*New Zealand Journal of Psychology,* 1984, *13,* 25-31

# Interpreting Stanford-Binet and WISC-R IQs in New Zealand: The need for more than caution.

Keith D. Ballard
University of Otago.

A recent caution on the use of the Stanford-Binet and WISC-R Scales in New Zealand was based on problems evident in using American norms. The present paper identifies further difficulties in interpreting these two tests arising from a lack of reliability and validity data and from problems relating to lack of stability in the trait being measured and evidence for poor prediction of individual achievement.

IQ scores derived from either the Stanford-Binet Intelligence Scale or the Wechsler Intelligence Scale for Children (Revised) are normally required by the New Zealand Department of Education for determining children's eligibility for admission to a Special Class for Backward Children, to Special Schools for the Intellectually Handicapped, and to some residential facilities for children with special needs (Notes 1, 2, 3). These tests may also be used in assessments for other purposes, such as identifying children having learning difficulties (e.g. Note 4) and recently the Tests and Standards Committee of the New Zealand Psychological Society commented on aspects of training in the use of both instruments (Tests and Standards Committee, 1983).

A recent report proposed caution in interpreting the Stanford-Binet and WISC-R in New Zealand on the basis that mean scores for four to six year old Dunedin children differed from American norms (Silva, 1982). Other available information, however, indicates that simple caution and concern with foreign norms is not an adequate position for the psychologist to hold regarding these two tests. It is suggested in the present paper that neither test should be used for individual assessment as the interpretation of scores cannot be made with scientific or ethical justification.

## IQ, intelligence, potential and aptitude: Enduring characteristics of the individual

IQ scores on the Stanford-Binet and WISC-R tests are often claimed to be measures of "general ability" or "potential" for academic achievement (Wright & Dappen, 1982). Such a claim implies that the IQ is related to a trait of "mental ability" (Messé, Crano, Messé & Rice, 1979) which determines ability to learn (Bradley & Howe, 1980). The idea that IQ measures a person's potential for learning implies that it taps a relatively stable personal characteristic, which further implies that an IQ score for an individual measured at one time should usefully predict the score that would be achieved on measurement at some future time. However, such stability of individual scores has not been found. It is generally acknowledged that stability of test scores in the preschool years is extremely poor, but Clarke (1978) and Mittler (1974) stress that what is frequently not appreciated is the lack of stability in IQ for individual children across the school years.

Clarke (1978) suggests that it may not always be appreciated that quite high stability coefficients still allow considerable variation in individual performance. Stanley and Hopkins (1972, pp 341-342), for example, cite data on predicting children's Stanford-Binet Grade 12 IQ's from their IQ's on the same test at Grade 3. The correlation between Grade 3 and Grade 12 IQ's is .80 in this instance. Stanley and Hopkins calculate the Standard Error of Estimate for the data as 10 IQ points. They point out that this means that one-third of the children will differ from their predicted Grade 12 IQ's by 10 or more points, while a .95 confidence interval for an individual's Grade 12 IQ would span a range of more than 40 IQ points.

There is well documented evidence for large changes in IQ scores for individual children, with individuals fluctuating by as much as 25 IQ points over the school years (Richards, 1951; Sontag, Baker & Nelson, 1958). In the primary school period alone Hindley and Owen (1978) found a quarter of their longitudinal sample of

British children changing by at least 15 IQ points, while between 11 and 17 years more than half the children changed by 10 points and a quarter by 19 or more.

Mittler (1974), Brody and Brody (1976) and Hindley and Owen (1978) have stressed that IQ scores for individuals must be recognised as unstable, and that insufficient emphasis has been given to the frequency of large changes in scores. Along with Clarke (1978), they have urged psychologists to recognise that test-retest correlations as high as .90 can conceal large changes over time in individual scores, making classification and placement decisions extremely hazardous.

Measurement of enduring traits or dispositions in individuals demands both high reliability and high stability, and, as Jensen (1980, p. 262) notes, for the prediction of the future status of an individual high stability is essential. Research data suggests that psychologists should not interpret IQ scores using terms such as "aptitude" or "potential", as these terms imply a relatively enduring trait or disposition. The 1974 edition of *Standards for Educational and Psychological Tests* (Note 5) reports that IQ is "frequently reified and interpreted as representing an unchanging and unchangebale characteristic" (p.23) and suggests that test users should avoid using the term IQ as it involves "spurious projections of growth" (p. 70).

## IQ and academic achievement

Another use of scores on the WISC-R and Stanford-Binet derives from the belief that such scores have a meaningful concurrent or predictive relationship with academic performance. However, the idea that such intellectual assessments are strongly related to academic attainments is difficult to support from the empirical literature. In part, this is a consequence of the lack of stability of IQ scores.

Nevertheless, a major basis of support for the use of IQ tests in assessing children lies in studies showing a correlation between IQ scores and scholastic achievement. Jensen (1980) states that ". . . even the harshest critics of mental testing wholly concede the substantial relationship between IQ and scholastic achievement . . ." (p. 316) which he identifies as typically a correlation of .50 (p. 317), a figure, says Jensen, that is "a good estimate of the overall average of all estimates of the predictive validity of IQ for scholastic achievement" (p. 318). Reviews of the literature regarding the relationship of IQ test scores to school attainments put the correlations usually obtained between IQ and academic attainment at around .5 to .6 (Jensen, 1980, p. 317; Lavin, 1965; Messé et al., 1979), and Jensen (1980) writes that it is ". . . important to note that IQ accounts for only about half, or less, of the variance in measured achievement at any given point in the course of schooling" (p. 321). Much higher correlations are reported between IQ scores and standardised achievement test scores. For example, Oakland (1982) obtained a correlation of .70 between WISC-R and California Achievement Test Reading test scores, and Stanley and Hopkins (1972, p. 346) cite correlations between IQ scores (Lorge Thorndike) and both vocabulary and reading achievement test scores (Iowa Tests of Basic Skills) of .82. It is not surprising, however, that one form of achievement test (IQ) correlates well with another (Miller & Davis, 1981; Wesman, 1968).

In their paper on the predictive validity of mental ability tests, Messé et al. (1979) state that their results showed a "reasonably impressive" relationship between mental ability test results and the classroom performance of just over 4,000 British school children in Grades 2 to 4, concluding that "the standardized test of mental ability provides a powerful instrument for the prediction of academic achievement" (p. 240). Over the total sample the correlation between the mental ability test and Grade Point Average (based on teacher rating of performance in reading and arithmetic) was .60, accounting for 36% of the common variance. It is important to ask if in fact this relationship suggests powerful prediction for academic tasks.

Given such a correlation it is possible to evaluate how well a mental ability test will predict academic achievement for individuals. Any such estimate of future performance will include an error component, which can be described by the Standard Error of Estimate (SEest), a statistic seen by Anastasi (1982) as "particularly well suited to the interpretation of individual scores" (p. 125). With the SEest, the psychologist can choose a level of confidence for prediction based on the normal probability curve. Given that prediction can have a profound effect on a child's placement in the shool curriculum, then it would seem essential to work with at least the 95% confidence interval (i.e. $\pm$ 2 SEest).

As an illustration, assume that an IQ test has a mean score of 100 and a Standard Deviation of 15, while School Certificate English scores have a mean of 50 and a Standard Deviation of 15. If the correlation between IQ and School Certificate English is .60 then the SEest is 12. Then, for Jane, who has an IQ score of 110, we can be 95% confident that her actual School Certificate English mark will fall in the range of 32-80 marks. For Simon, who achieved an IQ of 75, the corresponding range is 11-59. Using IQ as the predictor variable, therefore, it can only be said that it is possible that both Jane and Simon might pass School Certificate English, and it is also possible that they might both fail School Certificate English. Although it might appear that Jane is more likely to pass than Simon, there is no way of being confident that this will be the actual outcome. IQ scores have, at best, a tenuous relationship to school achievement when it comes to predicting individual performance.

## Some issues in test construction reliability and validity

The American Psychological Society et al. (Note 5) cite as an essential requirement of a test manual that it should describe the "rationale" for the test (p. 11). Littell (1960), in his review of literature on the WISC, wrote that the test lacked an adequate rationale, and termed the lack of evidence for predictive validity after a decade of research "appalling". The review of the WISC-R by Freides (1978) notes that there is little in the manual to explain or justify what is being measured. Freides claims that Wechsler is "deliberately inexplicit about what is being measured . . ." as explicitness "would require a comprehensive theoretical stance" and "the state of knowledge does not permit such accomplishment as yet . . ." (p.350). The reviews by Tittle (1975) and Whitworth and Petrosko, reported in Buros (1978, pp 351-355), all note that the absence of discussion on both construct and predictive validity is a major weakness of the WISC-R manual.

The test users then, must do their own research to justify what they are measuring and to support any inferences they may make on the basis of test scores. They have a substantial literature to choose from. Writers such as Jensen (1980) and Anastasi (1983), for example, seem confident that IQ tests measure something meaningful. Others, such as Cronbach (1969), Miller (1981) and Sabatino, Miller and Schmidt (1981) suggest that there is little agreement on the definition and nature of intelligence, which suggests problems for discussions of construct validity. If there is uncertainty about the construct, and therefore about the domain being sampled, then content validity presents a formidable challenge (Messick, 1980; Miller & Davis, 1981; Wesman, 1968). Validity is the most important characteristic of a measuring instrument (Ahmann & Glock, 1971; Messick, 1980), so it is interesting that in such a prominent test as the WISC-R he manual fails to address such critical issues. Miller and Davis (1981) may be correct when they claim in their review on measuring intelligence that apart from some agreement that the tests "evaluate the educational foundation established by the Binet Scales", the ". . . construct of intelligence, as it is measured by current techniques, remains a matter of speculation . . ." (p. 192) and "validity is apparently so problematic for the measurement of intelligence that discussions of the topic are often absent in the test manuals." (p. 196).

Perhaps, then, the test user will be comfortable with the idea of intelligence measurement as a "pragmatic accomplishment" (Freides, 1978, p. 349). The tests would thus measure how well a child answered some specific questions on one specific occasion compared to a specific group of people (Fischer, 1974), the questions essentially involving facts and logical problem solving, accomplishments determined as important by certain cultural values and standards (Samuda, 1975). It would be important to recognise that different cultures value different cognitive skills (Anastasi, 1983) and to question the notion of a uniform criterion of ability in a pluralistic society (Miller & Davis, 1981). It would also be important to accept that a lack of equivalent opportunity to learn the content areas of the test means that the test is not able to assess a child's "ability to learn" (Miller & Davis, 1981, p. 194).

A pragmatic perspective would help emphasise the fact that the WISC-R and Stanford-Binet are devices for ranking children according to how well they answer a given set of questions, and perhaps deemphasise attention to hypothetical ability constructs as if they were entities within each child. (Fischer, 1974).

To achieve a meaningful ranking it is essential to show that it is meaningful to compare the child tested with the children in the standardisation group (American Psychological Association et

al, Note 5; Salvia & Ysseldyke, 1978). It is necessary, therefore, for the psychologist to argue that it is sensible, in the case of the WISC-R, to compare a New Zealand child with American children tested in 1972. It would seem difficult to justify such a comparison, especially given Silva's (1982) data showing that the American norms do not apply to some New Zealand children.

With the Stanford-Binet the ranking process seems to lack any foundation. Not only are there no New Zealand norms, but for the 1972 revision the American normative sample is both inadequately constructed and described (Salvia & Ysseldyke, 1978). In her review of the Stanford-Binet since the 1972 restandardization, Waddell (1980) noted that details of the standardisation group are incomplete, reliability data are entirely missing from the manual and could not be located in the research literature, and validity data involve largely a comparison with other measures of intelligence. But especially notable is that no predictive validity studies are available in the literature, a fact seen by Waddell as presenting serious ethical, and possibly legal, implications when the test is used in placement decisions.

In *Standards for Educational and Psychological Tests* (American Psychological Association et al., Note 5) it is described as an essential practice for users of normative tests to "interpret an obtained score with reference to sets of norms appropriate for the individual tested and for the intended use" (p. 70). This is clearly not possible for the Stanford-Binet in any circumstances, as there is a lack of the most basic psychometric data on standardisation and reliability. For the WISC-R the user would need to justify using American norms knowing that they probably differ from New Zealand norms. Reliability for the WISC-R was calculated on standardisation sample data, so essentially the reliability of this test for New Zealand children is unknown.

Psychologists required to defend their use of the Stanford-Binet and WISC-R would have a challenging time confronting the literature on these tests, particularly the serious lack of information on the Stanford-Binet. Certain test design features could also present a challenge to the test user's logic. For example, in establishing norms for the 1937 revision of the Stanford-Binet, it was found that girls scored consistently superior to boys by a few points at every age. All items contributing to this superiorty were subsequently removed from the test. As Kaye (1973) points out, this was called making the test "sex-fair", "more fair for boys, less fair for girls" (p. 13). Or, supposing the psychologist had felt it relevant to compare a New Zealand child with 200 American children tested in 1972, and cited an IQ based on the WISC-R. If the child had been tested on the Stanford-Binet scale, would the child still have the same IQ? Bloom, Raskin and Reese (1976) found a correlation of .81 between these two tests, but also showed that in 27 of the 50 children in their study, scores on the two tests did not fall in the same classification category of mild retardation, moderate retardation, and so on. In 12 cases the true score ranges ($\pm$ 1 $SE_m$) for scores on the two tests did not overlap. A child placed into a special class on the basis of one of these tests could quite possibly be denied access to this placement on the basis of their score on the other test.

Interpretation of components of the Stanford-Binet (e.g. vocabulary) and subtests of the WISC-R presents further problems regarding the scientific basis for such analysis. The subtests cannot be seen as adequately sampling items or behaviours within such domains as "vocabulary", "comprehension" or "perceptual organisation", and were not designed with either construct or content validity in each area (Kaye, 1973). Coles (1978), in his detailed review of the literature, found little support for ideas concerning the relation of specific Wechsler subtest scores to academic behaviours. Likewise, reviews by Duffy, Salvia, Tucker and Ysseldyke (1981), Brody and Brody (1976) and Jensen (1980) found no justification for interpreting patterns or profiles in the Wechsler tests, contrary, says Jensen, "to much of the clinical folklore . . ." (p.279).

Overall, then, a case can be made that there is little room for even cautious interpretations of Stanford-Binet and WISC-R IQ scores that could stand up to basic psychometric requirements. Yet there is evidence that psychologists who use such tests may be heavily influenced by intelligence test data, especially in making placement decisions (Matuszek & Oakland, 1979). There is also evidence that some psychologists are so influenced by intelligence test scores that they reject other data regarding their client that conflicts with the test score data, in addition to rejecting facts regarding the technical inadequacy of the

tests (Ysseldyke, Algozzine, Regan & Potter, 1980). In the IQ testing area, as in the area of personality testing (Wade & Baker, 1977), there appears to be resistance to the implications of research data, and the use of IQ tests in making educational decisions has achieved a "functional autonomy" (Astin, 1961), independent of what is known about the tests.

*Some concluding comments on interpreting the WISC-R and Stanford-Binet IQ scores of New Zealand Children*

*Interpretation and test norms.* A child's score on a norm referenced test reflects his or her performance relative to the performance of children on whom the test was standardised If a test "is not properly standardised for a particular student such comparisons are meaningless" (Duffey et al., 1981, p. 431). With the Stanford-Binet, even the American comparison group is virtually unknown (Salvia & Ysseldyke, 1978; Waddell, 1981) and this fact, as well as the test's lack of reliability and validity data, should preclude the interpretation of this test with any children (Salvia & Ysseldyke, 1978, p. 462).

The comparison group for the WISC-R is 200 American children tested 11 years ago. It is not clear how, or why, test users would wish to justify comparing a New Zealand child with such a group. The evidence in Silva's (1982) report suggests that New Zealand norms would most probably differ from the American data. Apart from any comparison problems that this suggests, it will also be the case that the reliabiltiy of the WISC-R is unknown for New Zealand children (Salvia & Ysseldyke, 1978; Wesman, 1973). These grounds alone should preclude the interpretation of this test with New Zealand children. As Salvia and Ysseldyke (1978) suggest, "The zeal to help a child is not justification for using technically inadequate tests. Tests with inadequate norms must not be used to rank children in comparison to unspecified populations . . ." as there is no way of making "adequate norm referenced interpretations" (p. 464).

*IQ scores and the concept of "potential".* Interpretation of the test results that include the notion of "ability" or "potential" imply a prediction about future learning. These predictions cannot be supported by research data which, to the contrary, shows that an individual's future learning cannot be predicted from an IQ score. Messick (1980) claims that test validity essentially involves the inferences about people that can be made and justified on the basis of test scores, and that to ensure ethical interpretation of test scores it is critical to evaluate the potential social consequences of score interpretations. IQ scores are frequently seen as measures of "an encompassing ability related to many criteria . . ." (Miller & Davis, 1981, p. 193). The scores often have a major impact on placing limitations on what might be expected in learning and development from individual children (Miller & Davis, 1981; Ysseldyke et al., 1980). As there is no empirical basis for such interpretations, the use of the terms "ability" or "potential" in the context of IQ scores is not valid, and given the social consequences of such interpretations the use of these terms in this context may also be seen as unethical (Messick, 1980).

The most obvious prediction in which the WISC-R and Stanford-Binet is used in New Zealand occurs when a child is removed from the mainstream education system and placed in a Special Class for Backward Children or Special School. Such a placement involves the prediction that a child's future learning will be slowly paced and limited in extent. Evidence suggests that once placed in the special education system children are very rarely moved back into regular classes (Wilton, Glynn, Wotherspoon & McGinley, 1983). These placements then, clearly have profound significance for the children involved. Given the lack of an empirical base for interpreting WISC-R and Stanford-Binet scores in New Zealand, the use of these tests as part of such classification and decision making represents a serious threat to the credibility and ethics of both the test user and the categorical classifications prescribed by the education system.

*Assessment for classification or assessment for teaching?* Evidence from longitudinal studies shows that discontinuities are as common as continuities in human development, and that marked changes in development occur in response to changes in the individual's environment for both intellectually handicapped (e.g. Baller, Charles & Miller, 1967) and non handicapped individuals (Clarke, 1978). It would seem relevant, then, to replace attempts at predicting children's futures with attempts to provide optimum learning environments for children.

Present classification systems that attempt to use ability grouping as a way of identifying children with similar educational needs are of

doubtful validity, and it is increasingly recognised that similar teaching strategies are effective across various disability groups (Heward, Cooper, Heron, Hill, McCormick, Porter, Stephens & Sutherland, 1981). The classification of children for teaching purposes is a simplistic strategy that diverts attention from the need to provide the educational resources necessary to meet diverse individual needs.

As regards designing programmes for children, research has failed to support either the existence or relevance for teaching of hypothetical abilities and traits speculated to underlie children's academic performance. The evidence is considerably more in favour of direct evaluation of the difficulties children have with specific academic tasks, and for assessment to include ongoing evaluation of the appropriateness and quality of the teaching materials and strategies provided for children (Arter & Jenkins, 1979; Coles, 1978; Lahey, Vosk & Habif, 1981; McCarthy, Lund & Bos, 1983; Schroeder, Schroeder & Davine, 1978).

## Reference Notes


1. Department of Education. *Psychological Service handbook*. Section H20, Special and Experience Classes for Backward Children. Wellington: N. D.
2. Department of Education. Procedure for admission to and discharge from Campbell Park and Salisbury Schools. Memorandum E50/1/18, E50/4/3, E50/5/3. Wellington: November, 1983.
3. Wilden, C. J. The occupational and social adjustment of experience class leavers in Otago 1969-73. M. A. Thesis in Education, University of Otago, 1974.
4. Chapman, J. W., St. George, R. & Van Kraayenoord, C. E. Short WISC-R's, Shortcuts and Occam's Razor: Identification of learning disabled pupils. Paper presented at the Annual Conference of the New Zealand Psychological Society, Auckland, August 1983.
5. American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for educational and psychological tests*. Washington: American Psychological Association, Inc., 1974.


## References


Ahmann, J. S. & Glock, N. D. (1971). *Evaluating pupil growth: Principles of tests and measurements* (4th ed.). Boston: Allyn & Bacon.

Anastasi, A. (1982) *Psychological testing* (5th ed.). New York: MacMillan.

Anastasi, A. (1983). Evolving trait concepts. *American Psychologist, 38,* 175-184.

Arter, J. A. & Jenkins, J. R. (1979). Differential diagnosis — prescriptive teaching: A critical appraisal. *Review of Educational Research, 49,* 517-555.

Astin, A. (1961). The functional autonomy of psychotherapy. *American Psychologist, 16,* 75-78.

Baller, W. R. , Charles, D. C. & Miller, E. L. (1967). Life attainment of the mentally retarded: A longitudinal study. *Genetic Psychology Monographs, 75* 235-329.

Bloom, A. S., Raskin, L. M. & Reese, A. H. (1976). A comparison of WISC-R and Stanford-Binet intelligence scale classifications of developmentally disabled children. *Psychology in the Schools, 13,* 288-290.

Bradley, C. McD. & Howe, C. E. (1980). Administrative issues in evaluation of exceptionality. *Exceptional Education Quarterly, 1,* 1-12.

Brody, E. B. & Brody, N. (1976) *Intelligence: Nature, determinants and consequences*. New York: Academic Press.

Buros, O. K. *The eighth mental measurement yearbook*. Vol. 1. New Jersey: Gryphon Press, 1978.

Clarke, A. D. B. (1978). Predicting human development. *Bulletin of the British Psychological Society, 31,* 249-258.

Coles, G. S. (1978). The learning-disabilities test battery: Empirical and social issues. *Harvard Educational Review, 48,* 313-340.

Cronbach, L. J. (1969). Heredity, environment and educational policy. *Harvard Educational Review, 39,* 338-347.

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

Duffey, J. B. Salvia, J., Tucker, J. & Ysseldyke, J. (1981) Nonbiassed assessment: A need for operationalism. *Exceptional Children, 47,* 427-434.

Epstein, S. (1980). The stability of behavior II: Implcations for psychological research. *American Psychologist, 35,* 790-806.

Fischer, C. T. (1974). Exit IQ: Enter the child. In G. J. Williams & S. Gordon (Eds.), *Clinical child psychology: Current practice and future perspectives*. New York: Behavioral Press.

Freides, D. (1978) Review of the WISC-R. In O. K. Buros (Ed.); *The eighth mental measurement yearbook*. (Vol. 1). New Jersey: Gryphon Press, pp. 349-352.

Heward, W. L., Cooper, J. O., Heron, T. E., Hill, D. S., McCormick, S., Porter, J. T., Stephens, T. M. & Sutherland, H. A. (1981). Noncategorical teacher training in a state with categorical certification requirements. *Exceptional Children, 48* (3), 206-212.

Hindley, C. B. & Owen, C. F. (1978). The extent of individual changes in IQ for ages between six months and seventeen years in a British longitudinal sample. *Journal of Child Psychology and Psychiatry, 19,* 329-350.

Jensen, A. R. (1980) *Bias in mental testing*. New York: Free Press.

Kaye, K. (1973) IQ: A conceptual deterrent to revolution in education. *Elementary School Journal, 74,* 9-23 (150).

Lahey, B. B., Vosk, B. N. & Habif, V. L. (1981). Behavioral assessment of learning disabled children: A rationale and strategy. *Behavioral Assessment, 3,* 3-14.

Lavin, D. E. (1965). *The prediction of academic performance: A theoretical analysis and review of research*. New York: Russell Sage Foundation.

Littell, W. M. (1970) The Wechsler Intelligence Scale for Children: Review of a decade of research. *Psychological Bulletin, 57,* 132-156.

Locke, A. (1978). The case against assessment. *Bulletin of the British Psychological Society, 31,* 322-324.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

Matuszek, P. & Oakland, T. (1979). Factors influencing teachers' and psychologists' recommendations regarding

special class placement. *Journal of School Psychology, 17,* 116-125.

McCarthy, J. M., Lund, K. A. & Bos, C. S. (1983). Assessment of young children with special needs. *Focus on Exceptional Children, 15,* (5), 1-11.

Messé, L. A., Crano, W. D., Messé, S. R. & Rice, W. (1979). Evaluation of the predictive validity of tests of mental ability for classroom performance in elementary grades. *Journal of Educational Psychology, 71,* 233-271.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35,* 1012-1027.

Miller, T. L. (1981). Introduction to Special Issue: The training of intelligence, implications for special education. *The Journal of Special Education, 15,* 103-104.

Miller, T. L. & Davis, E. E. (1981). Can change in intelligence be measured by contemporary techniques? *The Journal of Special Education, 15,* 185-200.

Mittler, P. J. (1974). Assessment of handicapped children: Some common factors. In P. Mittler (Ed.), *The psychological assessment of mental and physical handicaps.* London: Tavistock Publications (pp. 343-373).

Oakland, T. (1983). Joint use of Adaptive Behavior and IQ to predict achievement. *Journal of Consulting and Clinical Psychology, 51,* 298-301.

Richards, T. W. (1951). Mental test performance as a reflection of the child's current life situation — a methodological study. *Child Development, 22,* 221-233.

Sabatino, D. A., Miller, P. F. & Schmidt, C. (1981). Can intelligence be altered through cognitive training. *The Journal of Special Education, 15,* 126-144.

Salvia, J. & Ysseldyke, J. E. (1978). *Assessment in special and remedial education.* Boston: Houghton Mifflin Company.

Samuda, R. J. (1975). *Psychological testing of American minorities: Issues and consequences.* New York: Dodd, Mead.

Schroeder, C. S., Schroeder, S. R. & Davine, M. A. (1978). Learning disabilities: Assessment and management of reading problems. In B. B. Wolman, J. Egan & A. O. Ross (Eds.), *Handbook of treatment of mental disorders in childhood and adolescence.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc. (pp. 212-237).

Silva, P. A. (1982) Interpreting Stanford-Binet and WISC-R IQs in New Zealand. *New Zealand Journal of Educational Studies, 17,* 195.

Sontag, L. W. , Baker, C. T. & Nelson, V. O. (1958). Mental growth and personality development: A longitudinal study. *Monograph of the Society for Research in Child Development, 23,* No. 2.

Stanley, J. C. & Hopkins, K. D. (1972). *Educational and psychological measurement and evaluation.* Englewood Cliffs, New Jersey: Prentice-Hall.

Tests and Standards Committee, New Zealand Psychological Society. A report. (1983, May). *Bulletin of the New Zealand Psychological Society,* No. 36, 5-7.

Waddell, D. D. (1980). The Stanford-Binet: An evaluation of the technical data available since the 1972 restandardization. *Journal of School Psychology, 18,* 203-209.

Wade, T. C. & Baker, T. B. (1977). Opinions and use of psychological tests: A survey of clinical psychologists. *American Psychologist,* 874-882.

Wesman, A. G. (1968) Intelligent testing. *American Psychologist, 23,* 267-274.

Wesman, A. G. (1973) Reliabiltiy and confidence. In L. R. Aiken, Jr. (Ed.), *Readings in psychological and educational testing.* Boston: Allyn & Bacon.

Wilton, K., Glynn, T., Wotherspoon, T. & McGinley, C. (1983). Special educational placement decisions for handicapped children: An initial New Zealand study. *The Exceptional Child, 30,* 133-141.

Wright, D. & Dappen, L. (1982). Factor analysis of the WISC-R and the WRAT with a referral population. *Journal of School Psychology, 20,* 306-312.

Ysseldyke, J., Algozzine, B., Regan, R. & Potter, M. (1980). Technical adequacy of tests used by professionals in simulated decision making. *Psychology in the Schools, 17,* 202-209.