Perceptual and Motor Skills, 1972, 35, 527-534.

IQ CONSTANCY AND AGE[1]

HARRY KLONOFF[2]
University of British Columbia

Summary.—The WISC was given to 173 normal children, ages 5 to 13 yr. on 3 consecutive occasions at yearly intervals. Only the Full-scale IQ data were presented in this paper. Of the 18 between-year comparisons of mean Full-scale IQ, only 1 was significantly higher. Of the 9 between 2-year comparisons, only 2 were significantly higher. Correlation analysis showed a perfect relationship between magnitude of correlations and lapsed time as well as a distinctive trend regarding the increase of magnitude of correlation with increasing age. No sex differences in the patterning of IQ changes were noted (only 1 of the 54 comparisons was significant). There was no consistent pattern of change in IQ scores of the low-IQ subgroup compared with the high-IQ subgroup. Three factors are discussed as being important in longitudinal studies, namely, maturation, sex and individual differences in IQs of *S*s.

The concept of IQ implies constancy and confirmation of constancy can be achieved only by longitudinal studies. Thorndike (1940) provided one of the first systematic reviews of longitudinal studies on the stability of intelligence test performance by the same individual over a lapse of time. He reported the general trend of high positive correlations between test and retest in children of school age. Bayley (1933, 1949), on the other hand, in her study of normal infants and subsequent longitudinal study of children from age 1 mo. to 18 yr., found no essential relationship between infant intelligence scores and intelligence test scores at school age. Honzik, Macfarlane, and Allen (1948) arrived at the same conclusion. Stability of IQ scores has, however, been reported for school-age normal children (Honzik, Macfarlane, & Allen, 1948; Bradway, Thompson, & Craven, 1958). Holowinsky (1962) also reported IQ constancy in a group of mental defectives over a period of 3 decades.

Bayley (1955) pointed out that IQ scores may be altered by such conditions as emotional climate, cultural milieu, environmental deprivation and developmental changes in the nature and composition of the behaviors tested. Other factors might be listed, such as inter-test differences (infant schedules subsequently compared with Binet, Binet subsequently compared with Wechsler Bellevue); inter-test limitations (Binet at younger age level subsequently compared with Binet when *S*s are older than age 16); and inter-examiner differences. For example, Bradway, Thompson, and Craven (1958) reported that the average increase of 10 points on the Binet after an interval of 10 yr. was in all probability due to limitations of the Binet. Anastasi and Foley (1954) and Clarke and Clarke (1958) in reviewing a number of longitudinal studies concluded that

[1]This study was assisted under Grant 609-7-138, National Health Grants, Ottawa, Canada. Appreciation is expressed to the Vancouver General Hospital for their continued support.
[2]Division of Psychology, Department of Psychiatry, University of British Columbia, Vancouver 8, British Columbia.

increase in IQ in all probability occurs as a result of positive environmental influences. Kagan, Sontag, Baker, and Nelson (1958) pointed out the relationship between personality factors and IQ change. More recently, Rees and Palmer (1970) studied a number of demographic and family background factors that were related to level and change in mental test performance of children assessed at 6, 12, and 17 yr. of age. Short-term changes on the WISC have been reported by Irwin (1966) who re-administered the WISC to 2 age groups of children after an interval of 1 mo., and also by Quereshi (1968) who retested 5 age groups of children after an interval of 3 mo.

One of the limitations to date in longitudinal studies of IQ constancy or lability is the variety of intelligence tests employed. No longitudinal study has been uncovered in the review of the literature in which the WISC has been used exclusively. The specific contribution of the present study lies in the repeated administrations of the WISC to normal children at each age level from 5 through 15 yr.

The purposes of the present study were to determine: (1) the nature of IQ change between 3 administrations of the WISC, the administrations having been separated by 1-yr. intervals; (2) the relationships between IQ change during 3 consecutive test administrations and initial age of test administration; (3) sex differences in the patterning of IQ changes; and (4) the relationship between intelligence of *Ss* and IQ constancy.

Method

Subjects

Ss were 173 children (106 boys and 67 girls). The age-sex distribution was as follows: age 5—14 boys and 13 girls age 6—9 boys and 4 girls; age 7—11 boys and 12 girls; age 8—16 boys and 11 girls, age 9—12 boys and 4 girls; age 10—12 boys and 10 girls; age 11—12 boys and 5 girls; age 12—11 boys and 6 girls; and age 13—9 boys and 2 girls. The children were referred by 4 pediatricians and designated as normal according to the following criteria: normal school progress for those children of school age, no physical anomalies, no neurological impairment, and no history of emotional disturbance.

Test Administration

The Full-scale WISC was administered to the 173 children on 3 occasions, the second and third administrations separated by 1-yr. intervals. This paper reports the Full-scale IQ data at 9 age levels: age 5 (then 6 and 7); age 6 (then 7 and 8); age 7 (then 8 and 9); age 8 (then 9 and 10); age 9 (then 10 and 11); age 10 (then 11 and 12); age 11 (then 12 and 13); age 12 (then 13 and 14); and age 13 (then 14 and 15).

Results

Mean Full-scale IQ scores of the 173 children were above average for the 3 trials; means and standard deviations for the respective trials were 113.4 (9.3), 116.0 (9.1), and 118.6 (9.8). These results are to be expected in view of the referral basis of the children, and furthermore these results are consistent with those of other studies (Sontag, Baker, & Nelson, 1958).

The distributions of mean IQs and *SD*s by trial for each of the 9 age levels are reported in Table 1. The range of mean Full-scale IQs for Trial 1 varied from 109.7 to 117.7; for Trial 2 from 113.7 to 118.9; and for Trial 3 from 116.9 to 123.0. The absolute range of Full-scale IQs varied from 109.7 (for Trial 1) to 123.0 (for Trial 3). The range of *SD*s was restricted regardless of age and trial, i.e., from 6.6 to 12.3. Mean Full-scale IQ increased consistently between trials, irrespective of initial age of WISC administration.

TABLE 1

MEAN FULL-SCALE IQ AND *SD* FOR 3 TRIALS AT 9 AGE LEVELS

Age	*N*	*M*	*SD*	Age	*N*	*M*	*SD*
5	27	109.7	10.4	11		123.0	8.5
6		115.1	9.4	10	22	112.8	9.3
7		117.3	10.4	11		113.7	12.3
6	13	111.2	11.5	12		117.1	11.8
7		116.3	9.7	11	17	112.1	7.8
8		119.5	9.6	12		115.1	10.2
7	23	114.7	8.7	13		118.5	9.3
8		115.6	6.6	12	17	115.8	10.4
9		116.9	8.6	13		117.5	9.6
8	27	113.7	8.2	14		121.6	11.0
9		115.9	7.6	13	11	114.7	7.4
10		117.0	8.5	14		117.5	8.7
9	16	117.7	7.8	15		120.7	10.0
10		118.9	7.6				

The range of IQ difference scores was between Trials 1 and 2—.9 to 5.4, between Trials 2 and 3—1.1 to 4.0, and between Trials 1 and 3—2.2 to 8.3. There was a consistent trend regarding the magnitude of positive increment between trials for the respective age levels. Specifically, the magnitude of positive increment was always highest between Trials 1 and 3, followed by greater increment between Trials 1 and 2 compared with Trials 2 and 3 in 6 of 9 comparisons.

The next step is to determine whether the increase in mean Full-scale IQ between trials is significant. Of the 18 comparisons between successive years, only 1 was significantly higher ($t = 2.01$, $p < .05$ between Trials 1 and 2 at age 5). Of the 9 comparisons between baseline Full-scale IQ and mean Full-scale IQ recorded 2 yr. later, 2 comparisons were significantly higher ($t = 2.69$, $p < .01$ between Trials 1 and 3 at age 5, and $t = 2.16$, $p < .05$ between Trials 1 and 3 at age 11).

Table 2 presents the correlations between trials for the 9 age levels. All of the correlations were significant ($p < .01$). There was a perfect relationship between magnitude of correlations and lapsed time (subsequent reexamination) in that the highest correlations were consistently found between Trials 1 and 2, followed by Trial 2 compared with Trial 3, and in turn followed by between Trial 1 compared with Trial 3. There was also a distinctive trend regarding the increase of magnitude of correlation with increasing age.

TABLE 2

CORRELATIONS BETWEEN TRIALS AT 9 AGE LEVELS

Age (Yr.)	N	Trials 1 & 2	Trials 2 & 3	Trials 1 & 3
5, 6, 7	27	76	75	72
6, 7, 8	13	77	76	73
7, 8, 9	23	79	78	74
8, 9, 10	27	82	81	76
9, 10, 11	16	84	82	75
10, 11, 12	22	84	80	74
11, 12, 13	17	85	80	77
12, 13, 14	17	89	86	78
13, 14, 15	11	92	89	87

Now turning to an examination of sex differences in the patterning of IQ changes, of the 54 t tests, only one was significant ($t = 2.53$, $p < .05$ between Trials 1 and 3 for the age 5 boys).

Table 3 summarizes the absolute change in Full-scale IQ scores of the 173 children between Trials 1 and 2, Trials 2 and 3, and Trials 1 and 3, regardless of magnitude of change as well as direction of change. The highest proportion of children, regardless of trial comparisons, changed 10 IQ points or less and the

TABLE 3

ABSOLUTE CHANGE AND DIRECTION OF CHANGE IN MEAN FULL-SCALE IQ SCORES

IQ Change	Trials 1 and 2			Trials 2 and 3			Trials 1 and 3		
	% Absolute	% +	% −	% Absolute	% +	% −	% Absolute	% +	% −
<5	63.01	(36.42	26.59)	60 70	(42.20	18.50)	44.51	(29.48	15.03)
6-10	23.12	(16.76	6.36)	27.17	(20.81	6.36)	35.84	(29.48	6.36)
11-15	11.56	(9.83	1.73)	12.13	(10.40	1.73)	11.56	(10.98	0.58)
16-30	2 31	(2.31	0)	0			8.09	(8.09	0)

distribution of change was as follows: 86.13% between Trials 1 and 2; 87.87% between Trials 2 and 3; and 80.35% between Trials 1 and 3. More than 16 IQ points change occurred relatively infrequently and as follows: 2.31% of the children between Trials 1 and 2; none of the children between Trials 2 and 3; and 8.09% of the children between Trials 1 and 3. Absolute change in terms of lowered IQ scores of 6 to 15 points occurred infrequently, i.e., 8.09% between

Trials 1 and 2, 8.09% between Trials 2 and 3, and 6.94% between Trials 1 and 3.

Still another variable to be considered is the effect of IQ of the children in the cohort on the stability of IQ scores during repeated examinations. The 173 children were accordingly subdivided as follows: low-IQ subgroup (IQ scores between 90 and 109) of 63 children; high-IQ subgroup (IQ scores of 110 or higher) of 110 children. IQs and *SD*s for the low-IQ subgroup were as follows: Trial 1—104.1 (5.5), Trial 2—108.7 (8.2), and Trial 3—111.4 (8.4). Scores for the high-IQ subgroup were as follows: Trial 1—118.8 (6.2), Trial 2—120.1 (6.6), and Trial 3—122.7 (8.0).

TABLE 4

ANALYSIS OF VARIANCE AND *t* TESTS FOR LOW- AND HIGH-IQ SUBGROUPS

Term	*F*	Trials	Low-IQ	High-IQ
Subgroups	39.78†	1 & 2	3.73†	1.57
Trials	24.66†	2 & 3	1.84	2.63*
Subgroups × Trials	5.04†	1 & 3	5.81†	4.11†

$^*p < .01$. $^†p < .001$.

Inspection of Table 4 indicates the following: IQ between subgroups is significantly different, IQ increases significantly between trials, and the increases in IQ between trials are significantly higher for the low- compared with the high-IQ subgroups. Consistently significant increases between trials were, however, found only for the longer interval between retesting (Trials 1 and 3). With the shorter intervals, IQ scores increased significantly between Trials 1 and 2 for the low-IQ subgroup and between Trials 2 and 3 for the high-IQ subgroup.

DISCUSSION

Concern about constancy of IQ arises from the possibility of predicting intelligence from a measure made at a given time. Estimates of constancy of IQ are, however, contaminated by practice effect. One might accordingly attempt to unravel those errors of measurement due to practice effect, generally obtained by comparing test-retest performance over a short period of time, from spurious errors due to fluctuations of mental abilities. Two studies have reported short-term changes in WISC Full-Scale IQ. Irwin (1966) reported test-retest coefficients of .94 after an interval of 1 mo. for 6-yr.-old children and .98 for 11-yr.-old children. Quereshi (1968), using 5 age groups and an interval of 3 mo., reported correlations between .72 and .92, and increases in IQ that ranged between a minimum of 4.3 to a maximum of 8.1 points. These findings may now be contrasted with the results of the present study, which deals with longer-term changes, e.g., retest 2 yr. later. The range of correlations was between .72 and .87, whereas differences in mean IQs ranged from 2.2 to 8.3. The striking congruence between the findings of the present study and those reported by

Quereshi (1968) would lead one to hypothesize that the instability in IQ found in the current study is related to practice effect rather than spurious fluctuations in mental ability.

The present study as previous longitudinal ones fails to find constancy in IQ over lapsed time, in this instance 1 and subsequently 2 yr. after the initial examination. If, on the other hand, consistency in IQ is substituted for constancy, and consistency is defined in terms of scores that do not differ significantly over lapsed time, then the present study as previous ones shows positive findings.

Indices of consistency or stability might include the following: increment of IQ scores between trials as measured by differences between mean IQ scores; magnitude and direction of change as measured by amount of change in IQ points; and the relationship between repeated test administrations as measured by correlation.

The present study satisfied the conditions for the initial index of consistency in that IQ scores were not significantly higher between successive trials, with the exception of the youngest age group. Conditions for the second index of consistency were also satisfied in that 12% of *S*s changed more than 10 IQ points and only 1% of *S*s changed more than 16 points between trials separated by 1 yr. (based on average change for combined Trials 1 and 2, and Trials 2 and 3). Rees and Palmer (1970), in comparing changes between trials separated by 5 yr. (ages 12 to 17), reported that 30% of their *S*s changed 10 IQ points or more and 10.5% of their *S*s changed 16 points or more. Comparison between these studies should take into account the difference in lapsed time between trials.

Conditions for the third index of consistency were also satisfied in that all correlations between trials proved to be uniformly high and significant. The range of correlations from .72 to .92, with increase of magnitude in correlation being observed for the older compared to the younger groups, is consistent with the findings of other studies. Honzik, Macfarlane, and Allen (1948) examined children over a 2-yr. period and reported correlations which range from .80 to .92. Sontag, Baker, and Nelson (1958) reported comparable inter-age Binet correlations derived from the Fels study, i.e., from .79 to .92, as well as comparable Binet correlations derived from the Brush study, i.e., from .69 to .93. Rees and Palmer (1970) reported inter-correlations of .81, .81, and .75, derived from Binet tests administered at 6, 12, and 17 yr. Two generalizations are indicated: the first, correlations increase with the inclusion of older children in a study; and the second, correlations increase as the interval between testing is shortened.

Rees and Palmer (1970) stress the importance of considering 4 parameters —age, sex, level of initial performance, and the specificity of the measuring used in studies dealing with change in mental test performance. Data have also been presented in the present study about a number of factors that should be taken into account in establishing and evaluating longitudinal studies of constancy of IQ.

The first factor is maturation. Bayley (1933, 1949, 1955) has demonstrated the absence of a relationship between infant intelligence test scores and intelligence test scores of school-age children. The present study also found that the youngest age group, namely 5, exhibited the greatest instability of IQ scores. It might accordingly be concluded that if the WISC is used as the measure of intelligence in longitudinal studies of IQ constancy, age 6 should be used as the baseline.

Sex is another factor that might be considered relevant in proposed longitudinal studies of IQ constancy. The findings in the present study regarding differences between boys compared to girls were essentially negative, with the exception of the youngest age group. Other studies, on the other hand, have found differences in IQ gain for boys compared to girls, the higher gain having been noted for the boys (Sontag, Baker, & Nelson, 1958).

The third factor is one which does not appear to have received sufficient attention in the literature, namely, the role of individual differences in IQ of *S*s selected in a longitudinal study of constancy of IQ. The present study failed to find a consistently different pattern of change in IQ scores of the low-IQ subgroup and the high-IQ subgroup. On the other hand, Ebert and Simmons (1943) as well as Sontag, Baker, and Nelson (1958) reported that *S*s with higher IQs showed higher gains in IQ between successive administrations of the Binet. The lack of congruence between prior studies and the present one regarding IQ change in relation to IQ level might be due to differences between the Binet and the WISC or in the nature of the samples of children compared. It still seems, however, that the basic intelligence of *S*s should be taken into account in evaluating studies of IQ constancy.

REFERENCES

ANASTASI, A., & FOLEY, J. *Differential psychology.* New York: Macmillan, 1954.

BAYLEY, N. Mental growth during the first three years: a developmental study of sixty-one children by repeated tests. *Genet. psychol. Monogr.*, 1933, 14, 1-92.

BAYLEY, N. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, 75, 165-196.

BAYLEY, N. On the growth of intelligence. *Amer. Psychologist*, 1955, 10, 805-818.

BRADWAY, K., THOMPSON, C., & CRAVEN, R. Preschool I.Q.'s after 25 years. *J. educ. Psychol.*, 1958, 49, 278-281.

CLARKE, A., & CLARKE, A. D. *The measurement of intelligence, its validity and reliability.* Glencoe, Ill.: Free Press, 1958.

EBERT, E., & SIMMONS, K. The British Foundation Study of Child Growth and Development. *Monogr. Soc. Res. Child Develpm.*, 1943, 8, Whole No. 35.

HOLOWINSKY, I. I.Q. constancy in a group of institutionalized mental defectives over a period of III decades. *Train. Sch. Bull.*, 1962, 59, 15-17.

HONZIK, M. P., MACFARLANE, J. P., & ALLEN, L. The stability of mental test performance between 2 and 18 years. *J. exp. Educ.*, 1948, 17, 309-324.

IRWIN, D. O. Reliability of the Wechsler Intelligence Scale for Children. *J. educ. Measmt*, 1966, 3, 287-290.

KAGAN, J., SONTAG, L., BAKER, C., & NELSON, V. Personality and I.Q. change. *J. abnorm. soc. Psychol.*, 1958, 56, 261-266.

QUERESHI, M. Y. Practice effects on the WISC subtest scores and IQ estimates. *J. clin. Psychol.*, 1968, 24, 80-85.

REES, A. H., & PALMER, F. H. Factors related to change in mental test performance. *Develop. psychol. Monogr.*, 1970, 3, 1-56.

SONTAG, L., BAKER, C., & NELSON, V. Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, 1958, 23, 1-85.

THORNDIKE, R. L. Constancy of the I.Q. *Psychol. Bull.*, 1940, 37, 167-186.

Accepted July 5, 1972.