Proceedings of the Iowa Academy of Science

Volume 54 | Annual Issue

Article 32

1947

An Evaluation of the Various Short Forms of the Wechsler-Bellevue Test

Bertram Cohen
State University of Iowa

Ira M. Steisel
State University of Iowa

Let us know how access to this document benefits you

Recommended Citation

An Evaluation of the Various Short Forms of the Wechsler-Bellevue Test

BERTRAM COHEN AND IRA M. STEISEL

Introduction

The need for a rapid and efficient method of estimating adult intellectual level on an individual examination basis has been felt with increasing urgency in the past few years. An instrument of this type should include, to as great an extent as possible, the statistical advantages of a well standardized individual test along with the advantage of administrative economy. A brief scale meeting these criteria should be useful as a clinical screening device. It should be of some service as a method of sifting out of a large group those individuals who should be given further examination. Such an instrument should also be useful for various placement problems where large groups must be dealt with, and as an aid in research where groups or individuals need to be matched for intelligence.

The advent of the Wechsler-Bellevue intelligence scale (7) has been welcomed by clinical psychologists both as a quantitative and as a qualitative method of appraising the intelligence of adults and older adolescents. This scale consists of ten separately scored subtests, each of which contains a series of individual items arranged roughly in order of difficulty. Any combination of subtests can be administered, scored, and after a simple prorating process the results can be made to yield a total weighted score. This score can be used to obtain a short form I.Q. from Wechsler's Full Scale tables in the same manner that the total weighted score derived from the complete scale is used. Furthermore, the validity of any short form may be tested by correlating I.Q.'s derived from it with I.Q.'s derived from the Full Scale.

Within the past four years, three different combinations of subtests have been offered to the psychological public as short form examinations. The first of these was the Rabin short form (5) consisting of three subtests, the Comprehension, Arithmetic, and Similarities tests. Rabin reports a correlation with the Full Scale of .80 for a group of 92 normals (nurses) ranging in age from 19 to 25, and in Full Scale I.Q. from 85 to 130. He also reports an r of .93 with a group of 200 psychiatric patients at the New Hampshire State Hospital. Most of these patients were schizophrenic and manic-depressive cases plus a considerable number diagnosed as psychoneurotic and "without psychosis." Ages of these patients ranged from 15 to 36 and I.Q.'s ranged from 39 to 122. The mean Full Scale I.Q. for this group was 90.71. The mean short form I.Q. was 89.03. The difference between the means is reported as statistically insignificant.

The results of other studies using Rabin's form have confirmed his findings. One of these, a study by Springer (6) tested the validity

221

of this short form on 100 men (naval personnel) between the ages of 20 and 24. He obtained a correlation of .92 with the Full Scale Wechsler. Full Scale I.Q.'s for these men ranged from 67 to 119 with a mean of 82.95. The Rabin I.Q.'s ranged from 61 to 124 with a mean of 81.20.

A study of Rabin's short scale by the senior author (1) at the Iowa Psychopathic Hospital yielded correlations with the Wechsler Full Scale ranging from .87 to .90 on four separate psychiatric groups (Schizophrenia, Psychoneurosis, Manic-Depressive, depressed, and Involutional Psychosis). An r of .88 was obtained for the total group of 172 cases. The total population ranged in Full Scale I.Q. from 68 to 129 with a mean of 100.88. Rabin I.Q.'s for this group ranged from 52 to 146 with a mean of 103.47. The difference, in this case, was statistically significant although small.

The second short form to appear in the literature comprises four subtests; the Comprehension, Similarities, Digit Span, and Block Design tests. This combination was suggested by Geil (2) who tested its validity against the Full Scale results of 250 unselected cases at the Medical Center for Federal Prisoners at Springfield, Missouri. These included physically handicapped, psychotic, and otherwise maladjusted individuals. Geil reports an r of .966 ± .003 with the Full Scale. This population ranged in Full Scale I.Q. from 40 to 133 with a mean of 90.30. The abbreviated scale I.Q.'s ranged from 37 to 137 with a mean of 90.66.

The third and last combination of subtests to be offered as a short form is that of Milton S. Gurvitz (3). This form includes two subtests; the Picture Arrangements and Digit Span tests. The population in this study was made up of 523 adult male prisoners. Two hundred and fifty of these cases were from the Medical Center for Federal Prisoners mentioned above, and the rest from the Psychological Clinic of the U. S. Penitentiary at Lewisburg, Pa. A correlation coefficient of .90 ± .01 was obtained for the relationship between the Full Scale and this short form. Since the population was overweighted with cases in the lower ranges, Gurvitz "normalized" the group by eliminating a number of borderline and defective cases and by adding or subtracting cases in the other levels. This latter group included 330 cases and the r for this group was .88 ± .03. (Gurvitz does not specify his method of choosing cases to be eliminated or added). The total population of 523 cases comprised individuals who were "selected solely on a routine basis." A great majority of them had been individually examined for some "special psychiatric or psychometric reason." The group ranged in age from 17 to 64 and in Full Scale I.Q. from 42 to 147.

With the exception of Rabin's group of nurses, (and perhaps, Springer's group of naval personnel) all of the studies mentioned include socialy and/or mentally deviant populations. In every case the populations are of average or below average mean intelligence with a noticable loading of the latter. None of the studies include

an attempt to estimate the test—retest reliability of any of the short forms.

These conditions leave open the following questions which the present study was designed to investigate:

(1)   How do the correlations hold up in a non-clinical and non-penal group of above average intelligence?

(2)   What is the reliability of the short forms as measured by test-retest correlation coefficients?

(3)   How do the various short forms compare with one another in terms of magnitude of correlation with the Full Scale, test-retest reliability, and discrepancy between the Full Scale I.Q.'s and the mean I.Q.'s as estimated by each of the short forms with an above-average population such as described above?

Subjects and Procedure

Thirty-four female students in the Elementary Psychology course at the State University of Iowa served as subjects. None of them had had any previous experience with any of the subtests or with the test as a whole. All tests were administered by one of the authors (I.S.). The age range of this group was 19-0 to 22-4 with a mean of 19-11 years. The Full Scale I.Q.'s ranged from 96 to 129 with a mean and standard deviation of 116.77 and 8.13 respectively. Half of the group was retested after an interval of two weeks and the other half after 11 weeks. (This is due to the fact that the testing had been done in relation to a different study.) In an attempt to keep motivation at an optimum and discussion of the test at a minimum, standard instructions were read at the two meetings.

After some of the retests had been given, it became apparent that several of the subjects had become acquainted with the answers to a few items which they had not known previously. As a result of this, the examiner asked the remaining subjects, after reading the retest instructions, whether they had in any way become familiar with the answers to the items since the previous meeting. Several subjects admitted that they had. However, the results for these subjects were not excluded from the study. This factor should be kept in mind in evaluating the retest reliabilities.

Pearson Product-Moment correlations were obtained for the relationship between the Full Scale I.Q.'s and those derived from each of the short forms. (Since the subtests which make up the short forms are included in the Full Scale, the correlations are expected to be inflated. This is of little concern for the present study inasmuch as our problem is simply one of comparing the relative efficiency of each of the abbreviated scales in estimating Full Scale I.Q.'s). Fisher's "t" for related samples* was applied to determine the significance of the mean of the differences between the individual I.Q.'s derived from each short form separately and their corresponding Full Scale I.Q.'s. Test-retest reliability was estimated

by computing Pearson Product-Moment correlations for each short form.

Results

Table I contains the results of the various statistical analyses.

TABLE I

STATISTICAL RESULTS FOR EACH SHORT FORM INCLUDING CORRELATIONS BETWEEN EACH SHORT FORM AND THE FULL SCALE, MEAN DIFFERENCES AND THEIR STATISTICAL SIGNIFICANCE, RANGES, AND TEST-RETEST CORRELATIONS (N = 34).

Author	Gurvitz	Geil Comp.,	Rabin Comp.,
Subtests used	Pict. Arr. and Dig.	Simil., Dig., and B.D.	Arith., and Sim.
Range of I.Q.'s	75 to 141**	104 to 140	94 to 142
Correlation with Full Scale I. Q.	.52 ± .13*	.87 ± .04	.78 ± .07
Mean Differences	6.30**	3.26	.65
S.E. of the Mean of the Differences	2.27	.79	1.55
t	2.78	4.13	.42
Level of confidence	1 percent	1 percent	N.S.
Range of retest I.Q.'s	86 to 144**	112 to 147	101 to 140
Correlation with retest I.Q.'s	.56 ± .12*	.53 ± .12	.72 ± .08

    *N = 32. See text.
    **N = 33. See text.

It can be seen that the correlations with the Full Scale I.Q.'s arrange themselves in the following order: Geil's short form is highest with an r of .87 ± .04, Rabin's next with an r of .78 ± .07, and that of Gurvitz is lowest with an r of .52 ± .13** This sequence might have been expected on *a priori* grounds, since a correlation between parts of a test and the entire test is a function of the number of parts (subtests in this case) employed. Therefore, we should expect that that short form which includes a greater number of subtests would correlate higher with the Full Scale than one which includes fewer subtests.

---

*For the formula see Lindquist (4, Pg. 59).
**In two cases the weighted scores obtained by prorating with this scale were beyond those in Wechsler's I.Q. tables, thereby necessitating their omission from the calculations. (N = 32).

The mean difference between the Full Scale I.Q.'s and the Gurvitz short form I.Q.'s is 6.30.* The "t" for this difference is 2.78 and is statistically significant beyond the 1 percent level of confidence. Geil's short form shows a mean difference of 3.26 I.Q. points from the Full Scale. The "t" is 4.13 which is also statistically significant beyond the 1 percent level of confidence. Rabin's short form shows a mean difference of .65 I.Q. points. The "t" of .42 is insignificant. It should be added that Geil's short form tends to overestimate the

Full Scale I.Q.'s, while that of Gurvitz and Rabin are in the reverse direction.

The test-retest correlation for Rabin's short form is the highest with an r of .72 ± .08. Gurvitz's and Geil's short forms are approximately equal in this respect, the r's being .56 ± .12** and .53 ± .12 respectively. The test-retest correlation for Geil's short form is surprisingly low in comparison with its relatively high correlation with the Full Scale as mentioned above. Inspection of the data indicates that this may be due to the inclusion of the Digit Span test in this short form. This subtest shows a greater amount of variability than any of the others in this particular group.

Our statistical results, therefore, indicate that Gurvitz's short form is inadequate when applied to a population such as ours. Of the other two forms evaluated, Rabin's seems to be the better, although its correlation with the Full Scale is somewhat lower than the

*One pair of scores were omitted for the reason stated in previous footnote.
**N = 32.

correlation for Geil's. Rabin's choice of subtests do not result in significantly different I.Q.'s from those obtained with the Full Scale and are more stable on retests.

All of the correlations obtained in the present study are lower than those reported by the original authors. This was to be expected inasmuch as the range of I.Q.'s for the present population is much narrower than those reported by the other authors. The group which is most comparable to ours is Rabin's group of nurses. The Full Scale I.Q.'s for this group range from 85 to 130, while ours range from 96 to 129. Our correlation of .78 between Rabin's short form and the Full Scale agrees very closely with the r of .80 that Rabin obtains for this group.

Aside from statistical considerations, the factor of economy of administration seems worthy of special evaluation in the choice of a short form. The amount of time necessary for administration and the materials required should be considered. As far as time is concerned, Gurvitz's short form is most rapidly administered (5-10 minutes), Rabin's next (15-20 minutes), and Geil's takes the most time (approximately 25 minutes). As to the materials necessary, Rabin's short form is the most efficient as it includes only verbal items all readily memorizable by an examiner. It should be mentioned, however, that while the other two forms include performance items requiring special materials, they are not bulky and are of no real hindrance. Also, the inclusion of a performance subtest might prove advantageous in balancing the verbal weighting in most short intelligence examinations.

Clinically, the authors feel that Geil's short form includes subtests which lend themselves readily to pattern analysis. Geil claims they are of special use in screening out for further psychological examination individuals with psychotic and organic thought disturbances. Rabin's short form is also of value in this respect, although it does not include the Block Design test which is included

by Geil and has proven extremely useful in yielding clinically significant information. Gurvitz's short form seems least fruitful in respect to the criterion of clinical usefulness.

In conclusion, there are three general factors to be considered in evaluating the relative adequacy of the various short forms of the Wechsler-Bellevue test. These are: (1) Statistical considerations, (2) considerations of economy of administration, and (3) clinical usefulness. In the present group of college students of high average intelligence, Rabin's short form holds up most consistently from the statistical standpoint. Furthermore, it can be administered rapidly and seems fairly adequate as a clinical screening device. If, however, the problem facing an examiner is one of screening out individuals with special mental disturbances, Geil's short form might be more profitably employed.*

BERTRAM COHEN,   IRA M. STEISEL,
DEPARTMENT OF PSYCHOLOGY,   DEPARTMENT OF PSYCHOLOGY,
THE STATE UNIVERSITY OF IOWA.   THE STATE UNIVERSITY OF IOWA.

References

1. Cohen, B. 1945. Validity of the short form of the Wechsler-Bellevue Intelligence Scale on four psychiatric groups. Unpublished Master's thesis, University of Iowa.
2. Geil, G. A. 1945. A clinically useful abbreviated Wechsler-Bellevue Scale *J. Psychol.*, 20, 101-108.
3. Gurvitz, M. S. 1945. An alternate short form of the Wechsler-Bellevue test, *Amer. J. Orthopsychiat.*, 15 No. 4, 727-732.
4. Lindquist, E. F. 1940. *Statistical analysis in educational research.* New York: Houghton Mifflin.
5. Rabin, A. I. 1943. A short form of the Wechsler-Bellevue test. *J. Appl. Psychol.*, 27, 320-324.
6. Springer, N. N. 1946. A short form of the Wechsler-Bellevue test as applied to naval personnel. *Amer. J. Orthopsychiat.*, 16 No. 2, 341-4.
7. Wechsler, D. 1945. *The Measurement of Adult Intelligence.* (3rd Ed.) Baltimore: Williams and Wilkins.

* After this paper went to press a similar study by Patterson was brought to the attention of the authors. His results are compatible with those presented here. See Patterson, C. H. 1946. A comparison of various "short forms" of the Wechsler-Bellevue scale, *J. Consult. Psychol.*, 10, 260-267.