

VARIABILITY IN THE RESULTS OF
INTELLIGENCE-TESTS

D. D. V. STUART, JR., M.D.

Assistant Psychiatrist, Johns Hopkins Dispensary

BALTIMORE

Although measures for the detection of mental deficiency, and the estimation of its degree when present, are being rather widely discussed at the present time, there is one very important factor which has seldom been emphasized, but which must be taken into consideration in forming conclusions based on the use of intelligence-tests. I refer to the possibility of wide variations in the results of such tests when applied to the same individual by different examiners, if the examiners do not take into account certain features which I shall state later.

In this paper I shall deal only with the Binet-Simon test-scale, as adapted by the late Dr. E. B. Huey for use in the Johns Hopkins Dispensary; but as this scale is the basis of all the commoner forms of intelligence-tests, it would seem that the possibilities for error which must be guarded against in its employment, must likewise be considered in connection with the other varieties.

As part of a general examination, and as a means of gaging *approximately* the extent of mental development in any given case, the Binet-Simon tests are of unquestionable value. Unfortunately, though, there is a wide-spread tendency among both physicians and educators to regard these tests as yielding estimates of the mental age of patients just as accurate from the scientific point of view as are the results of quantitative chemical analyses, and to disregard factors which, if given due weight, might alter materially their deductions.

Experience has shown that the Binet-Simon tests cannot be considered to be exact in the sense that chemical tests are exact; and it is more than doubtful if intelligence-tests can ever be made so. Many things, of which I shall mention those that seem to me the most important, have to be taken into consideration in the application of an intelligence-test scale. The possibility of fatigue on the part of the subject, and to a lesser degree, of the examiner, is one of these. The slowing-up of mental processes during exhaustion is a recognized fact, yet I have seen a physician put a thoroughly tired-out child through the tedious ordeal of a Binet test. The subject's home surroundings and the training he has received are others; for, though the scale is intended to be a measure of native intelligence and not of education, environmental conditions play a large part in an individual's ability to answer correctly certain of the tests, especially for the later ages. Again, and perhaps most important of all, the personal equation as regards both subject and examiner must be considered if these tests are to be properly used. For example, a timid or sulky child and an examiner with a brusque harsh manner, make a bad combination. Under such circumstances, the subject is apt to grow worried or bewildered, and to fail on questions which, under other conditions, he would pass without difficulty. This is of course especially true of those tests done on a time basis. On the other hand, a too sympathetic or inexperienced examiner will often unwittingly suggest the proper responses to tests of which the subject has not the slightest real

comprehension. Finally, a patient's general health may have considerable effect on his grading in an intelligence-test. It is surprising how often an apparently low-grade subject is found to be suffering from defective vision or hearing, or some other pathologic condition, the relief of which is followed by marked improvement in his mental status.

These factors are too frequently completely ignored, with the result that diagnoses of mental deficiency are made when no deficiency exists, and conversely, that parents are sometimes told that there is nothing the matter with a child who is in reality handicapped by an incurable constitutional defect. It is this point of variability in the results of the tests that I wish to illustrate by means of the appended table.

All the cases quoted are white males, pupils in the special class for defectives in one of the public schools of this city. In each instance, the first examination was made in the Johns Hopkins Dispensary; and on the result of this first test was based the recommendation which was the means of obtaining the child's admission to the special class. The second and third examinations were conducted at varying times after entry into the special class, the training given in which is largely manual in character. The physical age is that reported by the person accompanying the child at

VARIATIONS IN THE RESULTS OF INTELLIGENCE-TESTS,
CONDUCTED BY DIFFERENT EXAMINERS UNDER
DIFFERENT CONDITIONS

Case	Age, Years	First Examination Date	Mentality	Second Examination Date	Mentality	Third Examination Date	Mentality
1. J. M. ...	15	12/10/12	7+	1/23/13	7+	12/ 3/13	8+
2. L. K. ...	11½	1/ 3/13	7	1/15/13	5+	10/27/13	7
3. D. F. ...	12	1/15/13	9+	12/10/13	10	2/28/14	8
4. J. W. ...	10½	1/21/13	7+	12/ 3/13	9	4/20/14	10
5. E. D. ...	9	9/15/13	5+	12/10/13	5+	4/ 4/14	4
6. R. H. ...	13	9/29/13	10+	12/10/13	12+	4/ 4/14	10+
7. L. P. ...	9	11/ 1/13	5	12/10/13	6+	4/ 4/14	5
8. A. D. ...	14	12/15/13	10+	2/18/14	10+	4/18/14	9+
9. H. T. ...	12	12/31/13	7	2/18/14	7+	5/ 9/14	6+

the time of the first examination. In the grading, a plus-mark indicates that the subject received sufficient credits to pass at the mental age given, and any number of additional credits not exceeding four. Thus, "mentality 6+" shows that the subject responded correctly to enough tests to enable him to grade up to six years, plus one or more additional tests, up to and including four. The examiners were, with three exceptions, physicians, the exceptions being postgraduate students in psychology at Johns Hopkins University. In only a single case was the same individual tested more than once by any one examiner. This was Case 6, in which instance the first and third examinations were both conducted by me.

While the cases cited are too few in number to afford a basis for any general conclusions, it is nevertheless impossible to deny the significance of the results obtained. These show, as will be noted, variations of from one to three years in the estimated mentality of the individual subjects, the average being between one and two years. The final test, as compared with the first, shows apparent progress in two cases and apparent regression in four; while in three cases the first and last results coincide. In no case did all three tests agree. If each successive examination showed progressively higher mentality, the objec-

tion might be made that all the cases in the table happened to be subjects who had not reached their limit of development when first tested, and that the special-class training had improved their condition. The fact that the majority show apparent regression or remained stationary contradicts this, and the only conclusion possible is that the differences in the results were caused by the variability of the tests themselves.

If intelligence-tests, when conducted by examiners with a knowledge of psychology and psychiatry, show such great divergence in their results, the assumption seems justified that the variation would be even more marked in inexperienced hands. It would also seem that it is time to stop regarding such tests as an infallible method of determining a patient's degree of mentality, independently of other considerations.

935 St. Paul Street.

THE USE OF A NO-SOUND STROKE IN PERCUSSING OUT THE BOUNDARIES OF SUPERFICIAL DULNESS OF AIRLESS BODIES

HENRY LEE SMITH, M.D.

Instructor in Medicine, The Johns Hopkins University
BALTIMORE

In the recent editions of the more important books on physical diagnosis one sees at a glance that the strong percussion stroke is being gradually displaced by lighter percussion as a more accurate method of determining the deep dulness of air-free bodies.

Goldscheider¹ has been particularly active in this field of work. He uses the lightest percussion (*Schwellenwertperkussion*, "threshold of audibility") in ascertaining the deep heart-boundaries, liver edge, upper boundaries of lung and small areas of pulmonary infiltration.² The ordinary method of finger-finger percussion is employed except that the pleximeter finger is held in the position of Plesch³ or, preferably, his own method of *Griffelperkussion*,⁴ which consists in the use of a small glass rod as the pleximeter. The upper end is slightly curved, or covered with rubber, and is lightly percussed with the right middle finger; the lower end is placed in the intercostal spaces parallel to the margin of the borders to be delimited. He advises that the rod be held obliquely in order to give the *Schallstrahl eine gewisse Richtung*⁴—orthopercussion. He controlled the method by orthodiagraphic findings¹ and by experiments on the cadaver, claiming for it many advantages.⁴

Sahli vigorously opposes the method because "it does not depend on any generally applicable acoustic principle, and because it disregards the didactically important distinction between deep and superficial dulness."⁵ He says that "to appreciate superficial

boundaries it is a good general rule that we should percuss as lightly as possible, and a good criterion of the desired strength is to evoke practically no note over the dulled areas."⁶

On the other hand, Orlowski,⁷ among others, has gained satisfactory results by adopting Goldscheider's method of orthopercussion, particularly in mapping out the lower border of the liver. He modifies Goldscheider's method somewhat by using the right index finger as plexor and the left index-finger as pleximeter; but the latter is held in the position of Plesch.

Where fecal masses or gaseous distention of the large intestine interfere with accurate work a saline or enema is given. He claims correct findings in 86 per cent. of livers with soft edge and in 92 per cent. of those with hard edge.

Sahli⁵ tells us that "by light percussion the lower liver boundary is generally easy to determine." Cabot⁸ laconically remarks that "percussion of the lower edge of the liver is notoriously unreliable." Butler,⁹ who strongly upholds Goldscheider's method, admits that "it may be difficult to find the lower limit of the liver in the median line in front because of the thinness of the left lobe of the liver and the dulling effect of the muscular layers of the recti."

Owing to the inconstant results obtained by my students and myself in percussing out the areas of superficial heart and liver dulness, it seemed clear to me that the fault lay in the varying strength of the stroke, and that uniform results could be obtained only when each individual examiner percussed with a blow of equal force. To attain this end I devised some eight or ten years ago the following method, which I have since employed:

To determine the outline of the area of superficial cardiac dulness the patient is placed in the usual dorsal decubitus, and is directed to suspend respiration. The examiner's ear is held close to the patient's body. There should be absolute quiet in the room. A point is selected where the heart is known to be uncovered by lung. The pleximeter finger (left middle) is laid gently on this area, the other fingers remaining elevated. The second phalanx is now percussed with the right middle finger. A sound is at first produced by moderately strong percussion which is quickly graded down, by lessening the force of the blow to a *no-sound* or control stroke. This control stroke, which proves inaudible over that portion of the heart's surface uncovered by lung (area of superficial dulness), produces a faint sound over the area of the deep cardiac dulness, and a more decided one over the lung region just outside of it. The no-sound stroke having been obtained, one percusses with a blow of equal force from the area of deep dulness toward the area of superficial dulness. The borders of the latter are instantly recognized by the sudden disappearance of the faint sound obtained over the former. The outstretched pleximeter finger, resting gently in its entire length on the part percussed and parallel to the border to be determined, edges its way quickly in short lateral (pendulum-like) swings toward the objective border, and receives at the end of each swing two short control taps quickly delivered. The

1. Goldscheider: Ueber Herzperkussion, Deutsch. med. Wchnschr., 1905, xxxi, 333 and 382.

2. Goldscheider: Ueber abgestufte Lungenperkussion, Verhandl. d. Kong. f. inn. Med., 1907, xxvi, 284.

3. Plesch uses the right middle finger as plexor, and delivers the blow just above the flexed second phalangeal joint of the left middle finger, the first and second phalanges of which remain rigid and are held vertically, the finger-tip only coming in contact with the patient's body. Ueber ein Verbessertes Verfahren der Perkussion, München. med. Wchnschr., 1902, xlix, 620; Einiges über Perkussion, Deutsch. Arch. f. klin. Med., 1908, xciii, 201.

4. Goldscheider: Zur Schwellenwertperkussion des Herzens, Deutsch. med. Wchnschr., 1907, xxxiii, 1121.

5. Sahli, Hermann: Diagnostic Methods, Philadelphia, Saunders & Co., 1911, p. 216.

6. Sahli: *ibid.*, p. 210.

7. Orlowski: Contribution à l'exploration du foie: Au moyen de la percussion, Presse méd., 1912, xx, 643.

8. Cabot, R. H.: Physical Diagnosis, New York, Wood & Co., 1913, p. 363.

9. Butler: Diagnostics of Int. Med., New York, Appleton, 1913, p. 501.