THE STANDARDIZATION OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN[1]

HAROLD SEASHORE, ALEXANDER WESMAN AND JEROME DOPPELT
THE PSYCHOLOGICAL CORPORATION

THE *Wechsler Intelligence Scale for Children* has grown logically out of the *Wechsler-Bellevue Intelligence Scales* used with adolescents and adults [4]. In fact, most of the items in the *WISC* are from Form II of the earlier scales, the main additions being new items at the easier end of each test to permit examination of children as young as five years of age.

Even though the materials overlap, the *WISC* is a distinct test from the *Wechsler-Bellevue Scales* and is independently standardized. The scales overlap in usefulness since both scales can be used with adolescents. However, it is expected that the *WISC* will be preferred in testing adolescents up through the age of fifteen years.

This new Children's Scale (as it probably will come to be known in every-day clinical parlance) has been standardized with exceptional care over a five-year period of experimental tryouts, field testing, and statistical analysis. In this paper some of the principal research data are reported.

The *WISC* consists of twelve tests which, as in the Adult Scale, are divided into two subgroups identified as Verbal and Performance. In the standardization, there were six tests in each of the subgroups:

Verbal	*Performance*
General Information	Picture Completion
General Comprehension	Picture Arrangement
Arithmetic	Block Design
Similarities	Object Assembly
Vocabulary	Coding
(Digit Span)	(Mazes)

In the interest of shortening the time required for examination, the Scale is to be administered ordinarily on the basis of only ten tests. For various statistical and practical reasons, Digit Span is considered an alternate test in the Verbal series and Mazes an alternate in the Performance series. As a matter of fact, the reasons for including Coding or Mazes are about equally good except that the Mazes take a considerably longer time than Coding and will probably, therefore, not be preferred. The conditions for using the alternate tests are described in the manual.

The reader is referred to the manual and to Wechsler's earlier text for a more complete description and discussion of the tests.

THE STANDARDIZATION SAMPLE

Age of Children and Size of Sample. The *WISC* was standardized on a sample of 100 boys and 100 girls at each age from five through fifteen years. Each child was tested within one and one-half months of his mid-year; e.g., the five-year-olds were past 5 years, 4 months and 15 days but were not yet 5 years, 7 months and 15 days. The feebleminded cases were exceptions as an adequate sample could not be secured without permitting more variation; nearly all, however, were within two months of their mid-year. There were 1100 boys and 1100 girls in eleven age groups, a total of 2200 cases. Actually more cases were tested, but the final sample includes those who best satisfied the other sampling requirements described below. Only white children were examined.

Variables of Sampling. It was determined at the beginning that for each age (as closely as practicable) and for the total sample, the selected cases should meet certain sampling requirements based on U.S. Census Bureau data for 1940, with some adjustment for the recent shift of population toward the West.

[1]This report of the standardization is an expansion of technical sections in the test manual: David Wechsler, *Wechsler intelligence scale for children.* New York: Psychological Corporation, 1949. Pp. 113.

1. *Areas.* The states were divided into four geographic areas as defined in Table 1.

2. *Urban-Rural.* The urban-rural proportions of the total U. S. 1940 population were used as shown in Table 2.

3. *Parental Occupation.* The children's fathers were to be occupationally distributed similarly to all employed white males. The fourteen U. S. Census categories were reduced by combinations into nine, as shown in the footnote of Table 4. The quota for each geographic area was further defined in terms of Census reports on employment within each area.

Drawing the Sample. With these controls, tables of requirements were set up for examiners in each area. It was not expected that the occupational and urban-rural requirements would be exactly satisfied in each area, but that the over-all conditions would be met by the national sample.

Specific directions and worksheets for drawing a sample in a given school were provided so that cases would not be "volunteered" or "thrown in" at the whim of the examiner or school official.

Most of the 55 feebleminded cases were examined at the Illinois State School, Lincoln, Illinois; at Letchworth Village, New York; and at the Wayne County Training School, Michigan; a few selected cases from "special classes" of two public schools were included. The staff psychologists in the institutions aided in the selection of cases of the required ages who were rated as having IQ's under 70 and not below 50. Cases where postnatal disease or accident were considered causative of the deficiency were omitted. The number of feebleminded cases appearing in the regular school sampling was not determined. No examiner assigned to public schools reported that any case was officially labelled as feebleminded. In all, then, 2.5 per cent of the total number of cases in the standardization population is known to be feebleminded.

Analysis of the Obtained Sample. Tables 1, 2, 3, 4 and 5 present the sampling data on the 2200 cases finally included in the standardization group.

In Table 1 it is seen that the midwest sample (Area II) is slightly short of cases; this was deliberate in order to increase the western proportion in accord with wartime and postwar population shifts. All in all, the area sampling is eminently satisfactory.

TABLE 1
SAMPLE BY GEOGRAPHIC AREA

	Per Cent in U.S. Population	Wechsler %	Sample N
I New England and Middle Atlantic States	29.2	31.0	683
II North Central States	32.7	28.9	635
III South Atlantic and South Central States	26.8	26.5	583
IV Mountain and Pacific States	11.3	13.6	299
	100.0	100.0	2200

TABLE 2
SAMPLE BY URBAN-RURAL RESIDENCE

	Per Cent in U.S. Population	Wechsler %	Sample N
Urban	57.9	60.3	1327
Rural	42.1	37.2	818
Institutional*	—	2.5	55
	100.0	100.0	2200

*The 55 feebleminded cases from institutions were not reported as either rural or urban.

TABLE 3
PROPORTION OF URBAN AND RURAL CASES EXPECTED AND OBTAINED IN EACH AREA

AREA	URBAN			RURAL		
	Expected %	Wechsler %	Sample N	Expected %	Wechsler %	Sample N
I	38.4	34.5	458	16.2	23.4	191
II	32.9	32.9	437	32.5	21.6	177
III	17.2	17.5	232	40.0	42.9	351
IV	11.5	15.1	200	11.3	12.1	99
	100.0	100.0	1327	100.0	100.0	818

Table 2 reveals that in the national standardization an appropriate number of rural children was included. Table 3 presents the urban-rural data in a different analysis. This table shows what percentage of the total sample of urban cases was expected from each of the areas and what percentages were obtained. The rural sources are analyzed similarly. The only large differences are the relatively high rural contributions by Area I and a shortage in Area II.

The values in Table 3 are slightly distorted because the 55 feebleminded cases (2.5 per cent of all cases) were not allocated by residence.

The reader should also be reminded that the rural category includes all cases living on farms and in communities of less than 2500 population. "Bedroom villages" attached to large cities were generally avoided.

The statistical significance of the discrepancies between *expected* cases and *obtained* cases could be computed, but this seems inappropriate because the expected percentages themselves are quite imperfect criteria. Alloca-

TABLE 4

OCCUPATION OF FATHERS OF CHILDREN IN STANDARDIZATION SAMPLE

Occupational Groups*	Employed U. S. Males %	Wechsler Sample All Cases %	Wechsler Sample All Cases N	Wechsler Sample Boys %	Wechsler Sample Girls %
1	5.9	8.0	176	7.9	8.1
2	14.0	10.0	222	10.3	9.9
3	10.6	11.6	256	11.9	11.4
4	13.9	12.7	280	12.5	12.9
5	15.6	17.9	393	18.8	16.9
6	18.8	16.5	363	16.6	16.4
7	6.0	5.5	122	5.5	5.5
8	14.5	13.8	303	12.4	15.2
9	.7	1.4	30	1.5	1.2
(Feebleminded)	—	2.5	55	2.5	2.5
N			2200	1100	1100

*A consolidation of 14 Census groups, 1940:

1. (I and II) Professional and semiprofessional workers
2. (III) Farmers and farm managers
3. (IV) Proprietors, managers and officials
4. (V) Clerical, sales and kindred workers
5. (VI) Craftsmen, formen and kindred workers
6. (VII) Operatives and kindred workers
7. (VIII, IX and X) Domestic, protective and other service workers
8. (XI, XII and XIII) Farm laborers and foremen, and laborers
9. (XIV) Occupation not reported

TABLE 5

OCCUPATION OF FATHERS OF CHILDREN IN STANDARDIZATION SAMPLE

Occupational Groups	AREA I Expected %	AREA I Obtained* %	AREA II Expected %	AREA II Obtained %	AREA III Expected %	AREA III Obtained %	AREA IV Expected %	AREA IV Obtained %
1	7.0	8.6	5.3	6.0	4.7	10.6	6.8	7.0
2	3.6	3.7	17.2	12.9	22.9	16.5	10.0	7.7
3	11.1	11.2	10.0	10.3	9.8	13.2	12.0	14.4
4	16.6	16.9	13.1	9.3	11.7	10.3	14.0	17.7
5	18.1	18.2	15.4	22.8	12.7	16.1	16.0	13.7
6	23.3	19.6	18.1	18.7	15.8	13.4	16.4	14.4
7	7.7	7.9	5.8	4.1	5.1	3.4	8.8	8.7
8	11.9	12.0	14.4	14.5	16.6	14.9	15.6	16.4
9	.8	1.8	.7	1.5	.7	1.5	.4	—
N	640	649	720	614	590	583	250	299
(Feebleminded)		34		21		—		—

*Obtained percentages computed on N without feebleminded cases.

tions based on 1940 Census data are the best criteria the authors could set for control of the sampling. But we cannot readily know whether a discrepancy of, say, five percentage points, is due to obsolete Census data or to faulty sampling. For that reason one must be satisfied with reasonable approximations. There is reason to believe that there has been some trend toward urbanization since 1940, and that the five per cent shortage of rural cases is, therefore, an exaggeration as of 1947-1948.

Tables 4 and 5 describe the sample by occupation of the father. Again, reasonable agreement between the Census expectancy and the actual sample is evident. The greatest shortage is in Occupational Group 2, farmers and farm managers. If one mentally combines the percentages of Occupations 3 and 4, as seems logical from the descriptions, the expected percentage would be 24.5 and the obtained percentage 24.3. Similarly, if one combines Occupational Groups 5 and 6, the expected percentage is 34.4, which is identical to the obtained percentage. A slight error in the obtained percentages occurs because the 2.5 per cent of feebleminded cases were not allocated by parental occupation. In the last columns of Table 4, the analysis shows that the percentages of boys and girls with respect to parental occupation are very similar.

Table 5 presents the occupational sampling by area. The agreements are not as good here as for the national sample, but there are no gross miscarriages of sampling.

The field examiners used utmost care in ascertaining the father's occupation and were asked to write descriptions in considerable detail. The final classifications were made with the detailed Census descriptions at hand. Note that the field examiners were able to secure rather good samples in category 8, farm laborers and other laborers; this is usually a difficult task. Similarly, the excess in Occupation 1 (which is usually a category that one finds difficult to keep small enough) is considerable only for Area III.

The best available base for the selection of this sample was the occupational distribution as reported in the United States Census. However, one should not take percentages derived from those data as being absolute criteria against which to select cases. There obviously have been shifts in occupational percentages because of the war, and also shifts in occupational groups from area to area. Certain parts of the country have become more industrializ-

TABLE 6

CORRELATIONS OF EACH TEST WITH THE VERBAL, PERFORMANCE AND FULL SCALE SCORES
100 BOYS AND 100 GIRLS AT EACH AGE

	VERBAL*			PERFORMANCE†			FULL SCALE‡		
Age	7 1/2	10 1/2	13 1/2	7 1/2	10 1/2	13 1/2	7 1/2	10 1/2	13 1/2
Verbal Score	—	—	—	.60	.68	.56	—	—	—
Information	.64	.82	.80	.44	.59	.51	.59	.77	.73
Comprehension	.49	.70	.68	.46	.56	.37	.54	.69	.58
Arithmetic	.55	.70	.59	.46	.57	.38	.57	.69	.55
Similarities	.55	.72	.74	.41	.48	.52	.53	.65	.71
Vocabulary	.66	.82	.75	.47	.68	.51	.63	.83	.70
Digit Span	.48	.50	.44	.45	.40	.29	.52	.50	.42
Performance Score	.60	.68	.56	—	—	—	—	—	—
Picture Completion	.42	.45	.38	.34	.48	.55	.43	.51	.51
Picture Arrangement	.51	.58	.43	.51	.53	.51	.58	.62	.53
Block Design	.42	.55	.50	.53	.66	.65	.52	.64	.64
Object Assembly	.38	.38	.31	.59	.52	.68	.52	.47	.52
Coding‖	.31	.42	.42	.32	.35	.42	.35	.43	.48
Mazes	.31	.43	.40	.51	.55	.39	.46	.53	.44

*Sum of 5 tests, Digit Span omitted.
†Sum of 5 tests, Mazes omitted.
‡Sum of 10 tests, Digit Span and Mazes omitted.
‖Coding A at age 7 1/2; Coding B at ages 10 1/2, 13 1/2.

ed. There has also been a considerable shift in the west coast total population.

The sampling requirements were defined in terms of the occupations of the fathers, and of urban-rural residence of workers, but this does not mean that children across the country are distributed in the same percentages. It is generally known that rural families, laboring families, and perhaps southern families have more children than urban, upper middle class, and northern families. The data for making adjustment in quotas are complicated and incomplete. Modification of the percentages in each occupational category, and in the urban and rural categories to account for differential fecundity did not seem feasible.

TABLE 7

MEDIAN CORRELATION COEFFICIENTS BETWEEN TESTS AND VERBAL, PERFORMANCE AND FULL SCALE SCORES

	Verbal Score	Performance Score	Full Scale Score
Median r of Verbal tests	.67	.46	.61
Median r of Performance tests	.42	.51	.52

INTERCORRELATIONS OF THE TESTS WITH VERBAL, PERFORMANCE AND FULL SCALE SCORES

Table 6 is constructed to show the relationship of each test with the three special scores. When a test is correlated with the composite of which it is also a contributing member (e.g., Vocabulary with the Verbal Score) a correction [2] for spuriousness has been applied. The data are presented for three representative ages. In the manual more complete tables are printed showing the intercorrelations of the tests themselves. Table 7 may be helpful in comprehending the mass of coefficients in Table 6.

The Verbal tests correlate more highly with the Verbal Score than with the Performance Score, and likewise the Performance tests correlate more highly with the Performance Score than with the Verbal Score. This is as one would expect. It does appear that the Verbal tests are somewhat more homogeneous as to abilities measured.

The correlations between the Verbal Score and the Performance Score are sufficiently high (.60, .68, .56 for the three ages) to indicate considerable common variance, yet are low enough to suggest that the abilities included in *V* and *P* cannot be readily inferred from each other. Both classes of abilities need to be tapped in an over-all appraisal of abilities.

These data also indicate that the Digit Span is the least like the other Verbal tests; for this reason it was made an alternate. The Coding and Maze tests are about equally eligible to remain in the Performance Scale, with the preference going to Coding on the basis of ease of scoring and brevity.

RELIABILITY

The reliability coefficients of the individual

TABLE 8

RELIABILITY AND STANDARD ERROR OF MEASUREMENT* OF THE WISC TESTS

N = 200 FOR EACH AGE LEVEL

	Age 7 1/2		Age 10 1/2		Age 13 1/2	
	r	SE_m	r	SE_m	r	SE_m
Information	.66	1.75	.80	1.34	.82	1.27
Comprehension	.59	1.92	.73	1.56	.71	1.62
Arithmetic	.63	1.82	.84	1.20	.77	1.44
Comparison	.66	1.75	.81	1.31	.79	1.37
Vocabulary	.77	1.44	.91	.90	.90	.95
Digit Span	.60	2.45	.59	1.92	.50	2.12
Verbal Score (without Digit Span)	.88	5.19	.96	3.00	.96	3.00
Picture Completion	.59	1.92	.66	1.75	.68	1.70
Picture Arrangement	.72	1.59	.71	1.62	.72	1.59
Block Design	.84	1.20	.87	1.08	.88	1.04
Object Assembly	.63	1.82	.63	1.82	.71	1.62
Coding†	.60	1.90	—	—	—	—
Mazes	.79	1.37	.81	1.31	.75	1.50
Performance Score (without Coding and Mazes)	.86	5.61	.89	4.98	.90	4.74
Full Scale Score (without Digit Span, Coding and Mazes)	.92	4.25	.95	3.36	.94	3.68

*SE_m is in Scaled Score units for the tests and in IQ units for the Verbal, Performance and Full Scale Scores.

†Based on correlating Coding A and Coding B, 115 cases. See text for explanation. For age 8 1/2 the value is .56 for 91 cases.

tests and of the Verbal, Performance and Full Scale Scores are presented in Table 8 for ages 7½, 10½ and 13½. These three ages were selected for presentation as being probably most representative of the age range for which the *Wechsler Intelligence Scale for Children* is designed. Reliability coefficients have been computed by the split-half technique, with appropriate correction for full length of the test by the Spearman-Brown formula.

This technique could not legitimately be used for estimating the reliability of the Coding test, which is essentially a speed test; nor did the Digit Span test lend itself to such treatment because of its administration as two separate subtests—Digits Forward and Digits Backward. The reliability coefficients reported for Coding were made possible because, for age 7½ and 8½, many of the children were given both Coding A and Coding B. (See administrative manual for description of these tests.) The reported values thus are based on an alternate test situation. The coefficients presumably would be a little higher if scores on Coding A were correlated with scores on a strict alternate form. The reliability coefficients shown for the Digit Span test are based on the correlation between scores on Digits Forward and scores on Digits Backward corrected according to the Spearman-Brown formula.

For the composite scores (Verbal, Performance and Full Scale Scores) the sum of the scores on odd items in the contributing tests were correlated with the sum of the even items.

The reliability coefficients presented in these tables should be carefully considered by the conscientious clinician when interpreting the scores earned on separate tests, or differences between scores. The smaller the reliability of a given score, the less confidence one can have in the judgments made concerning a child's true ability based on that particular test. Judgments with respect to *differences* between scores on two tests of moderate reliability must be made with considerable caution—the lower the reliability of the scores, the more likelihood there is that the difference between them is due to chance rather than to any real difference in the abilities possessed by the child. As may be seen by reference to the reliability table, this caution is more necessary for some tests than for others. It is least necessary when working with the composite Verbal, Performance and Full Scale Scores, which are highly reliable.

As another statement of the stability of scores on the *Wechsler Intelligence Scale for Children,* Table 8 presents the standard error of measurement by test and age. This measure indicates the band of error which surrounds the child's test score. Thus, a SE_m of 1.75 for 7½-year-olds on Information indicates that the chances are about two out of three that a true score on this test is within 1.75 points of the obtained scaled score. One can be highly certain that the true score is within 5.25 points of the obtained score (5.25 is three times the SE_m of 1.75). Note that there are considerable differences in size of SE_m from test to test. For example, confidence in the stability of Block Design for 7½-year-olds is permissible within limits of ± 1.20 (chances two out of three) and ± 3.60 (high certainty). Obviously, the smaller the SE_m, the less allowance one needs to make for unreliability of the score. Differences between Block Design and Vocabulary scores are less likely to be due to chance than are differences between scores on Object Assembly and Comprehension. These facts call for special wariness in attempts to compare differences between test *profiles.*

The reader of Table 8 should not be confused by the discrepancy between the size of the SE_m for the individual tests, as contrasted with the SE_m for Verbal, Performance and Full Scale IQ's. For individual tests, the SE_m is in *scaled score units*; for the IQ's the SE_m is in *IQ units,* which are the ones in which most test users are interested. Thus, the SE_m of 5.19 for the Verbal IQ of 7½-year-olds indicates that the true IQ is probably (chances two out of three) within 5 points of IQ of the obtained IQ.

SCALED SCORES

The scaled scores have been so derived as to provide, at each age and for each of the separate tests, a mean scaled score of 10 and a standard deviation of 3. This was accomplished by preparing a cumulative frequency distribution of raw scores for each test at each age level and

setting each percentile point at its appropriate standard score value on a theoretical normal curve with a mean of 10 and standard deviation of 3. Scores for all ages on a single test were then listed in parallel columns and minor irregularities in the progression of scaled score equivalents from age to age were smoothed. The assumption that these irregularities were chance results of population sampling seemed to be the only tenable position. Few instances of such minor deviations were found, and the Scales are essentially a direct translation from raw scores to a normalized distribution of scaled scores with a mean of 10 points and standard deviation of 3 points.

The manual presents 33 tables for converting raw scores on each test into scaled scores. The tables for the mid-years (age of testing) were first made. Then tables for each four-month span were constructed by interpolation; thus there are tables for 6-0 through 6-3; 6-4 through 6-7, 6-8 through 6-11, etc. After securing raw scores on each test, the examiner converts them to scaled scores by using the appropriate age table for the subject. These scaled scores, then, are the basis for determining the IQ.

THE DEVIATION INTELLIGENCE QUOTIENT

One of the most important innovations in the standardization of the present Scale is that IQ's are obtained by comparing each subject's test performance not with a composite age group but exclusively with the scores earned by individuals in a single (that is, his or her own) age group.[2] With one stroke, the deviation IQ method cuts away much of the underbrush which has encumbered the problem of the variability of the individual's IQ. By keeping the standard deviation of IQ's identical from year to year, a child's obtained IQ does not vary unless his actual test performance as compared with his peers varies; if the standard deviations were not made identical, a child's obtained IQ might vary considerably from year to year, even though his relative ability remained constant. Apart from test unreliabilities, IQ's obtained by successive retests with the *WISC* automatically give the subject's relative position in the age group to which he belongs at each time of testing. If any changes are observed they may be ascribed to changes in the subject and not in the structure of the test nor its standardization, since in IQ units the standard deviations as well as the means of all age groups are identical. It is no longer a matter of discovering how many children test above or below a given IQ in the population, since the deviation IQ is by definition dependent on the normal distribution of the test scores.

Each person tested is assigned an IQ which, at his age, represents his relative intelligence rating. This IQ, and all others similarly obtained, are deviation IQ's since they indicate the amount by which a subject deviates above or below the average performance of individuals of his own age group. The IQ of 100 on the *WISC* is set equal to the mean total score for each age, and the standard deviation is set equal to 15 IQ points. In terms of percentile limits, the highest one per cent will have IQ's of 135 and above, and the lowest one per cent IQ's of 65 and below. The middle fifty per cent of children at each age will have IQ's from 90 to 110.

The IQ tables were constructed as follows: For each age the five Verbal scaled scores for each subject were summed and a mean and standard deviation of such sums computed. These sums were transformed into a distribution of IQ's with a mean of 100 and a standard deviation of 15. The same process was followed for translating the sums of the five Performance scaled scores to an IQ scale with a mean of 100 and an SD of 15. The IQ's based on a Full Scale Score of ten tests were similarly determined.

One set of three IQ tables (*V*, *P*, and *FS*) suffices for all ages since the process of scaling each test for each age resulted in similar means and sigmas of the sums of five or ten tests at all ages. This was not a fortuitous result, but is one that should be expected because each of the tests was standardized for each age so that the raw scores are converted into scaled scores with a mean of 10 and a standard deviation of 3.

[2]The deviation IQ concept has been similarly employed in some group tests, notably the Otis Tests and Pintner General Ability Tests.

INTELLIGENCE QUOTIENTS AND AGE

Having carried on all the standardization processes described above, a test of the statistical transformations from the original responses of the child to the final IQ's is whether the mean IQ's and their standard deviations for all ages approximate 100 and 15, respectively, when each subject's scores are now converted to IQ's. Table 9 shows the data for boys and girls and for the three Scales.

TABLE 9

THE MEAN AND S D OF IQ'S ON THE THREE SCALES BY AGE AND SEX

100 BOYS AND 100 GIRLS AT EACH AGE

Age*	VERBAL Mean	VERBAL SD	PERFORM-ANCE Mean	PERFORM-ANCE SD	FULL SCALE Mean	FULL SCALE SD
Boys						
5	99.5	13.8	98.6	16.9	99.0	15.4
6	100.6	14.8	98.9	16.0	99.7	15.7
7	99.6	14.2	100.0	15.2	99.8	14.6
8	101.7	15.5	102.1	15.4	102.1	15.5
9	99.2	16.2	99.2	16.9	99.1	17.1
10	102.1	15.2	101.7	15.3	102.1	15.5
11	102.0	16.4	99.8	15.4	101.1	16.1
12	102.4	16.7	101.5	15.3	102.1	16.5
13	101.9	14.7	100.6	15.2	101.4	14.1
14	101.6	16.6	100.7	15.1	101.3	15.9
15	102.4	13.5	100.4	14.5	101.6	13.7
All	101.2	15.3	100.3	15.6	100.8	15.6
Girls						
5	99.8	13.4	101.0	13.9	100.4	13.2
6	100.6	13.9	101.4	13.4	101.0	13.6
7	100.5	11.7	100.9	12.0	100.7	11.4
8	97.8	15.8	97.9	15.3	97.6	15.8
9	100.4	13.0	100.4	14.0	100.4	13.1
10	98.4	16.7	98.4	13.8	98.2	15.3
11	97.9	14.5	99.6	13.8	98.6	13.8
12	97.5	14.7	100.0	14.4	98.5	14.5
13	98.3	16.2	98.7	15.8	98.3	15.9
14	97.6	14.3	100.2	15.3	98.7	14.4
15	97.3	16.4	97.9	15.4	97.4	16.5
All	98.7	14.7	99.7	14.4	99.1	14.4
Boys and Girls						
(N = 2200)	100.0	15.1	100.0	15.0	100.0	15.0

*Read these ages as 5½, 6½, etc.

The bottom line of values shows that for all cases—2200 in all—the requirement is exactly met. However, for boys and girls separately and for different ages, small discrepancies occur. The age-to-age discrepancies are caused in part by the fact that to secure one IQ table for all ages small discrepancies in means and standard deviations of scaled scores for the eleven ages were eliminated by averaging. It was assumed that these discrepancies were due to sampling and that smoothing of the data in making scaled scores and the IQ tables was justified.

However, the sex differences are not so easy to explain. Scaled score and IQ tables were made by treating boys and girls as members of one sample. The mean IQ's, however, show that boys in the standardization sample generally were slightly superior to girls. The superiority is primarily in the older ages, and the differences are small. On the Verbal Scale, the boys excel the girls by more than three points at ages 8, and 10 through 15. On the Performance Scale, the difference favors the boys by more than three points at ages 8 and 10, and the girls are ahead at ages 5, 6, 7, and 9. On the Full Scale, boys have higher IQ's than girls by 2.5 to 4.5 points at 7 ages, while girls are ahead by smaller amounts at four ages.

How shall one interpret these sex differences? Three explanations come to mind:

A. The tests are fair to both boys and

TABLE 10

DISTRIBUTION OF IQ'S OF RURAL AND URBAN CHILDREN

IQ	VERBAL Rural	VERBAL Urban	PERFORM-ANCE Rural	PERFORM-ANCE Urban	FULL SCALE Rural	FULL SCALE Urban
150-154		1				
145-149		3		1		2
140-144		7		2		2
135-139	2	8	3	5	3	7
130-134	9	16	2	12	2	16
125-129	13	40	14	45	10	41
120-124	22	78	29	82	27	69
115-119	37	115	52	101	46	117
110-114	67	143	92	180	58	167
105-109	85	189	81	159	102	195
100-104	110	211	137	193	122	194
95- 99	130	170	93	160	119	194
90- 94	116	150	103	172	119	134
85- 89	94	96	89	108	68	87
80- 84	69	67	35	38	66	55
75- 79	31	22	50	42	39	28
70- 74	15	7	23	14	25	12
65- 69	11	1	6	10	6	5
60- 64	6	2	8	3	4	1
55- 59	1	1	1		2	1
N	818	1327	818	1327	818	1327
Mean	97.3	103.3	98.5	102.5	97.6	103.2
SD	13.5	13.4	13.8	13.5	13.5	12.9

girls, and boys actually do excel girls, especially at the later ages.

B. Boys and girls are the same in mental ability, but the chosen test items turned out to be slightly biased in favor of the boys.

C. Again, assuming that general ability is not sex differentiated, the sampling of boys was somehow chosen with a slight bias.

The data at hand do not permit a resolution of these three choices. The safest assumption is that factors described in (B) and (C) are involved. Terman and Merrill [1, 3] found the same situation in their 1937 Revision of the Stanford-Binet examination, and likewise could find no definitive answer from their data.

All in all, the preliminary studies leading to inclusion of test items and the sampling itself were fortunate enough to result in mean IQ's of boys and girls which are essentially equal. For all practical purposes the clinical examiner can ignore sex differences. A difference in mean scores of three points, for example, is really a plus and minus difference of 1 ½ points from the actual norms based on both sexes.

INTELLIGENCE QUOTIENTS OF RURAL AND URBAN CHILDREN

Table 10 distributes the IQ's of urban and rural children for all ages, the 55 known feebleminded cases being excluded. As has been found in many researches, urban children score higher on mental tests. The Full Scale difference in the standardization sample is 5.6 points. The Verbal Scale difference is 6 points, and the Performance Scale difference is 4 points. Terman and Merrill report an urban-rural differential of 6.5 points for the 1937 Revision of Stanford-Binet.

TABLE 11

DISTRIBUTION OF VERBAL IQ'S FOR EACH OCCUPATIONAL GROUP, FOR THE FEEBLEMINDED GROUP, AND FOR ALL CASES

Verbal IQ	Father's Occupation*										
	1	2	3	4	5	6	7	8	9	FM	All
150-154								1			1
145-149	1		1						1		3
140-144	4	1	1	1							7
135-139	3		4	2			1				10
130-134	6	3	4	6	6						25
125-129	19	4	4	7	9	7	1	2			53
120-124	19	9	23	14	9	15	5	3	3		100
115-119	20	11	24	26	30	20	6	12	3		152
110-114	20	20	41	42	39	27	5	15	1		210
105-109	20	19	39	44	56	44	17	32	3		274
100-104	29	21	40	48	59	57	19	46	2		321
95- 99	17	26	25	45	57	62	16	44	8		300
90- 94	8	35	21	24	55	60	23	39	1		266
85- 89	6	29	17	11	39	31	11	42	4		190
80- 84	2	24	9	6	21	23	12	37	2		136
75- 79	1	12	3	2	7	9	2	16	1	2	55
70- 74		2		2	4	4	3	7		8	30
65- 69		4			1	2		4	1	9	21
60- 64	1	2				2	1	2		8	16
55- 59					1			1		14	16
50- 54										9	9
45- 49										4	4
44 and below										1	1
N	176	222	256	280	393	363	122	303	30	55	2200
Mean	110.9	96.8	105.9	105.2	100.8	98.8	97.6	94.6	100.6	59.6	100.0
SD	14.0	14.6	12.9	11.7	12.6	12.2	12.8	12.8	16.1	8.1	15.1

*See Code in Table 4.

INTELLIGENCE QUOTIENTS AND FATHER'S OCCUPATION

Distributions of Verbal, Performance, and Full Scale IQ's are given in Tables 11, 12, and 13 for each of the nine occupational categories, and for the known feebleminded cases. Particular interest attaches to the differences in mean scores of these various groups. Table 14 approximates the data for the full scale.

These differences in mean IQ's for occupational groups are considerable but not as great as those reported by Terman and Merrill for the 1937 Revision. The categories are not exactly comparable. Group A corresponds fairly closely to Terman and Merrill's groups I and II, for which they give eight mean IQ's for different age groupings, the median of these being about 115 as compared with 110 from the Wechsler data. For several age groupings in their category IV, rural owners, the median of the mean IQ's is about 94 as compared with 97 for Group E. Their group VII and the Wechsler group F are comparable, and yield mean IQ's of about 97 (median for mean IQ's of four ages) and 94, respectively.

For the lower socioeconomic groups, the two tests yield similar mean IQ's of the order of 95. The somewhat higher mean IQ of Terman and Merrill's higher sociometric groups, 115 vs. 110, may be accounted for, in part, by the greater verbal loading in the Stanford-Binet tests. For *WISC,* it is noted in Tables 11 and 12 that socioeconomic differences are greater on the Verbal Scale than on the Performance Scale.

Whatever other social implications one may consider, these mean differences between groups should not be allowed to overshadow

TABLE 12

DISTRIBUTION OF PERFORMANCE IQ'S FOR EACH OCCUPATIONAL GROUP, FOR THE FEEBLEMINDED GROUP, AND FOR ALL CASES

Performance IQ	Father's Occupation										
	1	2	3	4	5	6	7	8	9	FM	All
150-154											
145-149	1										1
140-144	1			1							2
135-139		1	4	1	1	1					8
130-134	3	1	3	3	3	1					14
125-129	16	2	8	5	10	12	4	2			59
120-124	13	11	17	19	17	21	3	9	1		111
115-119	22	12	26	26	27	20	7	11	2		153
110-114	31	26	40	48	60	30	7	24	6		272
105-109	21	21	43	38	45	36	12	22	2		240
100-104	20	41	32	45	69	57	17	43	6		330
95- 99	18	22	30	34	44	47	15	42	1	1	254
90- 94	15	24	24	29	48	57	27	47	4		275
85- 89	8	23	15	18	36	37	10	47	3	2	199
80- 84	1	11	6	8	14	16	5	11	1	1	74
75- 79	4	17	6		10	17	8	28	2	3	95
70- 74	1	8	1	3	4	8	3	8	1	5	42
65- 69	1	1	1	1	2	2	4	3	1	9	25
60- 64		1		1	3	1		5		11	22
55- 59								1		7	8
50- 54										7	7
45- 49										6	6
44 and below										3	3
N	176	222	256	280	393	363	122	303	30	55	2200
Mean	107.8	98.6	105.3	104.3	101.6	99.5	96.9	94.9	98.3	61.6	100.0
SD	13.4	13.9	12.9	12.2	13.0	13.5	13.7	13.3	13.6	12.2	15.0

the fact of great overlap in the distributions of IQ's for the various occupational groups.

THE FEEBLEMINDED

Wechsler classes IQ's under 70 as evidence of feeblemindedness. On the basis of a mean of 100 and a standard deviation of 15, 2.2 per cent of cases should have IQ's below 70. This is a statistical definition; it says that arbitrarily 2.2 per cent of the cases are feebleminded. The clinical and social significance of feeblemindedness is not defined by these numbers except that if this percentage is very far away from the proportion who are actually classified as feebleminded in clinical practice, the statistical concept will have to be changed. Roughly, the test author considers that about 3 per cent of the population could well be classified as feebleminded by a test; this seems to be reasonable in the light of actual practice.

In the standardization it was decided to let about 2.5 per cent of the sample be made up of known feebleminded cases. The resultant distribution of IQ's of these 55 cases is shown in the columns headed FM in Tables 11, 12, and 13. Some of these 55 institutionalized and school-identified feebleminded cases tested above 70. On the *V* and *P* Scales, 10 and 12 cases, respectively, tested over 70; but when *V* and *P* scores were made into Full Scale scores for these subjects, only 4 showed up as having IQ's over 70 on the Full Scale.

Some children in the general sample tested below the level set for feeblemindedness; in the Full Scale the number is 19, which, added to the 51 cases (55—4), equals 70 cases, or 3.2 per cent of the total sample.

In making the IQ scales on a deviation basis it is possible to assign IQ's as low as the sum

TABLE 13

DISTRIBUTION OF FULL SCALE IQ'S FOR EACH OCCUPATIONAL GROUP, FOR THE FEEBLEMINDED GROUP, AND FOR ALL CASES

Full Scale IQ	Father's Occupation										
	1	2	3	4	5	6	7	8	9	FM	All
150-154											
145-149	1							1			2
140-144			1	1							2
135-139	2	2	2	1	1	1	1				10
130-134	7	1	4	2	3				1		18
125-129	15	2	12	9	5	7		1			51
120-124	23	8	16	15	14	11	1	6	2		96
115-119	23	17	26	27	30	21	8	8	3		163
110-114	26	14	43	35	48	28	12	16	3		225
105-109	18	24	40	61	63	46	14	28	3		297
100-104	24	27	39	47	61	62	11	43	2		316
95- 99	18	28	37	35	59	58	22	51	5		313
90- 94	8	34	11	29	39	58	22	49	3		253
85- 89	6	21	9	9	31	32	11	34	2		155
80- 84	2	23	10	6	23	16	12	26	3	1	122
75- 79	1	12	6	2	10	13	1	21	1	2	69
70- 74	1	7			4	7	7	10	1	1	38
65- 69		2		1				7	1	8	19
60- 64	1				1	3				6	11
55- 59					1			2		14	17
50- 54										11	11
45- 49										6	6
44 and below										6	6
N	176	222	256	280	393	363	122	303	30	55	2200
Mean	110.3	97.4	106.2	105.2	101.3	99.1	97.0	94.2	99.5	56.6	100.0
SD	13.3	14.0	12.4	11.1	12.4	12.2	12.5	12.8	15.2	9.5	15.0

TABLE 14
APPROXIMATE WECHSLER IQ'S FOR OCCUPATIONAL CATEGORIES

	Approximate Wechsler IQ's
A. (1) Professional and semi-professional workers	110
B. (3) Proprietors, managers and officials, and (4) Clerical, sales and kindred workers	105+
C. (5) Craftsmen, foremen and kindred workers, and (6) Operatives and kindred workers	100
D. (7) Domestic, protective and other services workers	97
E. (2) Farmers and farm managers	97
F. (8) Farm laborers and foremen, and laborers	94

of the scaled scores obtainable. It was felt that IQ's under 45 would not be discriminatively meaningful, so the IQ tables stop at that point. Persons who have scaled scores yielding IQ's below 45 can be recorded as "44 or below."

THE SUPERIOR

No attempt was made to isolate a unique group of very superior children as was done with the feebleminded. It was assumed that the superior children were in the school systems and would enter the sample in proper sampling proportions. On the Full Scale, 1.5 per cent tested above an IQ of 130, whereas 2.2 per cent were expected. On the Verbal Scale the percentage was 2.1 per cent, and for the Performance Scale, 1.1 per cent. In the published IQ tables the highest IQ assigned is 156, but children whose scaled scores are higher than those required to attain this IQ can be recorded as being 156 or above. Differentiation above this point probably is not necessary.

To examiners who have been accustomed to secure IQ's on the order of 20 to 25 or 170 to 180, lack of very low and very high IQ's on *WISC* may at first be a little disturbing. They should be reminded that the range of IQ's on a deviation scale can be quite arbitrary. For example, if the mean were set at 100 and the standard deviation at 20 (instead of 15), lower and higher IQ's would be secured arbitrarily. The reason for setting the standard deviation at 15 is that it approximates the empirical standard deviation of about 16 secured by Terman and Merrill by an age-scale method. With standard deviations so similar, *WISC* will approximate in meaning (as far as size of the number is concerned) the IQ's secured by the Stanford-Binet Revision.

Received December 6, 1949.
Early publication.

REFERENCES

1. MCNEMAR, Q. *The revision of the Stanford-Binet scale.* Boston: Houghton Mifflin, 1942.
2. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
3. TERMAN, L. M., AND MERRILL, MAUDE A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
4. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.