

RATER AGREEMENTS IN ASSIGNING STANFORD-BINET ITEMS TO GUILFORD'S STRUCTURE OF INTELLECT OPERATIONS CATEGORIES

CALVIN O. DYER, CYNTHIA NEIGLER, and
JOHN E. MILHOLLAND

University of Michigan

Summary: Nineteen school psychologists assigned the 142 items in the Form L-M of the Stanford-Binet Intelligence Scale to the five Operations categories of Guilford's Structure of Intellect model, following flow charts prepared for this purpose by Meeker (1965). On the average, one rater agreed with another on about half the items, and their modal assignments agreed with Meeker's (1969) assignments on only 81 (57%) of the items. These levels of agreement are judged not to be high enough to justify classifying Stanford-Binet items in accordance with the Structure of Intellect Operations categories.

For years the Stanford-Binet Intelligence Scale has been widely used by school psychologists and others in the field of psychological testing. Because of its established reputation, it is commonly used as a measure of intellectual ability. However, there are some psychologists who consider its single global measure of intelligence inadequate for the purpose of identifying and dealing with specific abilities in children and look for ways to penetrate beyond the single score.

By contrast, Guilford's Structure of Intellect (Guilford, 1967) offers a detailed partition of intellectual abilities, classifying them under the dimensions of Operations, Content, and Products. His model can accommodate 120 different abilities, each specified by a trigraph consisting of a letter for one Operation, one Content, and one Product.

Some psychological testers have been using performance on individual Stanford-Binet test items to assess specific intellectual abilities, although Terman and Merrill (1960) specifically warned against such a practice in their manual for the 1960 (Form L-M) revision of the test. Ramsey and Vane (1970) investigated Valett's (1963) scheme of item classification on the basis of five factors. They, however, found seven factors and also concluded that, contrary to Valett's assertion, performance on subtest items does not depend on only one aspect of intelligence. Darrah (Note 1) likewise concluded that a group factor structure was indicated. He identified five factors—one verbal, two memory (recall and concentration), a judgment and reasoning factor, and one more which was complex and not interpreted. The memory and reasoning factors that Darrah obtained agreed well with similar factors reported by

A more extended report of this study is available from Calvin O. Dyer, University of Michigan, Ann Arbor, Michigan 48104.

Sattler (1965) and Valett. He also agrees with them on the presence of a verbal factor although the item content for this factor seemed unclear.

Newland and Meeker (1963) used a nonstatistical method in classifying the Stanford-Binet items, assigning a trigraph from Guilford's model to each. They were satisfied with the outcome and Meeker (1969) used this procedure in California where the state has legislated funds for the education of its gifted children. Trigraphing the items enables her to evaluate each child's abilities in many specified areas and thus to plan individual programs to better educate each person.

In 1965 Meeker presented flowcharts for standardizing a method of trigraphing Stanford-Binet items. Following a map of questions with yes or no answers leads to the Operation, Content, and Product which is most appropriate for the item. However, Newland and Meeker recognized the need for interjudge reliability for this to become a useful, accepted process.

The purpose of this study is primarily to determine whether a group of school psychologists can agree on trigraphing test items according to Meeker's procedure. The extent to which this group of school psychologists agree with Meeker's own suggested trigraphs and with other studies of item classification of the Stanford-Binet will also be examined.

METHOD

At the close of a preconvention institute conducted by the APA Division of School Psychology in 1965, the senior author agreed to ask members of the institute to assign all the items of the Stanford-Binet Intelligence Scale, Form L-M, to trigraphs of the Structure of Intellect model. Many persons in this group did not respond, so other school psychologists with whom the author was acquainted were enlisted. In all, there were 20 persons, some of whom were graduate students, who were asked to provide ratings.

There are 6 items and 1 alternate for each of 19 age levels, plus 8 items and 1 alternate at the average adult level; thus, 142 ratings were required. The raters were all provided with Meeker's flowcharts to assist them. Although they were given the option of supplying more than one response per item, this study deals with only their first choices and with just the Operations dimension. The Operations responses of one rater were not complete so the data consist of 19 sets of ratings instead of 20.

RESULTS

The numbers of test items achieving various levels of agreement among raters are shown in Table 1. Since there are five categories of Operations, the greatest possible disagreement among raters would result in approximately four having to agree on each item. The actual lowest frequency was six, for two items. More than half the items, however, were assigned to the same categories by 12 or more raters; about three fourths of them by at least 10 (a majority). All 19 raters concur on only eight items.

The average number of assignments on which one rater agreed with another ranged from 57.1 (40.2%) to 79.2 (55.8%), with the group average

being 71.5 (50.4%). Thus it may be said that one judge is likely to agree with another judge on about half the items. The largest number of items on which any two judges agreed was 105 (75%), the smallest number, 31 (22%).

A comparison of Meeker's (1969) assignments with those of the 19 raters in this study is given in Table 2. The assignments of the 19 raters were more evenly distributed over the five operations than Meeker's were. The largest discrepancies appear in Cognition, which Meeker used for 80 items as opposed to our raters' 52, and Convergent Production, to which Meeker assigned only 13 items compared to our raters' 41. Most of those differences are accounted for by the fact that 28 of the items that Meeker assigned to Cognition were assigned to Convergent Production by our raters. The main diagonal of the table, from lower left to upper right, includes the numbers of items agreed upon by Meeker and our raters in the various categories. Only 81 (57%) of the 142 items appear in this diagonal, and one would hardly consider this a spectacularly high correspondence.

Among our raters Cognition, Memory, and Convergent Production dominate the responses. Cognition is not only chosen the most frequently but also the most consistently. There is only one age level where no item is assigned to it. Convergent Production is the second most frequent choice and is also fairly consistent over all age levels.

DISCUSSION AND CONCLUSION

In order for the cataloguing of Stanford-Binet test items in accordance with Guilford's Structure of Intellect model to be useful, agreement on the assignment of items to Structure of Intellect categories would seem to be necessary. The data of this study, dealing with the Operations category only, provide evidence against the existence of a satisfactory level of agreement.

Table 1
Agreement Among Raters in the Assignment
of Items to Operations Categories

Number of raters agreeing	Number of items	Cumulative frequency	Cumulative per cent
19	8	8	5.6
18	9	17	12.0
17	5	22	15.5
16	6	28	19.7
15	3	31	21.8
14	18	49	34.5
13	9	58	40.8
12	15	73	51.4
11	10	83	58.5
10	21	104	73.2
9	20	124	87.3
8	10	134	94.4
7	6	140	98.6
6	2	142	100

Table 2
Number of Items Appearing in Each Combination of
Categories as Assigned by Meeker (1969) and
by the Raters in the Present Study

Category assigned in the present study	Category assigned by Meeker					
	Cognition	Memory	Convergent Production	Divergent Production	Evaluation	Total
Evaluation	1 1/2*	0	0	0	8 1/2	10
Divergent Production	2	0	1	2	1	6
Convergent Production	28	4 1/2	5	1	2 1/2	41
Memory	5	22	5	0	1	33
Cognition	43 1/2	3 1/2	2	0	3	52
Total	80	30	13	3	16	142

*Where two operations for one test item received an equal number of the raters' responses, it is represented in this table by one half for each operation involved.

Overall, about half the judgments of one rater would agree with those of another. The highest level of agreement of any one rater with any other was on about 56% of the items. Items varied somewhat in their ease of classification, with three fourths of them showing agreement by at least 10 raters, and a little more than half of them had 12 raters agreeing.

There is a trend for the raters to agree more on the higher age levels, but it is not so strong as to warrant the use of trigraphing even in the later years.

Our raters' responses differed from those recommended by Meeker (1969), largely in the operations of Cognition and Convergent Production. Our raters chose the latter for many of the items that Meeker assigned to the former. Not only do these 19 raters not agree well among themselves, but they also don't agree with Meeker's suggestions. It is possible that some of this disparity may be attributed to an underlying difference between the operations of Cognition, Memory, and Evaluation and those of Convergent and Divergent Production. The latter two refer to methods of arriving at certain answers, whereas the other operations refer to how a person acquires certain knowledge. Guilford himself acknowledges the possible confusion between the concepts of Cognition and Convergent Production. According to him, Cognition involves a person's having certain knowledge and being able to understand it, whereas in Convergent Production a person arrives at an answer, given certain specifications, through inferences.

Comparison of our raters' assignments with factor analytic studies of the Stanford-Binet supports the position that the test is heavily loaded on a Cognition factor. Ramsey and Vane (1970) found that the first factor they extracted accounted for much of the variance on 12 of the 18 subtests they examined, and they related this factor to the Cognition operation in Guilford's model. The raters in the present study felt also that the Stanford-Binet tests Cognition more than any other operation, and Cognition was present at all age levels except one.

REFERENCES

- Guilford, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.
- Meeker, M. A procedure for relating Stanford-Binet behavior samplings to Guilford's structure of the intellect. *Journal of School Psychology*, 1965, 3, 26-36.
- Meeker, M. *The structure of intellect*. Columbus, Ohio: Charles E. Merrill, 1969.
- Newland, T. E., & Meeker, M. Binet behavior samplings and Guilford's structure of the intellect. *Journal of School Psychology*, 1963-64, 2, 55-59.
- Ramsey, P. H., & Vane, J. R. A factor analytic study of the Stanford-Binet with young children. *Journal of School Psychology*, 1970, 8, 278-83.
- Sattler, J. M. Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, Form L-M. *Journal of Clinical Psychology*, 1965, 21, 173-179.
- Terman, L., & Merrill, M. A. *Stanford-Binet Intelligence Scale*, Form L-M. Boston: Houghton-Mifflin, 1960.
- Valett, R. E. A clinical profile for the Stanford-Binet. *Journal of School Psychology*, 1963-64, 2(1) 45-54.

Reference Note

1. Darrah, D. *Factor analytic validation of two item classification schemes of Stanford-Binet Intelligence Scale*. Unpublished research report, University of Michigan, 1973.

Calvin O. Dyer
 Professor of Education
 Combined Program in Education
 and Psychology
 2339 SEB
 University of Michigan
 Ann Arbor, Michigan 48104

Cynthia Neigler
 Graduate Student
 Combined Program in Education
 and Psychology
 2339 SEB
 University of Michigan
 Ann Arbor, Michigan 48104

John E. Milholland
 Professor of Psychology
 Psychology Department
 529 Thompson
 University of Michigan
 Ann Arbor, Michigan 48104

Received: May 13, 1974

Revision Received: August 15, 1974