuld cholinergic profile to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 14*, 566-574.

Gutkin, T.B., Reynolds, C.R., & Galvin, G.A. (1984). Factor analysis of the Wechsler adult intelligence scale - revised (WAIS-R): An examination of the standardization sample. *Journal of School Psychology, 22*, 83-93.

Hunt, H.F. (1949). The relative rates of decline of Wechsler - Bellevue "hold: and "don't" tests. *Journal of Consulting Psychology, 13*, 440-443.

Kaplan, E., Fein, D., Morris, R., & Delis, D.C. (1991). *WAIS-R as a Neuropsychological Instrument.* New York: Psychological Corporation.

Kelley, Y.L. (1923). A new method of determining the significance of differences in intelligence and achievement scores. *Journal of Educational Psychology, 14*, 321-333.

Knight, R.G. (1983). On interpreting the several standard errors of the WAIS-R: Some further tables. *Journal of Consulting and Clinical Psychology, 51*, 671-673.

Knight, R.G., & Godfrey, H.P.D. (1984). Assessing the significance of differences between subtests on the Wechsler adult intelligence scale - revised. *Journal of Clinical Psychology, 40*, 808-810.

Knight, R.G., & Longmore, B.E. (1991). What is an amnesic? In W.C. Abraham, M.C. Corballis, and K.G. White (Eds.), *Memory mechanisms: A tribute to G.V. Goddard.* New York: Lawrence Erlbaum Associates.

Kolb, B., & Milner, B. (1981). Performance of complex arm and facial movements after focal brain lesions. *Neuropsychologia, 19*, 491-503.

Kramer, J.H. (1990). Guidelines for interpreting WAIS-R subtest scores. *Psychological Assessment, 2*, 202-205.

Krauskopf, C.J. (1991). Pattern analysis and statistical power. *Psychological Assessment, 3*, 261-264.

Lawson, J.S., & Inglis, J. (1983). A Laterality Index of cognitive impairment after hemispheric damage: A measure derived from a principal-components analysis of Wechsler Adult Intelligence Scale. *Journal of Consulting and Clinical Psychology, 51*, 832-840.

Lawson, J.S., Inglis, J., & Stroud, T.W.F. (1983). A laterality index of cognitive impairment derived from a principal - components analysis of the WAIS-R. *Journal of Consulting and Clinical psychology, 51*, 841-847.

Leckliter, I.N., Matarazzo, J.D., & Silverstein, A.B. (1986). A literature review

of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology, 42*, 332-342.

Lezak, M. (1994). *Neuropsychological testing*. (3rd. ed.). New York: Oxford University Press.

Logsdon, R.G., Teri, L., Williams, D.E., Vitiello, M.V., & Prinz, P.N. (1989). The WAIS-R profile: A diagnostic tool for Alzheimer's disease? *Journal of Clinical and Experimental Neuropsychology, 11*, 892-898.

Lord, F.M., & Novik, M.R. (1968). *Statistical theories of mental test scores*. Reading, MS: Addison-Wesley.

Matarazzo, J.D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th. ed.). New York: Oxford University Press.

Matarazzo, J.D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45*, 999-1017.

Matarazzo, J.D. (1991). Psychological assessment is reliable and valid: Reply to Ziskin and Faust. *American Psychologist, 46*, 882-884.

Matarazzo, J.D., Bornstein, R.A., McDermott, P.A., & Noonan, J.V. (1986). Verbal IQ vs. Performance IQ difference scores in males and females from the WAIS-R standardization sample. *Journal of Clinical Psychology, 42*, 965-974.

Matarazzo, J.D., Daniel, M.H., Prifitera, A., Herman, D.O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *Journal of Clinical Psychology, 44*, 940-949.

Matarazzo, J.D., & Herman, D.O. (1984a). Relationship of education and IQ in the WAIS-R standardization sample. *Journal of Consulting and Clinical Psychology, 52*, 631-634.

Matarazzo, J.D., & Herman, D.O. (1984b), Base-rate data for the WAIS-R: Test-retest stability and VIQ-PIQ differences. *Journal of Clinical Neuropsychology, 6*, 351-356.

Matarazzo, J.D., & Herman, D.O. (1985). Clinical uses of the WAIS-R: Test-retest stability and VIQ-PIQ differences in the WAIS-R standardization sample. In B.B. Wolman (Ed.), *Handbook of intelligence; Theories, measurements, and applications*. (pp. 899-932). New York: Wiley.

Matarazzo, J.D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS-R standardization sample. *Psychological Assessment, 1*, 186-191.

McDermott, P.A., Jones, J.N., Glutting, J.J., & Noonan, J.V. (1989). Typology and prevailing composition of core profiles in the WAIS-R standardization sample. *Psychological Assessment, 1*, 118-125.

McFie, J. (1975). *Assessment of organic intellectual impairment*. London: Academic Press.

McLean, J.E., Kaufman, A.S., & Reynolds, C.R. (1989). Base rates of WAIS-R subtest scatter as a guide for clinical and neuropsychological assessment. *Journal of Clinical Psychology, 45*, 919-926.

McLean, J.E., Kaufman, A.S., & Reynolds, C.R. (1990). WAIS-R subtest scatter using the profile variability index. *Psychological Assessment, 2*, 289-292.

McNemar, Q. (1957). On WAIS difference scores. *Journal of Consulting Psychology, 21*, 239-240.

Mittenberg, W., Hammeke, T.A., & Rao, S.M. (1989). Intrasubtest scatter on the WAIS-R as a pathognomonic sign of brain injury. *Psychological Assessment, 1*, 273-276.

Mittenberg, W., Thompson, G.B., Schwartz, J.A., Ryan, J.J., & Levitt, R. (1991). Intellectual loss in Alzheimer's dementia and WAIS-R intrasubtest scatter. *Journal of Clinical Psychology, 47*, 544-547.

Moore, A.D., Stambrook, M., Hawryluk, G.A., Peters, L.C. et al. (1990). Test-retest stability of the Wechsler Adult Intelligence Scale - Revised in the assessment of head-injured patients. *Psychological Assessment, 2*, 98-100.

Naglieri, J.A. (1982). Two types of Tables for use with the WAIS-R. *Journal of Consulting and Clinical Psychology, 50*, 319-321.

Nebes, R.D., Martin, D.C., & Horn, L.C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology, 93*, 321-330.

Nelson, H.E. (1982). *The National Adult Reading Test (NART): Test Manual*. Windsor, England: NFER - Nelson.

Nelson, J., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex, 4*, 234-244.

O'Carroll, R.E., & Gilleard, C.J. (1986). Estimation of premorbid intelligence in dementia. *British Journal of Clinical Psychology, 25*, 157-158.

O'Grady, K.E. (1983). A confirmatory maximum likelihood factor analysis of the WAIS-R. *Journal of Consulting and Clinical Psychology, 51*, 826-831.

Paolo, A.M., & Ryan, J.J. (1992). Unusual combinations of Verbal and Performance IQs in normal and neurologically impaired elderly. *Journal of Clinical Psychology, 48*, 230-233.

Parker, K. (1983). Factor analysis of the WAIS-R at nine age levels between 16 and 74 years. *Journal of Consulting and Clinical Psychology, 51*, 302-308.

Parker, K.C.H., & Atkinson, L. (1995). Computation of Wechsler Adult Intelligence Scale-Revised factor scores: Equal and differential weights. *Psychological Assessment, 7*, 456-462

Payne, R.W., & Jones, G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13*, 115-121.

Petrie, K., Dibble, C., Long-Taylor, M., & Ruthe, G. (1986). A New Zealand information subtest for the WAIS-R. *New Zealand Journal of Psychology, 15*, 23-26.

Piedmont, R.L., Sokolove, R.L., & Fleming, M.Z. (1989a). On WAIS-R difference scores in a psychiatric sample. *Psychological Assessment, 2*, 155-159.

Piedmont, R.L., Sokolove, R.L., & Fleming, M.Z. (1989b). An examination of some diagnostic strategies involving the Wechsler Intelligence Scales. *Psychological Assessment, 1*, 181-185.

Plake, B.S., Reynolds, C.R., & Gutkin, T.B. (1981). A technique for the comparison of the profile variability between independent groups. *Journal of Clinical Psychology, 37*, 142-146. Prifitera, A., & Ryan, J.J. (1983). WAIS-R/WAIS comparisons in a clinical sample. *Clinical Neuropsychology, 5*, 97-99.

Prigatano, G.P. (1978). Wechsler Memory Scale: A selective review of the literature. *Journal of Clinical Psychology, 34*, 816-832.

Randolph, C., Mohr, E., & Chase, T.N. (1993). Assessment of intellectual function in dementing disorders: Validity of WAIS-R short forms for patients with Alzheimer's, Huntington's, and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology, 15*, 743-753.

Rapaport, D., Gill, M.M., & Schafer, R. (1968). *Diagnostic Psychological Testing* (rev. ed.). New York: International Universities Press.

Rogers, R., Bagby, R.M., & Perera, C. (1993). Can Ziskin withstand his own criticism? Problems with his model of cross-examination. *Behavioural Sciences and the Law, 11*, 223-233.

Russell, E.W. (1992). Comparison of two methods for converting the WAIS to the WAIS-R. *Journal of Clinical Psychology, 48*, 355-359.

Ryan, J.J., & Bohac, D.L. (1994). Neurodiagnostic Implications of Unique Profiles of the Wechsler Adult Intelligence Scale - Revised. *Psychological Assessment, 6*, 360-363.

Ryan, J.J., Georgemiller, R.J., Gleisser, M.E., & Randall, D.M. (1985). Test-retest stability of the WAIS-R in a clinical sample. *Journal of Clinical Psychology, 41*, 552-556.

Ryan, J.J., & Paolo, A.M. (1992a). Base rates of intersubtest scatter in the old age standardization sample of the WAIS-R. *Archives of Clinical Neuropsychology, 7*, 515-522.

Ryan, J.J., & Paolo, A.M. (1992b). Verbal-Performance IQ discrepancies on the WAIS-R: An examination of the old-age standardization sample. *Neuropsychology, 6*, 293-298.

Ryan, J.J., & Paolo, A.M. (1992c). A screening procedure for estimating premorbid intelligence in the elderly. *Clinical Neuropsychologist, 6*, 53-62.

Ryan, J.J., Paolo, A.M., & Brungardt, T.M. (1990a). Standardization of the Wechsler Adult Intelligence Scale - Revised for person 75 years and older. *Psychological Assessment, 2*, 404-411.

Ryan, J.J., Paolo, A.M., & Brungardt, T.M. (1990b). Factor analysis of the Wechsler Adult Intelligence Scale - Revised for persons 75 years and older. *Professional Psychology Research and Practice, 21*, 177-181.

Ryan, J.J., Paolo, A.M., & Brungardt, T.M. (1992). WAIS-R test-retest stability in normal persons 75 years and older. *Clinical Neuropsychologist, 6*, 3-8.

Ryan, J.J., Paolo, A.M., & Smith, A.J. (1992). Wechsler Adult Intelligence

Scale - Revised intersubtest scatter in brain-damaged patients: A comparison with the standardization sample. *Psychological Assessment, 4*, 63-66.

Ryan, J.J., Prifitera, A., & Larsen, J. (1982). Reliability of the WAIS-R with a mixed patient sample. *Perceptual and Motor Skills, 55*, 1277-1278.

Ryan, J.J., Rosenberg, S.J., & DeWolfe, A.S. (1984). Generalization of the WAIS-R factor structure with a vocational rehabilitation sample. *Journal of Consulting and Clinical Psychology, 52*, 311-312.

Ryan, J.J., & Schneider, J.A. (1986). Factor analysis of the Wechsler adult intelligence scale - revised (WAIS-R) in a brain-damaged sample. *Journal of Clinical Psychology, 42*, 962-964.

Schinka, J.A., Vanderploeg, R.D., Curtiss, G. (1994). Wechsler Adult Intelligence Scale - Revised subtest scatter as a function of maximum subtest scaled score. *Psychological Assessment, 6*, 364-367.

Sharpe, K., & O'Carroll, R. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: A Canadian study. *British Journal of Clinical Psychology, 30*, 381-384.

Sherman, E.M.S., Strauss, E., Spellacy, F., & Hunter, M. (1995). Construct validity of WAIS-R factors: Neuropsychological test correlates in adults referred for possible head injury. *Psychological Assessment, 7*, 440-444.

Siegert, R.J., Patten, M.D., Taylor, A.J.W., & McCormick, I.A. (1988). Factor analysis of the WAIS-R using the factor replication procedure: FACTOREP. *Multivariate Behavioural Research, 23*, 481-489.

Silverstein, A.B. (1982a). Factor structure of the Wechsler Adult Intelligence Scale - Revised. *Journal of Consulting and Clinical Psychology, 50*, 661-664.

Silverstein, A.B. (1982b). Pattern Analysis as simultaneous statistical inference. *Journal of Consulting and Clinical Psychology, 50*, 234-240.

Silverstein, A.B. (1984a). Pattern Analysis; The question of Abnormality. *Journal of Consulting and Clinical Psychology, 52*, 936-939.

Silverstein, A.B. (1984b). Standard errors for short forms of Wechsler's intelligence scales with deviant subjects. *Journal of Consulting and Clinical Psychology, 52*, 913-914.

Silverstein, A.B. (1985). Verbal-Performance IQ discrepancies on the WAIS-R: Estimated vs. empirical values. *Journal of Clinical Psychology, 41*, 694-697.

Silverstein, A.B. (1989). Reliability and abnormality of scaled-score ranges. *Journal of Clinical Psychology, 45*, 926-929

Silverstein, A.B. (1991). Reliability of score differences on Wechsler's intelligence scales. *Journal of Clinical Psychology, 47*, 264-266.

Silverstein, A.B. (1993). Type I, Type II, and other types of errors in pattern analysis. *Psychological Assessment, 5*, 72-74.

Smith, D.H. (1988). The British adaptation of the WAIS-R: The vocabulary scale. *British Journal of Clinical Psychology, 27*, 261-262.

Snow, W.G., Tierney, M.C., Zorzitto, M.L., Fisher, R.H., & Reid, D.W. (1989). WAIS-R Test-Retest Reliability in a Normal Elderly Sample. *Journal of Clinical and Experimental Neuropsychology, 11*, 423-428.

Stebbins, G.T., Wilson, R.S., Gilley, D.W., Bernard, B.A., & Fox, J.H. (1990). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *Clinical Neuropsychologist, 4*, 64-68.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sturmey, P., Gatherer, A., Ghadiali, E., Hallett, S. & Searle, Y. (1993). The Wechsler Adult Intelligence Scale - Revised (WAIS-R): Factor structure in a British, neurologically impaired population. *Personality and Individual Differences, 14*, 255-257.

Sweet, J.J., Moberg, P.J., & Tovian, S.M. (1990). Evaluation of Wechsler Adult Intelligence Scale - Revised premorbid IQ formulas in clinical populations. *Psychological Assessment, 2*, 41-44.

Urbina, S.P., Golden, C.J., & Ariel, R.N. (1982). WAIS/WAIS-R: Initial comparisons. *Clinical Neuropsychology, 4*, 145-146.

Waller, N.G., & Waldman, I.D. (1990). A reexamination of the WAIS-R factor structure. *Psychological Assessment, 2*, 139-144.

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

Wechsler, D. (1941). Diagnostic and clinical features of the Wechsler-Bellevue scales. In D. Wechsler (Ed.), *The measurement of adult intelligence*, (2nd. ed.), (pp. 145-162). Baltimore: Williams & Wilkins.

Wechsler, D. (1944). The problem of mental deterioration. In D. Wechsler (Ed.), *The measurement of adult intelligence*, (3rd. ed.), (pp. 54-69). Baltimore: Williams & Wilkins.

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

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th. ed.). Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale - Revised*. New York: Psychological Corporation.

Weins, A.N., Bryan, J.E., & Crossen, J.R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test - Revised in normal subjects. *Clinical Neuropsychologist, 7*, 70-84.

Willshire, D., Kinsella, G., & Prior, M.R. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: A cross-validation. *Journal of Clinical and Experimental Psychology, 13*(2), 204-216.

Zillmer, E.A., Waechtler, C., Harris, B., Khan, F., & Fowler, D. (1992). The effects of unilateral and multifocal lesions on the WAIS-R. *Archives of Clinical Neuropsychology, 7*, 29-40.

Ziskin, J., & Faust, D. (1988). *Coping with psychiatric and psychological testimony* (Volumes 1-3, 4th. ed.). Loss Angeles: Law and Psychology Press.

Address for correspondence:
Department of Psychology
University of Otago
Box 56, Dunedin
E-Mail: SYCO10@OTAGO.AC.NZ
Fax: (03) 479-8335