VERBAL AND ABSTRACT ELEMENTS IN INTELLIGENCE EXAMINATIONS

JOHN P. HERRING

Bureau of Educational Research, Bloomsburg State Normal School, Bloomsburg,
Pennsylvania

This is a study of relations existing between human intelligence on the one hand and certain definite abilities on the other. These definite abilities are comprised in two somewhat distinct scales; *first*, a scale of situations extending from very concrete to very abstract; and *second*, a scale of situations extending from very non-verbal to very verbal.

The conclusions of such a study as this should relate both to pure and to applied psychology. In human nature, original and acquired taken together, what relations exist among the abilities named? And in practice, for the purposes of prediction and of control, what relations? How may we best estimate intelligence—through samplings of concrete or of abstract work—non-verbal or verbal?

We will call the scales C for concrete and N for non-verbal.

Scale C involves two concepts: *concrete* and *abstract*. *Concrete* is taken to mean: pertaining to things or objects; having no, or little, reference to the investigation of relations between objects; not problematical in character. *Abstract* is taken to mean: pertaining to relations between objects; involving higher intellectual processes, particularly analytic and synthetic thinking; problematical in character. The distinction is not one between working *with* and working *without* material objects; for the same material things may enter equally into responses that are most widely separated in this scale. The Stenquist Mechanical Tests probably involve far less of the concrete and more of the abstract than is obvious.

As illustrative of concrete situations we may take the following:

(1) Threading the mazes of the Army Beta test. More difficult and complicated mazes would be more abstract, more problematical; (2) Classifying blue squares and red triangles as in the Dearborn Intelligence Examination. Tests differing from these, *not* by involving closer discriminations of the same sort but by involving similar and more complicated discriminations, would be more abstract; (3) Shingling a roof after the process is well mechanized. Learning to shingle a roof by experimentation without instruction would be more problematical; (4) Street-sweeping.

As further illustration of *abstract* situations we may take these:

filling a hard completion test; solving hard analogies; reading difficult scientific literature; planning and supervising the construction of large bridges; conducting the business affairs of Standard Oil.

Scale N also involves two concepts, *non-verbal* and *verbal*. Verbal is taken to mean: involving the use of words and other symbols such as numbers, mathematical signs of equality and inequality or any subjective imagery taken to represent and to retain during further study earlier outcomes of investigation. Non-verbal is taken to mean: not involving the use of symbols; pertaining to experiences in which the subject deals immediately with its object; not involving representative symbolism.

Verbal or symbolic situations may be illustrated, obviously enough, by verbal completion and analogies tests and arithmetical problems, and somewhat less obviously by teachers' estimates of children's intelligence; non-verbal non-symbolic situations by writing S between two numbers or symbols that are the same, and D between two that differ; by threading mazes; and by learning digit symbol combinations.

It is assumed for these two scales: *first*, that it is false to classify human behavior dualistically as necessarily either concrete or abstract; necessarily either non-verbal or verbal; and *second* that human behavior may be described in these two phases as comprising responses, the distribution of which is *continuous* throughout both scale C and scale N, some responses being, for instance, extremely concrete; some just a little less concrete, and so on, without gap, throughout the gamut to extremely abstract. These assumptions are roughly borne out by a number of subjective judgments which will be presented.

The definitions of the concepts *abstract*, *concrete*, and *verbal* are not regarded as final or even as necessarily valid, but much rather as starting points for the work of classifying the samplings of behavior elicited by a number of group examinations. These examinations were of two classes, selected for two distinct purposes: *first*, a group, seven in number, employed as a weighted composite criterion of intelligence; and *second*, a group, five in number, employed as the means of measuring concrete-abstract abilities and non-verbal-verbal abilities. The first group, the criterion of intelligence, comprised the following with weights as stated: *First*, the Stanford Revision of the Binet Simon Tests, weight 3; *second*, educational age found by averaging the results of the Thorndike Reading Test Alpha 2, the Woody-McCall Arithmetic Test, the Monroe Arithmetical Reasoning

Test, and an Ayres Buckingham Spelling Test, weight 3; *third*, a measure of intelligence based upon the age and grade reached at the time of the study, weight 3; *fourth*, the National Intelligence Tests, Forms A and B, weight 2; *fifth*, the Thorndike Reading Test Alpha 2, weight 1; *sixth*, teachers' estimates of intelligence, weight 1; and *seventh*, the Kelley Trabue Completion Test Alpha, weight 1.

This weighted composite criterion of intelligence comprises a very wide range of activities, involves several modes of observation obtained upon a number of different and widely separated days, and exhibits self-correlations between 0.9 and 0.8 in age groups. The validity of this criterion is assumed. It is just the sort of criterion psychological investigators everywhere assume. Any who do not accept its validity, may, on that account, differ with the conclusions offered.

The second group, used as a measure of the dependent variables, C and N, comprised the following:

The Dearborn Group Intelligence Examination;
The Army Beta Examination
The Pressey Primer Mental Survey
The Indiana Cross-cut Tests
The Thorndike Visual Vocabulary

All the examinations of both groups were divided into short units, each about a page in length, assumed to be homogeneous with respect both to scale C and to scale N. These units were bound into a book of 63 pages, one page per unit, which was submitted to five judges, who classified its content by assigning each unit a position from 1 to 7 upon both scales. These judgments combined two elements: *first*, an averaging of opinion concerning the content of the concepts themselves (abstract, concrete, verbal and non-verbal), and *second*, an averaging of opinion concerning the content of the situations represented in the book of units. It was important to give each judge a voice in the determination of the meaning of the concepts as well as in the classification of the test material. Relatively little stress is laid, therefore, upon the rather formal definitions already presented. To one asking for better definition would be given the average findings of the judges themselves. The term "abstract" comes now to mean—having the character of situations like those marked at or above 5 in this book of units. The uniformity of judgment is indicated by the average correlation, judge with judge, of 0.542 ± 0.019 .

A result of this classification is the existence of two subjective scales, by comparison with which any homogeneous portion of test material or of behavior, may be rated for abstractness and for symbolism. Among the less obvious comparisons which are thus made possible are these:

1. Digit symbol tests are distinctly concrete and very non-symbolic. The "symbols" in such tests are the things themselves, not symbols of other things.

2. The Army Beta Maze Test is as concrete as anything found, and (expectedly) as non-verbal as it is concrete.

3. The Beta Cube-counting Test is distinctly an abstract one and, as would be expected, is very non-verbal.

4. The Woody McCall Arithmetic Test is rated 5 in abstractness and about $3\frac{1}{2}$ in verbal quality.

5. Arithmetical problems are rated very abstract.

6. The Stanford Binet Tests average as a whole rather abstract than concrete, rather verbal than non-verbal, but they do *not* occupy extreme position in either respect. In both scales which extend potentially from 1 to 7 they are rated about 5, leaving in both series about 15 out of 63 samples above.

7. The position of the composite criterion of intelligence is just about 5, like the Stanford Binet, upon both scales.

8. The scale of non-verbal—verbal quality runs from the Beta Maze Test at one extreme to the Thorndike Reading Test Alpha 2 at the other. The scale of concreteness—abstractness extends from the Beta Maze at one end to the Analogies of the National Intelligence Tests at the other.

In the field of applied psychology, for the purposes of prediction and control, the following results and conclusions may be tentatively stated.

The raw correlations include those of four grades in a public school and those of three age groups. These are grades IV, V, VI, and VII. They have membership ranging from 23 to 41 per grade and totalling 118. They are extraordinarily homogeneous with respect to mental age, the standard deviations averaging only 4 months, as against perhaps three or four times that spread in many public school grades. This reduction of variation operates to reduce correlations but leaves them comparable. Averaging the several grades, we have correlations as follows:

Intelligence:

With concrete tests	0.25
With the middle portion of scale C	0.36
With abstract tests	0.58
With non-verbals	0.27
With the middle of scale N	0.31
With the abstract tests	0.54

Similar correlations for age groups consisting of averages of ten year olds, eleven year olds and twelve year olds, are as follows:

Intelligence: with

Concrete tests	0.47
Middle Scale C	0.58
Abstract tests	0.84
Non-verbal tests	0.49
Middle scale N	0.70
Abstract tests	0.88

The P.E.'s of the correlations entering into all these averages range from 0.056 to 0.130 and average 0.10. The average *r*'s themselves are of course much more reliable. Further, these correlations vary almost as do the S.D.'s of the mental ages; the larger the S.D., the higher the *r*. So close is the correspondence that the average correlation of the *r*'s with the S.D.'s is 0.906 for raw and 0.886 for corrected coefficients. Increase in the S.D.'s involved in a correlation results in an increase of correlation which is paralleled by no fact or change in human nature; it is a function of increase in the variability of the group measured. With regard to the correlation existing between the two traits, it is wholly spurious. A correlation of 0.4 in a group having a standard variability of 5 months of mental age may mean just the same degree of mutual implication as a correlation of 0.8 in a group having a much higher standard variability. It appears, therefore, that these correlations have a reliability much greater than the customary statistical formula of probable error reveals.

The correlations point in the direction of the following conclusions:

1. Abstract and verbal tests afford better means for the prediction of human intelligence and the control of human situations than do concrete and non-verbal tests.

2. Abstract and verbal tests will make the best material that can

be selected from such tests as those employed in this investigation for the purpose of inclusion in intelligence examinations.

3. The middle mongrel portions of the two scales are to be distinguished by means of their correlations from the concrete and non-verbal portions; they are in general distinguishably better material for intelligence examinations than N1 and C1 and very inferior to N3 and C3.

4. These abstract and verbal tests, besides being correlated more highly with intelligence, have higher coefficients of reliability and are, for this reason, other things equal, better content for intelligence examinations.

It is by no means argued that it is never proper to use concrete and non-verbal tests of intelligence, for there are circumstances in which they are the feasible form of procedure; nor that they should never be included in a battery of tests for literate and intelligent adults, for the data do not permit the conclusion that concrete and non-verbal tests may not properly supplement the others in a battery of tests.

Perhaps the outstanding conclusion is the following: it seems to be the more purely abstract and the more purely verbal tests that afford the closer measures of intelligence.

We come now to examine the corrected correlations between intelligence and the different portions of these two scales.

COEFFICIENTS OF CORRELATION CORRECTED FOR ATTENUATION, r_{MC1} , . . . r_{MN3} ,
RAW AND CORRECTED
(M means intelligence)

	RAW	CORRECTED
C1	0 474	0 800
C2	0 578	0 686
C3	0 837	0 963
N1	0 492	0 738
N2	0 697	0 865
N3	0 880	0 951

The following interpretation is suggested:

Total human nature and the mutual demands of human beings have become such that intelligence, as it is required for success in contemporaneous human society, comprises largely the ability to deal effectively with situations involving the use of language and of mathematical and other symbols, both subjective and conventional, and also the ability to control situations through the analysis and inter-

pretation of novel and complicated phenomena. It is hard to think of any important posts of social responsibility, from railway surveying to international law, of which this does not seem true. It is the hod-carriers who typically deal with concrete situations. The master architect controls the situations but he does so through such as the hod-carrier, whom he reaches by the way of many and abstract processes involving complex symbolisms.