# Technical Papers

## Different I.Q.'s for the Same Individual Associated with Different Intelligence Tests

Ralph Mason Dreger
*Department of Psychology, Florida State University, Tallahassee*

The argument over the "constancy of the I.Q." in respect to the Stanford-Binet tests, the most well-known intelligence scales, seems for the time being fairly well settled (*1*). For any one individual, the I.Q. ratio may vary widely if personal and environmental factors are changed considerably over a period of time (*2*); for the group, however, there seems to be a constancy, i.e., most individuals will not vary more than about 5 I.Q. points from one time to another.

A further question arises, however, in the light of clinical experience. What about the constancy of the same individual's I.Q. as reported on different tests at approximately the same time? Clinicians especially (*3*) have pointed out that one must report not only the ratio computed but also on what test the I.Q. was obtained, in order for a proper evaluation of the person's intelligence to be made. Even with respect to one composite test, the Wechsler Intelligence Scale for Children, Seashore (*4*) warns that clinicians should be cautious about attaching unusual meaning to differences in Verbal and Performance I.Q.'s derived from the test.

The present investigation gives a limited answer to the question concerning different I.Q.'s from different tests. Ten graduate students in psychology were trained by the writer to give individual tests; then each one tested one subject with 4 intelligence tests. Table 1 discloses the fact that selection of subjects was from above average intelligence levels for the most part; chronological ages ranged evenly from 5 years and 8 months (5–8) to 13 years and 10 months (13–10). In one case only 3 tests were given. The design was intended to duplicate the actual clinical situation where one examiner gives several tests to the same person.

TABLE 1

I.Q.'s of Ten Children on Four Intelligence Tests Each

| S | CA | S–B* | | Good. | Kuhl. | Arth. I | Arth. II | Arth. Leit. |
|---|---|---|---|---|---|---|---|---|
| | | L | M | | | | | |
| 1 | 5– 8 | | | 154 | | | | |
| | 5–10 | 137 | | | | 147 | 139 | |
| 2 | 6– 2 | 114 | | 89 | | 111 | | 111 |
| 3 | 8– 1 | 153 | | 123 | | 101 | 102 | |
| 4 | 8– 2 | 96 | | | | | | |
| | 8– 5 | | | 101 | | 116 | 101 | |
| 5 | 8– 4 | | | 119 | | | | |
| | 8– 5 | | 103 | | | 115 | 107 | |
| 6 | 8–11 | 100 | | 117 | 95 | | 108 | |
| 7 | 10– 0 | 145 | | 130 | | | 120 | 121 |
| 8 | 10– 1 | 101 | | | | | | |
| | 10– 3 | | | 102 | | 115 | | |
| 9 | 12– 3 | 110 | | | | | | |
| | 12– 5 | | | 105 | 92 | | 111 | |
| 10 | 13– 7 | 119 | | | | | | |
| | 13–10 | | | 102 | 112 | | 122 | |

* Abbreviations refer to intelligence scales described in text.

TABLE 2

Coefficients of Concordance and Rank Correlations Among Several Intelligence Tests*

| Tests | $W$† |
|---|---|
| Stanford-Binet, Goodenough, Arthur I | .496 |
| Stanford-Binet, Goodenough, Arthur I, Arthur II | .525 |
| | $\tau$ |
| Stanford-Binet & Goodenough | .360‡ |
| Stanford-Binet & Arthur I | −.414 |
| Stanford-Binet & Arthur II | .354 |
| Goodenough & Arthur I | .138 |
| Goodenough & Arthur II | .214 |
| Arthur I & Arthur II | .333 |

* Probability levels are read from Kendall (*6*).
† Not significant in either case. $W$ ranges only from 0 to 1.
‡ $P = 0.036$. $\tau$ ranges from −1 to 1. All significance tests are one-tailed.

To allow for examiners' confidence in their own skill and to increase thereby examining validity, examiners were given the following tests from which to choose: Stanford-Binet, 1937 revision, Forms L and M (one of these was required in all cases as a baseline test); Goodenough Draw-a-Man Test; Kuhlmann Tests of Mental Development; Arthur Point Scale of Performance Tests, Forms I and II; and the Arthur Adaptation of the Leiter International Performance Scale. Not more than 3 months elapsed between the administration of the first test and the fourth.

Table 1 shows that I.Q.'s derived from the same individual on different tests do differ from one another. This conclusion is supported by the fact that one of the correlations shown in Table 2 is negative and only one of the remaining positive correlations is statistically significant.

An analysis of variance and Bartlett's test for homogeneity of variance (*5*) indicate that in the case of the entire group of children there are probably no significant differences between the tests. Differences appear to be between tests on the same child.

On the assumption that the tests which yielded the highest and lowest I.Q.'s for each subject could have been by chance the two tests administered if only two were administered in any clinic, the extremes of I.Q.'s for each individual were entered in two columns as his high and low scores. Fisher's test for difference

scores (7) resulted in a $t$-ratio beyond the 0.001 level. If this procedure is not too unsound, the inference is that the highest and lowest I.Q. scores made by the same children are significantly different from each other.

Although the present experiment confounds the reliability of the examiners' test-administration with the reliability of the subjects' I.Q. scores, it is exactly from this kind of confounded situation that we are forced to draw practical conclusions about I.Q. test scores obtained in our clinics today. Therefore, it can be concluded that whereas group means on different tests of intelligence may not differ except by chance from one another, individual's I.Q.'s may differ widely and significantly from one another on different tests.


References

1. ENGLISH, H. B., and KILLIAN, C. D. *J. Consult. Psychol.*, **3**, 30 (1939).
2. SKODAK, M., and SKEELS, H. M. *J. Genet. Psychol.*, **75**, 85 (1949).
3. WATSON, R. I. *The Clinical Method in Psychology*, p. 117. New York: Harper, 1951.
4. SEASHORE, H. G. *J. Consult. Psychol.*, **15**, 62 (1951).
5. EDWARDS, A. L. *Experimental Design in Psychological Research*, p. 197. New York: Rinehart. 1950.
6. KENDALL, M. G. *Rank Correlation Methods*, p. 150. London: Charles Griffin. 1948.
7. FISHER, R. A. *Statistical Methods for Research Workers*, p. 121. New York: Hafner. 1950.



Manuscript received August 10, 1953.


## An Improved Method for the Determination of Blood Volume Using Radioactive Iodinated Human Serum Albumen[1]


John R. Erickson, James B. McCormick, and Lindon Seed

*Isotope Laboratory, Augustana Hospital, Chicago, Illinois*


Improving techniques for the determination of blood volume have moved this procedure from that of an investigative tool to one of direct clinical value in the management of a wide variety of clinical problems. The earliest studies involved exsanguination of experimental animals. The introduction of essentially nondiffusible dyes, such as vital red and Evans blue, and the use of carbon monoxide permitted determination of blood volumes on living patients. Technical complexity and the inability to apply these methods repeatedly at short intervals in a given patient limits their clinical usefulness.

Tagging erythrocytes with $P^{32}$, $K^{42}$, and $Fe^{58}$ opened a new approach which, however, was still time-consuming.

In 1951, Aust *et al.* (*1*) devised a technique using 50–100 μc of $I^{131}$ tagged human albumen.

The sensitivity of the recently available well-type scintillation counter[2] is such that 0.001 μc of $I^{131}$ gives 895 ct/min.


[1] Supported by funds from the Nelson M. Percy Research Foundation, Augustana Hospital, Chicago, Ill.


[2] Manufactured by Nancy Woods, Chicago, Ill.

$I^{131}$-tagged human albumen as received from the supplier is diluted with sterile 0.85% saline solution so that 1 ml contains 3 μc. It is essential that sterile technique be observed in making the dilutions and that both the stock and diluted reagent be refrigerated.

1. A volume of the diluted stock solution calculated to contain approximately 3 μc of $I^{131}$ is accurately drawn up to the mark in a sterile syringe, and transferred quantitatively to an oxalate tube.
2. Using the same needle and syringe (to obviate error in standardization) an identical volume (drawn to the same mark as under 1) is injected into the subject's antecubital vein.
3. After 10 min (15–20 if patient is in shock) 10 cc of blood is withdrawn from the opposite antecubital vein and placed in an oxalate tube.
4. The control volume is diluted to 1000 ml in a volumetric flask with distilled water. This dilution gives approximately 10,000 ct/min/5 ml.
5. Background is determined for two 10-ml graduated test tubes. Average 200 ct/min.
6. Five milliliters diluted control solution is pipetted into one of the 10-ml tubes, and 5 ml of the oxalated blood (well mixed) is pipetted into the second 10-ml tube.
7. Each sample is counted in the well counter for 1 min.
8. Calculation:

$$\text{Total blood volume} = \frac{(\text{control count} - \text{background}) \times 1000}{\text{test count} - \text{background}}$$

9. Centrifuge blood at 2500 rpm 20 min and hematocrit determined.
10. Procedure for repeated determinations of blood volumes: a control blood specimen is drawn before the injection of the second dose of tagged albumen to determine residual radioactivity. Then repeat entire procedure.

$$\text{Total blood volume} = \frac{(\text{control count} - \text{background}) \times 1000}{(\text{blood \#2} - \text{background}) - (\text{blood \#1} - \text{background})}$$

The total time involved in this procedure is 45 min. The total dose of radiation is 3 μc/determination.

The only serious disadvantage to this technique is the original cost of the equipment. Its advantages are speed, complete safety, and reproducibility (Table 1).

TABLE 1

BLEEDING DUODENAL ULCER WITH NO FLUID INTAKE DURING PERIOD OF STUDY

| Date | Time | Hematocrit | T V, cc | TBV, cc/kg | RCV, cc | RCV, cc/kg | PV, cc | PV, cc/kg |
|---|---|---|---|---|---|---|---|---|
| 5/14 | 1:30 P.M. | 24 | 3120 | 41.6 | 748 | 10 | 2372 | 31.6 |
| 5/14 | 4:30 P.M. | 23 | 3250 | 43.3 | 747 | 10 | 2502 | 33.5 |
| 5/15 | 9:30 A.M. | 19 | 3870 | 51.6 | 735 | 9.8 | 3145 | 41.8 |