

Brief Reports

Disparities in Mentally Retarded Persons' IQs Derived From Different Intelligence Tests

HERMAN H. SPITZ

E. R. Johnstone Training and Research Center

For mildly and moderately mentally retarded individuals, the Wechsler Adult Intelligence Scale—Revised measures at about the same level as the Wechsler Adult Intelligence Scale, and, consequently, it also produces much higher IQs than do the Wechsler children's scales or the Stanford-Binet Intelligence Scale, Form L-M, despite high intertest correlations. This unsatisfactory state of affairs is due primarily to inadequate standardization sampling, particularly in the lower range of intelligence. Rather than abandon the IQ, corrective action should be taken.

Evidence has been accumulating that lack of adequate standardization samples in the lower range of intelligence has resulted in intertest inconsistency of IQs derived from widely used individual intelligence tests (Flynn, 1985; Spitz, 1983). The Wechsler scales (Wechsler, 1949, 1955, 1974, 1981) have been particularly subject to criticism because in the retarded range the Wechsler Adult Intelligence Test (WAIS) measures appreciably higher than does the Wechsler Intelligence Scale for Children (WISC) and the revised version of the WISC (the WISC-R) (Spitz, 1983). Also, the WISC measures higher than the WISC-R (Flynn, 1985).

With rather small samples, Spitz (1983) and Simon and Clopton (1984) reported that the intertest IQ constancy of the WAIS and WAIS-R is very high, and therefore the WAIS-R fails to correct the disparity between the adult and children's scales. In the present study I used the files of a residential institution for mentally retarded adolescents and young adults to make a large sample comparison of WAIS and WAIS-R

IQs. For smaller samples, I also assessed the IQ constancy between a number of other widely used intelligence tests. The results are given in Tables 1 and 2. Only full-scale IQs are given. None of the data were used in Spitz (1983).

When the WAIS-R was administered an average of 2.57 years after the WAIS, the mean IQ remained about the same, corroborating the findings of Spitz (1983) and Simon and Clopton (1984). By inference, the WAIS-R must also measure higher than the Wechsler children's scales. I now report direct evidence for this inference with, however, smaller samples. (Note that for this population the WAIS-R is very reliable and stable, as can be seen in Table 1.)

Fourteen subjects had a mean IQ of 64.07 on the WAIS-R an average of 2.07 years after they had obtained a mean IQ of 52.14 on the WISC-R. Similarly, 9 subjects' mean WISC IQ of 50.89 rose to 61.56 on the WAIS-R given, on the average, 2.33 years later. The data given in Table 1 are for the combined 23 subjects and indicate that the WAIS-R, like the WAIS, measures considerably higher than the Wechsler children's scales; in this case higher by .76 of the test's standard deviation (*SD*). These results are consistent with studies in which investigators compared the WAIS with the

Requests for reprints should be sent to Herman Spitz, E. R. Johnstone Training and Research Center, Bordentown, NJ 08505.

TABLE 1
INTER- AND INTRATEST IQ DATA

Test order ^a	N	Age, 1st test ^b		IQ, 1st test		Retest interval ^b		IQ, 2nd test		Diff.	r
		Mean	SD	Mean	SD	Mean	SD	Mean	SD		
WAIS, WAIS-R	91	23.09	3.73	60.02	8.38	2.57	.50	60.75	5.40	.73	.72*
WAIS-R, WAIS-R	31	22.93	4.14	58.42	5.92	2.53	.83	59.03	5.63	.61	.92*
WISC & WISC-R, WAIS-R	23	15.27	1.06	51.65	9.56	2.17	.69	63.09	6.22	11.44*	.77*
S-B, WAIS-R	12	22.74	4.35	38.92	9.04	2.46	.75	58.17	9.09	19.25*	.86*

^a Listed in order of administration.

^b In years.

* $p < .001$.

WISC or WISC-R; many of these investigators used a counterbalanced design (reviewed by Spitz, 1983). Recently, Ring (1985) reported an average rise of 11.8 points on the WAIS-R administered 3 to 6 years after the WISC-R; this finding is also very consistent with the present results.

An interesting question is whether the Wechsler adult scales are measuring too high or the children's scales too low, in terms of some outside criterion of general intellectual functioning. The best way to answer this question would be to compare the scores with various real-life indices, but a psychometric approach is to compare the scores with a third test, the Stanford-Binet Intelligence Scale, Form L-M (Terman & Merrill, 1973).

As Table 1 shows, the mean WAIS-R IQ obtained, on the average, 2.46 years after the Stanford-Binet is an astonishing 19.25 points higher, but the *N* is small and the size (though not the direction) of this difference is probably a sampling artifact. For only 7 retarded adolescents, Ring (1985) re-

ported that the average Stanford-Binet score was about the same as the WISC-R, but was 10 points lower than the WAIS-R, a finding more commensurate with the results given below for Stanford-Binet/WAIS comparisons.

One might argue that the higher score on the WAIS-R in comparison with the WISC, WISC-R, and Stanford-Binet was a true reflection of an actual rise in intelligence. Fortunately, in a sample of 23 subjects given both the WAIS and Stanford-Binet, 13 persons received the Stanford-Binet first and 10, the WAIS first. As Table 2 shows, even when the Stanford-Binet was given second, the mean IQ was substantially lower than it was on the WAIS. A Tests \times Order analysis of variance indicated that tests was a reliable effect, $F(1, 21) = 67.49$, $p < .0001$, but not order nor the interaction. For the combined data, the WAIS measured 12.43 IQ points higher than the Stanford-Binet, with an intertest correlation of .70, $p_s < .001$.

These results are consistent with reports

TABLE 2
INTERTEST IQ DATA FOR WAIS AND STANFORD-BINET (S-B) COUNTERBALANCED

Test order ^a	N	Age, 1st test ^b		IQ, 1st test		Retest interval ^b		IQ, 2nd test		Diff.	r
		Mean	SD	Mean	SD	Mean	SD	Mean	SD		
S-B, WAIS	13	17.41	3.37	48.54	10.78	1.90	1.01	58.85	11.78	10.30**	.82**
WAIS, S-B	10	18.76	2.32	56.20	10.32	1.70	1.07	41.00	8.34	15.20**	.65*

^a Listed in order of administration.

^b In years.

* $p < .05$. ** $p < .001$.

that the mean WAIS IQ of retarded groups is, on the average, 12 to 15 points higher than it is on the Stanford-Binet (L-M) (Cochran & Pedrini, 1969; Kroske, Fretwell, & Cupp, 1965). Moreover, this disparity increases from about 15 points in the 18- to 54-year age range to about 23 points in the 55- to 73-year age range (Fisher, Kilman, & Shotwell, 1961). Wechsler had overadjusted the IQs of older retarded persons because he believed that the normal age decline in intelligence would be more marked in mentally retarded than in non-retarded individuals. This belief was incorrect, according to Bensberg and Sloan (1950), because it was based on faulty sampling procedures.

Using different forms of the Stanford-Binet, several investigators reported acceptable IQ stability of retarded groups over retest intervals of 20 to 40 years (e.g., Earhart & Warren, 1964; Fisher, 1962; Holowinsky, 1962), but in Fisher and Zeaman's (1970) semi-longitudinal analysis, Stanford-Binet scores dropped precipitously from ages 2 to 16 years, then rose somewhat in mildly and moderately retarded groups. Fisher and Zeaman developed a transformation (*K* score) that increased the constancy of their data over the entire life span.

Higher scores on the Wechsler adult scales than on the Stanford-Binet might be attributed to the more verbal nature of the latter. In the present sample of 35 subjects tested on both the Stanford-Binet and either the WAIS or WAIS-R, however, mean Verbal IQ (60.40) was about the same as mean Performance IQ (60.09). Furthermore, Verbal IQs correlated .80 and Performance IQs correlated .56 with Stanford-Binet IQs, *ps* at least < .001. For this sample, then, the high verbal content of the Stanford-Binet can claim only limited responsibility for the test score disparities. It is likely that the primary sources of the disparities are the dissimilar and inadequate standardization samples used in the lower range of intelligence (Flynn, 1985).

Surely the various interest score disparities create an intolerable situation, but abandoning the IQ would seem too drastic a

solution. It would leave us with only a measure of adaptive behavior plus subjective impression, and it was the inadequacy of subjective impressions that provided the impetus for the development of intelligence tests in the first place. Flynn (1985) suggested that we adjust the mental retardation cutting line of various tests to be equivalent to some criterion test (e.g., WISC or WAIS) and that for a new or revised test, a large sample of retarded persons should be given both the new test and the older criterion test in order to equate the new score to the old cutting line. Valid test-equating requires a major, continuing effort (as with the Scholastic Achievement Tests) with large numbers of subjects and the use of sophisticated sampling and statistical techniques, and the question of which test to use as a criterion would be a difficult one. An alternative would be to return to the Stanford-Binet as the test of choice, but using (if research confirms its value) the Fisher-Zeaman *K* score. A new version of the Stanford-Binet is expected shortly, however, perhaps making the *K* score obsolete and, in any case, requiring new comparisons. Unless some standard method of adjustment can be agreed upon, the best that can presently be done is to alert users of these tests to the presence of test disparities. To do nothing about this situation would be to abrogate our responsibilities.

References

- Bensberg, G. J., & Sloan, W. (1950). Wechsler's concept of "normal deterioration" in older mental defectives. *Journal of Clinical Psychology*, 6, 359-362.
- Cochran, M. L., & Pedrini, D. T. (1969). The concurrent validity of the 1965 WRAT with adult retardates. *American Journal of Mental Deficiency*, 73, 654-656.
- Earhart, R. H., & Warren, S. A. (1986). Long term constancy of Binet IQ in retardation. *Training School Bulletin*, 61, 109-115.
- Fisher, G. M. (1962). A note on the validity of the Wechsler Adult Intelligence Scale for mental retardates. *Journal of Consulting Psychology*, 26, 391.
- Fisher, G. M., Kilman, B. A., & Shotwell, A. M. (1961). Comparability of intelligence quotients of mental defectives on the Wechsler Adult Intelligence Scale and the 1960 revision of the Stanford-Binet. *Journal of Consulting Psychology*, 25, 192-195.

- Fisher, M. A., & Zeaman, D. (1970). Growth and decline of retardate intelligence. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 4, pp. 151-191). New York: Academic.
- Flynn, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency, 90*, 236-244.
- Holowinsky, I. (1962). IQ constancy in a group of institutionalized mental defectives over a period of 3 decades. *Training School Bulletin, 59*, 15-17.
- Kroske, W. H., Fretwell, L. N., & Cupp, M. E. (1965). Comparison of the Kahn Intelligence Tests: Experimental Form, the Stanford-Binet and the WAIS for familial retardates. *Perceptual and Motor Skills, 21*, 428.
- Ring, B. C. (1985, January). Test scores for retarded young adults on the WAIS-R. *Oregon School Psychologists' Association Bulletin, 5-6*.
- Simon, C. L., & Clopton, J. R. (1984). Comparison of WAIS and WAIS-R scores of mildly and moderately mentally retarded adults. *American Journal of Mental Deficiency, 89*, 301-303.
- Spitz, H. H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and WAIS Full Scale IQs in a mentally retarded population. *Journal of Special Education, 17*, 69-80.
- Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: 1972 norms edition*. Boston: Houghton-Mifflin.
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corp.
- Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corp.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corp.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corp.