DOCUMENT RESUME

ED 036 808 CG 004 783

AUTHOR Dunn, James A.
TITLE The OST: Theory, Instrument and Norms.
INSTITUTION American Institutes for Research, Washington, D.C.; American Psychological Association, Washington, D.C.
SPONS AGENCY Public Health Service (DHEW), Washington, D.C. Health Services and Mental Health Administration.
PUB DATE Aug 69
NOTE 26p.; Paper presented at the American Psychological Association Convention, Washington, D.C., August 31--September 4, 1969

EDRS PRICE EDRS Price MF-$0.25 HC-$1.40
DESCRIPTORS *Cognitive Development, *Cognitive Measurement, Concept Formation, Intelligence, *Intelligence Factors, Intelligence Tests, *Test Construction, Test Results, Test Validity

ABSTRACT

The OST is a technique quite different from anything reported elsewhere in the literature. The model underlying OST work assumes that an individual's intellect, at any given point in time, is the set of all information he has at his disposal at that point in time. The set of concepts an individual has at any point in time may radically change over time. The OST has reasonable face and construct validity; and yields scores that make a unique contribution to achievement score variance. The OST scores also reveal clear cut develcpmental differences that are at the same time both understandable and compatible with current theories of cognitive development. Because of its simplicity, and its inherent and nonthreatening appeal to children, there is the distinct possibility it may find its way to applied diagnostic use in clinical assessment areas. Its heuristic character has already been demonstrated and its research possibilities are being considered. (KJ)

THE OST: THEORY, INSTRUMENT AND NORMS[1,2]

James A. Dunn

American Institutes for Research

[1]Presented in Symposium, APA, Washington, D. C., 1969.

[2]Portions of research reported herein were supported by MH Grant #01428. The majority of data was collected where the author was on the faculty of the University of Michigan and analyzed while a Visiting Fellow at Harvard University.

There is a fairly high degree of agreement in psychology as to what constitutes conceptual behavior. The statements of Keller and Schoenfeld (1950) regarding this topic comprise the generally accepted view. Keller and Schoenfeld have argued that "one demonstrates conceptual behavior by acting in a certain way;" that behavior is considered conceptual "when a group of objects (or events) get the same response, when they form a class the members of which are reacted to similarly;" and that the essence of conceptual behavior is "generalization within classes."

While there has been general agreement as to what constitutes conceptual behavior, there has been far less agreement as to the underpinings of that behavior; i.e. what constitutes a concept, and whether some form of mediation is involved.

LITERATURE

Concept has been variously defined as: a response (Osgood, 1953); a construct (Heidbreder, 1946); a mental process (Morgan, 1956); a property (Ruch, 1948); and a relationship (Hilgard, 1967).

Early on Hull (1920) challenged the use of mediation as the explanatory construct for conceptual behavior.

Hull, it will be recalled, taught subjects to respond with nonsense words to stimuli similar to Chinese ideographs. Six lists of stimulus characters of increasing complexity were used. The same common elements were embedded in the characters of each list, and the same response was required whenever a given element was involved, regardless of stimulus complexity. The object was to see whether the learning of each successive list was facilitated. In brief, learning time to criterion was progressively shortened. In addition, many subjects who evidenced strong "concept formation behavior" (in the sense that their learning was facilitated) could not identify the character components to which they had been correctly responding. As a result it was concluded that subjects formed "concepts" only in the sense that their behavior generalized on the basis of the critical elements in the different characters.

Reversal shift and multiple alternation studies are generally felt to support mediation theory, however (Osgood, 1953; Hunt, 1961).

A reversal shift problem is one in which the subject is required to solve two concept learning tasks each involving several attributes with only two values. For example, the attribute "shape" might involve the two values triangle and circle. In the first task the individual might learn that positive instances are always triangles. In the second task the individual must reverse the polarity of the concept and respond to circularity instead of triangularity. With conceptual mediation individuals ought to learn the reversal shift concept faster than a second concept involving some other attribute as size or color. If there is no mediating response then there should be no difference in time (or trials) to criterion between the two tasks.

A double alternation problem is a problem in which the "meaning" of each preceeding response changes. In simple maze learning, for example, the first left turn may signal a coming left turn but the second left turn then becomes a signal for a coming right turn. That right turn then signals a second right turn but the second right turn signals a coming left turn, and so forth. The pattern in its simple form, would be: L L R R L L. The paradigm, of course, can generalize to more complex patterns such as: L R R L L L R R R R, and so forth. In double alternation problems the signal value of a response is nil unless the organism can do something analogous to counting; and, in more complex problems, something analogous to differential incrementing.

By way of comparative psychology, as one descends the phylogenetic scale, alternation problems become increasingly difficult for the subject to perform. Adult humans learn the use of alternation patterns very quickly, and can generalize readily to new series. Chimpanzees and children find it somewhat more difficult, but can still generalize from complex alternation patterns. Cats and dogs can learn simple alternation problems but cannot extend the series. No rat has learned even the simplest double alternation problem.

Verbal mediation theory receives support from work such as that of Riess (1940, 1946). Riess (1940) conditioned a galvanic skin response (GSR) to stimulus words and then tested for generalization to their synonyms and homonyms. Mean GSR gain to stimulus words was approximately 350 percent; gain associated with synonym presentation was approximately 140 percent; and gain to homonym presentation was only approximately 95 percent. The conclusion was that GSR gain to homonyms was simple stimulus generalization while GSR gain to synonyms was generalization based on verbal mediation. Some time later Riess (1946) showed that the effect of the mediating processes increased with growing maturity. He found that subjects below 8 years of age showed greater GSR gain to homonyms than to synonyms; subjects approximately 11 years of age showed equivalent homonym and synonym gain; subjects 14 years of age showed results approaching the adult pattern; and subjects 18 years of age showed results clearly in the pattern of his 1940 study.

In recent years attention has shifted from such questions as: whether mediation is a necessary construct, whether sufficiently demonstrable behavior can be generated to warrent its inference, and whether the mediation is verbal or not, to such questions as: the development of the cognitive structures through which mediation is effected (Piaget, 1950, 1964); the types of operations on which the structures can be focused (Guilford, 1959, 1967); and the nature and form in which the information organized in those structures is encoded and the strategies that can be employed (Bruner, 1964, 1966).

PROBLEM

The work reported herein was motivated by an interest in both the form and focus of concept formation, recognition, and utilization; and how that form and focus changes with age. Of particular interest were questions dealing with such issues as:

1) verbal vs. non-verbal (iconic) concept encodement;
2) the relative availability of concepts within one's cognitive structure;
3) reversability and the ability to multiple-classify;
4) concept formation vs. concept recognition;
5) equilibrality of definition of the reference points on verbally mediated classification dimensions;
6) concept and classificatory dimension hierarchies within the cognitive structure;
7) divergent as contrasted to convergent concept formation and recognition;
8) fluency and flexibility; and the like.

The origin of some of these issues, such as form of concept encodement, divergent vs. convergent processes, and fluency vs. flexibility, are well known. Interest in the others derives from the heuristic character of a model developed to integrate various ideas suggested by Piaget, Bruner, Guilford, Torrance, and others. Before commenting further on that model however, it would be well to recall the distinction between heuristic-exploratory research and analytic-evaluative research and to comment briefly on the nature of theory.

ON THEORY

A theory is a set of hypothetical equivalence statements asserted to be true. These equivalence statements may be verbal, mathematical, or analogue in language form. The same content can be expressed, albeit in varying degrees of precision and succinctness, in any language form. Although the content of a theory is independent of its language form, the availability of the content inherent in that theory is not.

Simon and Newell (1956) have argued that: "One theory can contain exactly the same logical content as another, but be infinitely more valuable than the other if the content is stated in such a way as to be easily manipulated, so that its logical content is actually (psychologically) available to the inquirer." Simon and Newell then go on to point out that analogue theory generally tends to make its content more readily available to the user than do mathematical or verbal language theories, and for that reason is generally heuristic and therefore useful for early exploratory studies.

The use of analogue theory is not without its dangers however. There is error associated with theoretical decision-making, just as there is with statistical decision-making. Theories differ in the degree to which they omit details which are true and admit details which are false. Analogue theory is very susceptible to the latter type of error. But, when one wishes to conduct research in largely unexplored and/or novel areas, he may wish to opt for Type I error.

THE OST MODEL

The model underlying OST work is a spacio-temporal analogue matrix model. In brief, it assumes that an individual's intellect, at any given point in time, is the set of all information he has at his disposal at that point in time. The set of concepts an individual has at any point in time, his time-point set, may change radically over time, however.[3]

A time-point set may be represented as an ordered distribution of points in an n-dimensional hyperspace; each point representing a unit, or element, of information, i.e. concept. Two units of information then define a classificatory dimension; a set of points, a region; a set of regions, a cognitive domain, or structure.

Since this concept space is being considered in terms of a quasi-physical analogue, one can consider not only such "pure" geometric characteristics as: a) number of points, b) proximity of points, and c) pattern of point distributions; but also d) degree of point resolution (precision of point definition), e) the speed at which a point may be located in the structure, f) the effort required for that location (the accessibility of a concept may be a function of speed and effort), g) the processes of point additions, differentiations and consolidations, etc.

Clearly the model is in no way assumed to parallel the real world. Like all

---

[3]A detailed description of the model and the variety of questions it has raised, "Toward an Heuristic for Research on the Cognitive Process," is being prepared for separate publication.

models it is simply a convenient, and expendable, mneumonic device to see how far one can go in generating various interesting research questions. There is little point in going into detail regarding the model since its function was to serve as an heuristic; as a source of researchable questions (Kerlinger (1964) calls these problems) rather than as a source of rigorously derived hypotheses.

THE OST INSTRUMENT

To explore questions raised by the matrix model of cognitive structure, it was obvious that a data collection procedure was needed which would, among many other things: a) yield behavioral as well as verbal data; b) afford scores on both divergent as well as convergent thought processes; c) allow for the investigation of concept availability, and so forth.

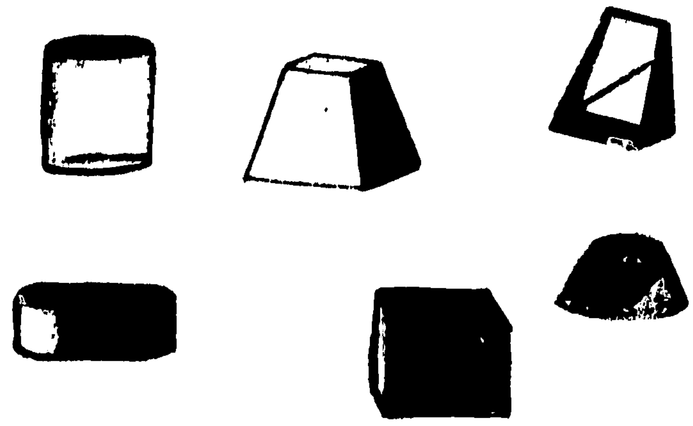
Many sorting tasks were considered and found wanting. The Vygotsky and the Goldstein-Sheerer tests came the closest to what was needed but they actually yield very little objective data. Moreover they do not consider divergent concept formation at all. In point of fact no objective object sorting task was found that did. Divergent process tests are almost exlusively in the verbal domain. In addition most object sorting tasks required sorting on only very gross, and relatively simple, conceptual bases and were designed for measuring such broad dimensions as concrete vs. abstract, or color vs. form preferences. The end result was the development of what appears to be a unique object sorting task.[4]

The procedure developed was an object sorting task (OST) in which six plastic blocks, differing on a variety of attribute dimensions, are repeatedly classified into equal dichotomous groups on the basis of those dimensions.

The blocks and the procedure have evolved over two prototypes and several pilot studies. The blocks used in the present version of the OST are precision milled from bulk plastic to tolerances of ±.05", vary in weight from 6 to 17 grams, in volume from 33 to 87 cc., and are either dip dyed or spray painted.

The objects vary on as many dimensions as there are possible equal dichotomous groupings minus one. Thus there are nine valid dichotomous combinations of the six objects taken three at a time. The tenth combination has no attribute justification and is used as the "blank" starting sort; i.e., the sort used by the examiner to start each new classification attempt of the subject.

Figure 1 -- OST BLOCKS



[4]Several extensive reviews of the literature have failed to discover a similar procedure. If the reader should know of one, the author would greatly appreciate the reference.

The sorts, rank ordered in order of difficulty, are:

1) curvilinearity
2) hue
3) transparency
4) height
5) perpendicularity of side
6) radial equality of the base
7) saturation
8) volume
9) area of the base.

These nine attributes represent three supra-ordinate attribute families: color, unidimensional form, and dimensional relationships.[5]

ADMINISTRATION

The OST procedure requires six pencils, six plastic blocks, a cardboard administration mat, a stop watch, and a protocol blank. Directions for administering and scoring are given in Appendix A.

The pencils are used to establish a multiple classification set and to provide the subject with a fixed amount of controlled practice prior to the actual task. They differ in three ways: three are sharpened and three are not; three have erasers and three do not; and three have pocket clips and three do not.

The pencils are placed on the administration mat and the examiner goes through a standard training procedure. The child is shown how the six pencils may be sorted into two equal dichotomous groups on the basis of pointedness. The pencils are then mixed up and the child is next shown how they may be classified on the basis of their eraser attribute. The pencils are again mixed up and the child is asked to sort them in a third way. If the child says he cannot, or fails to do so properly, the examiner points to the pocket clip on one examplar and asks, "Is there something special about these?". If the subject still cannot execute the sort, the examiner helps and prompts the subject until the third sort is achieved.

After the practice experience the pencils are removed, the six plastic blocks are positioned on the administration mat, and the "divergent processes" phase of the test is begun.

After each sort, the subject is asked to explain why he sorted the way he did, the blocks are returned to their original starting positions, and the subject is asked to sort them again. This is continued until the subject meets a specified failure criterion; at which time the second, or "convergent phase", of the test is begun. In testing for concept recognition the examiner sorts the blocks in order of difficulty, according to the attributes not employed by the subject in the divergent phase. After each sort the examiner says to the subject, "Some people sort them this way; can you tell me why?".[6] This continues until all possible combinations of the blocks have been exausted.

Latency times and verbal responses are recorded on protocol sheets and the entire

[5]The most common object classification polarity in the literature is color-form. Neither color nor form is a very precise concept however. They are generic. Hue, shade, saturation and transparency (or translucency) may all be loosely subsumed under the more generic rubric of color. Similarly, a large number of attributes such as angularity, curvilinearity, perpendicularity of sides, proportions of one dimension to another and so forth, which may be considered formal attributes.

[6]The complete instructions for administration and scoring are attached as Appendix A.

dialogue transcribed on a tape recorder for subsequent verbalization scoring. Administration time for the entire test ranges from ten to twenty minutes depending on the age and perseverance of the subject.

SCORING

The OST procedure is a very "rich" procedure in that it can yield a large number of objective scores.

In the early stages of its development the OST was used with an extremely heterogeneous group of subjects to test its limits. Subjects ranged from early childhood to adult, gifted to retarded, and normal to emotionally disturbed. On the basis of the anecdotal information accrued during this period, 37 computationally independent scoring categories were defined. On the basis of pilot study results, 17 scoring dimensions were retained for futher investigation.

Data for 8 of these dimensions are reported in this paper.

CONSTRUCT VALIDITY

Analytic studies using a non-metric factoring procedure (SSAR-III) developed by Guttman and Lingoes (Lingoes, 1968) has generally confirmed the three supra-ordinate concept familes indicated earlier and has suggested reasonable reliability in the sorting procedure.

The construct validity of the OST also receives support from the analysis of intercorrelational data. When McQuitty's (1957) Elementary Linkage Analysis is applied to the correlation matrix of Table 1 below, three discrete clusters appear. The largest cluster includes the various divergent scores, non-verbal as well as verbal. The second cluster is based on concept recognition, or convergent process, scores, and the third cluster is based on erroneous sorting scores. These results are summarized in Figure 2.

Table 1

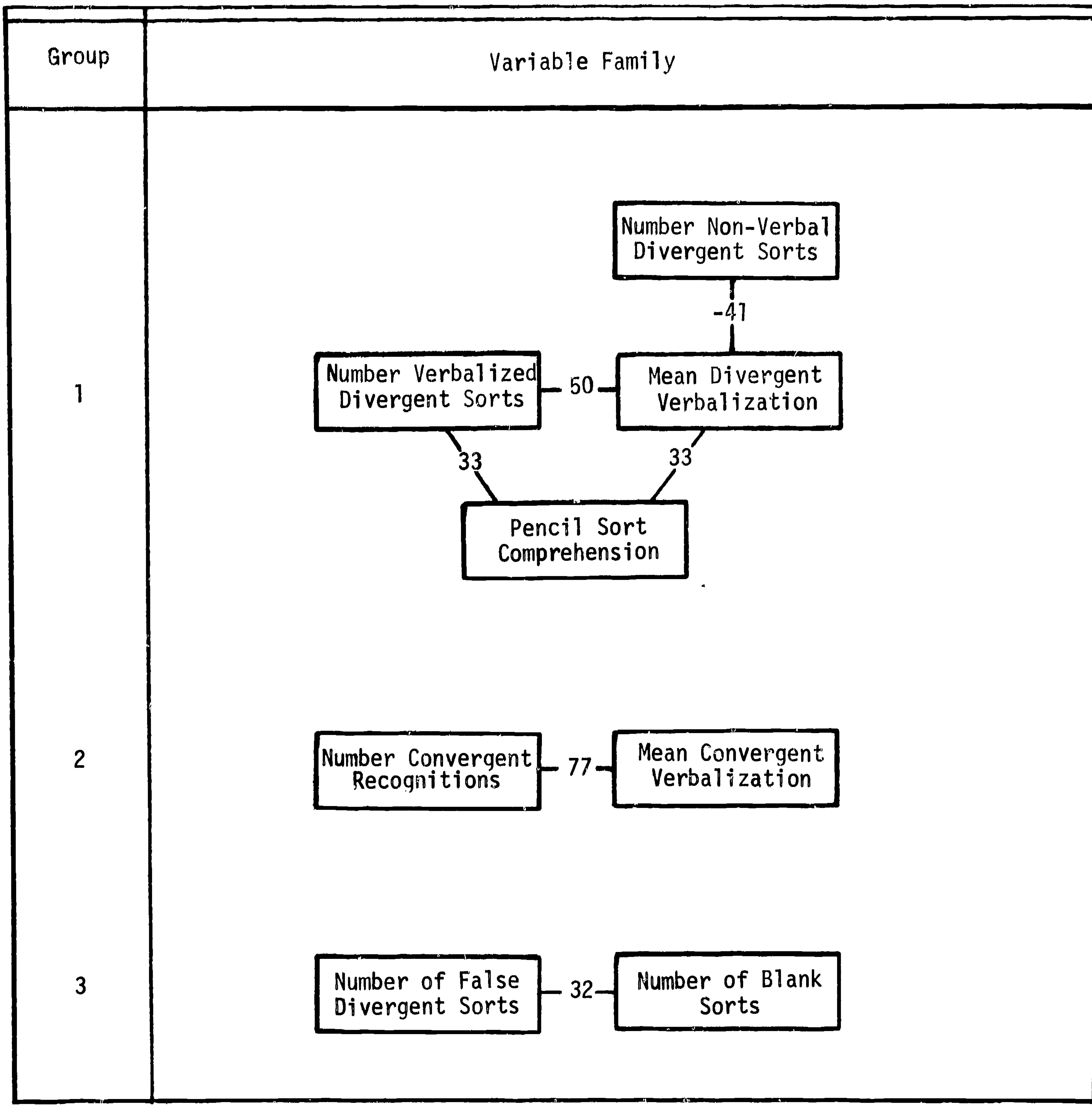
OST SCORES INTERCORRELATIONS

Variable	1	2	3	4	5	6	7
1. Pencil Sort Comprehension							
2. No. Verbalized Divergent Sorts	33**						
3. No. Non-Verbal Divergent Sorts	-25**	-25**					
4. No. False Divergent Sorts	-15*	-18**	-02				
5. No. Blank Sorts	-28**	-26**	28**	32**			
6. No. Convergent Recognitions	03	-19**	-13*	-18**	-17**		
7. Mean Divergent Verbalization	33**	50**	-41**	-06	-29**	01	
8. Mean Convergent Verbalization	05	-16**	-16**	-17**	-18**	02	77

* $p \leq .05$; ** $p \leq .01$

Figure 2

OST SCORE LINKAGE ANALYSIS



Note: Numbers represent the correlation coefficients linking the variables to their respective families.

PRELIMINARY STUDIES

Prior to undertaking a more extensive effort with the OST, three correlational and three experimental pilot studies were conducted.

Field Study 1 was a very brief study conducted on a small sample of 18 subjects. The purpose of the study was to test the OST procedure and obtain some preliminary notion of score relationships. The prototype blocks were made of wood, and to very gross dimension tolerances. In general, the test, in this early form, correlated rather highly with WISC Full Scale IQ. Number of positive sorts and total verbalization score correlated .77 and .79, respectively, with IQ.

Field Study 2 was designed to see whether or not there were gross sex and/or developmental differences in OST performance. Data were collected from 40 subjects: five boys and five girls from each of grades ? 6, 9, and 12. In general, there were no significant sex differences in OST performance but there were significant increases with age in scores on such variables as number of positive sorts, speed of first sort, and mean verbalization score.

Field Study 3 was the first study in which the current OST objects and procedure were used[7] (Safford, 1967; Safford and Dunn, 1967). The study was based on a stratified random sample of ten boys and ten girls from each of grades 1, 2, 3, 4, and 6. Clear-cut differences in OST performance due to age and IQ were obtained. A few very modest sex differences were also noted. In general, total number of positive sorts correlated .31 and .51 with Stanford-Binet IQ and WISC vocabulary scores respectively; and .28, .29, and .33 with Torrance flexibility, fluency, and originality, respectively.

The most interesting result was the strong relationship between the performance on the OST and academic achievement. The OST was a better predicter of academic achievement, for this sample, than either Weschler vocabulary or Stanford-Binet IQ scores. The probable reason for the lower IQ-Achievement correlation evidenced in this study is that the subjects were very bright, and in addition to attenuated heterogeneity, at the upper levels of intelligence, academic achievement may load more heavily on personal-social-motivational factors. Mean IQ of this sample was 126; SD was 19. Correlation between total number of positive sorts and achievement in the California Achievement Test Battery was .48. When mean speed and proportion of positive sorts were included, a multiple R of .55 was obtained.

The purpose of the experimental pilot studies was to investigate the conditions of administration on OST performance.

Experimental Study 1 was carried out by Robert Egri[8] on a sample of 50 elementary school children. The object was to see whether student set had an effect on OST performance. In the experimental condition, the OST was administered as a formal test with the expectation that test anxiety would be elicited and OST performance thereby impeded. In the control condition, the OST was administered in such a way as to minimize as much as possible any activation of student anxiety. There were no statistical differences in the performance of the two groups.

---

[7] The OST objects and procedures were revised on the basis of experience in Field Study 1 and again after experience with Field Study 2. The present objects were produced at the University of Michigan Psychology Department instrumentation shop.

[8] Unpublished doctoral student research project, The University of Michigan.

Experimental Study 2 was carried out by Roger Scott (1967) with a sample of 20 second grade students matched on IQ. The purpose of the study was to investigate the effects of data processing demands on the subject in the execution of OST sorts. Cahill and Hovland (1960) found the more memory required, the greater the difficulty subjects had in forming concepts. Data processing considerations also play an important role in the effectiveness of optimum strategies (Bruner, 1956). It was hypothesized that if the objects were presented sequentially instead of simultaneously, the successive presentation mode would minimize the amount of information with which the subject was dealing at any one time and OST performance would be increased. Significant results in the predicted direction were obtained.

Experimental Study 3 was conducted by Connie Bohannon[9] on a sample of 224 children. Because elementary school children operate primarily at Piaget's level of "concrete operations" and typically use what he has called "pre-concepts", i.e. concepts that exist only vis-a-vis the real object, physical exemplars were thought to be necessary for the OST procedure. Experimental Study 3 asked the question whether the blocks *per se* were necessary to the OST or whether photographs could be substituted. The subjects were 28 boys and girls drawn from each of grades 3, 4, 5, and 6. The experimental group sorted colored photographs; the control group sorted the blocks. In general, although older and brighter subjects were found to sort equally well regardless whether they had the physical objects or their photographs, there was sufficient argument to continue with the objects themselves.

RESULTS

SAMPLE

The sample on which the present norms for the OST were based was 21 boys and 21 girls drawn at random from grades kindergarten through six, in a middle-class Midwestern suburban school system. Total N was 294. Three examiners administered the OST. Each examiner tested seven males and seven females at each grade level. There were no significant differences in the test scores obtained by the different examiners.

Table 2

SAMPLE CHARACTERISTICS, BY AGE GROUP

Age Group Designation	Age in Months			Number of Subjects	CTMM IQ*	
	$\bar{X}$	SD	Range		$\bar{X}$	SD
5	62	2.6	54-66	32	-	-
6	73	3.6	67-78	40	-	-
7	84	3.3	79-90	43	109	8.99
8	98	3.6	91-102	38	106	9.63
9	109	3.2	103-114	35	106	10.73
10	121	3.5	115-126	39	105	12.30
11	132	3.4	127-138	42	103	11.23
12	141	2.1	139-150	14	101	13.62

* $\bar{X}$=105; SD=11.0 for total sample

[9]Unpublished doctoral student research project, The University of Michigan.

NORMS

Means and standard deviations of OST scores for the various age groups are summarized in Table 3. The last column of Table 3 also gives the probability values associated with Anova F-tests for each score. All scores from the divergent testing phase increased with age.

There was only one significant score difference attributable to sex; that was on number of convergent sorts recognized. Girls tended to recognize more convergent sorts than boys. Means and standard deviations were .4 and .74 respectively for the boys as contrasted to .7 and .88 for the girls. The p value was ≤.01. In addition this score does not appear to be age related.

Table 3

NORMS FOR OST SCORES

Variable		Age Group 5	6	7	8	9	10	11	12	p*
Pencil Sort Comprehension	$\bar{X}$	.3	.5	.9	.8	.8	1.0	1.0	1.0	≤.001
	SD	.47	.51	.52	.37	.38	0.0	.15	0.0	
No. Verbalized Diverg. Sorts	$\bar{X}$	1.1	1.6	2.2	2.7	2.9	3.6	3.6	3.5	≤.001
	SD	1.29	1.12	1.44	1.46	1.45	1.52	1.43	1.60	
No. Non-Verbal Diverg. Sorts	$\bar{X}$	1.1	.6	.2	.2	.1	0.0	0.0	0.0	≤.001
	SD	1.60	1.41	.72	.55	.37	0.0	0.0	0.0	
No. False Diverg. Sorts	$\bar{X}$	1.0	1.0	.8	.7	.7	.3	.4	.1	≤.05
	SD	1.26	1.62	1.40	1.31	1.14	.80	.96	.36	
No. Blank Sorts	$\bar{X}$	1.9	.8	.6	.4	.5	.1	.1	0.0	≤.001
	SD	1.68	1.44	1.07	.86	1.07	.41	.26	0.0	
No. Converg. Recognitions	$\bar{X}$	.5	.5	.6	.4	.6	.9	.8	.5	NS
	SD	.62	.72	.85	.55	.88	.98	.99	.94	

*1-way Anova F-test

Figures 3, 4, and 5 abstract and/or combine selected information from Table 3. Figure 3 reflects the age patterning of valid and invalid divergent sorts. A valid divergent sort may be either a verbalized or non-verbalized sort. An invalid sort is either a "blank" sort or a "false" sort, i.e. a sort predicated on a false verbal basis.

As one would expect, the number of valid sorts goes up with age and the number of invalid sorts goes down with age.

If one examines the components of these valid and invalid scores, interesting patterns obtain. For instance, the frequency of occurrence of random sorting behavior, i.e., the number of "blank" sorts, drops drastically between age 5 and 6. See Figure 4. For all age levels except the earliest, the frequency of this form of classificatory behavior is slightly less than for fallaciously justified classifications. Zona Hansen, in her doctoral thesis (in progress for the University of Michigan) has been working to extend these norms downward to 3 and 4 year olds.

Figure 3

VALID AND INVALID DIVERGENT SORTS

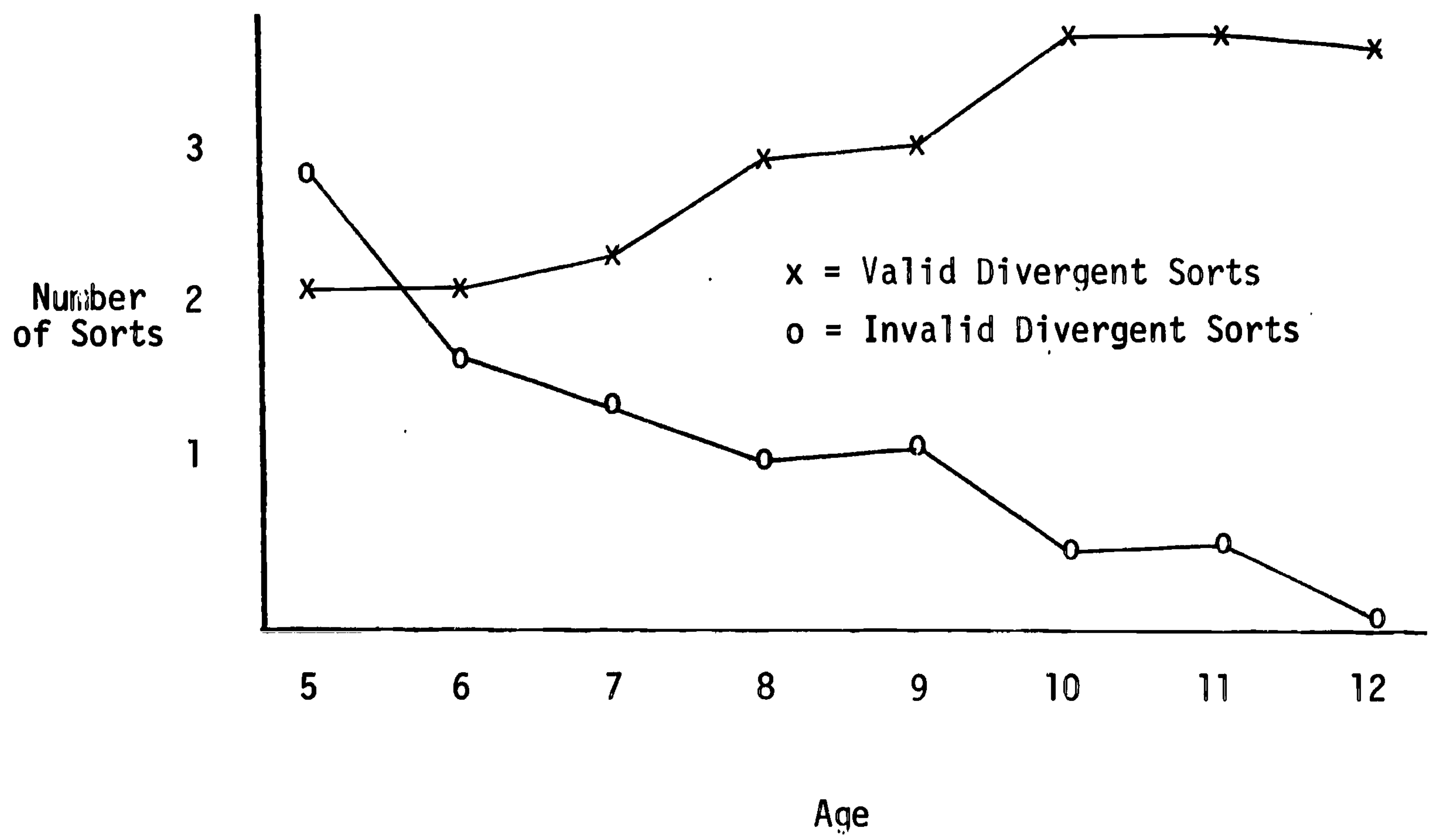


Figure 4

INVALID DIVERGENT SORT COMPONENTS

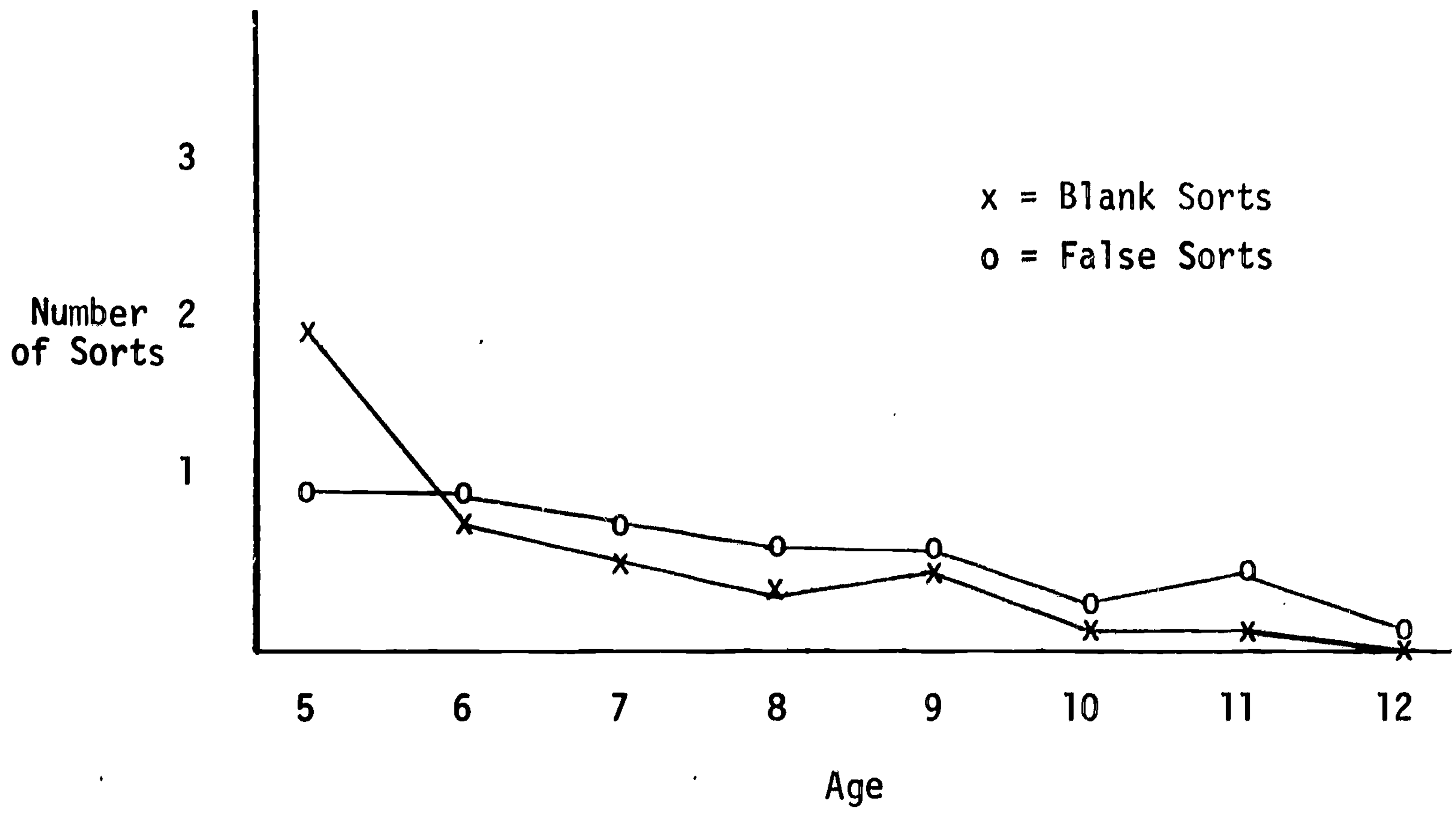
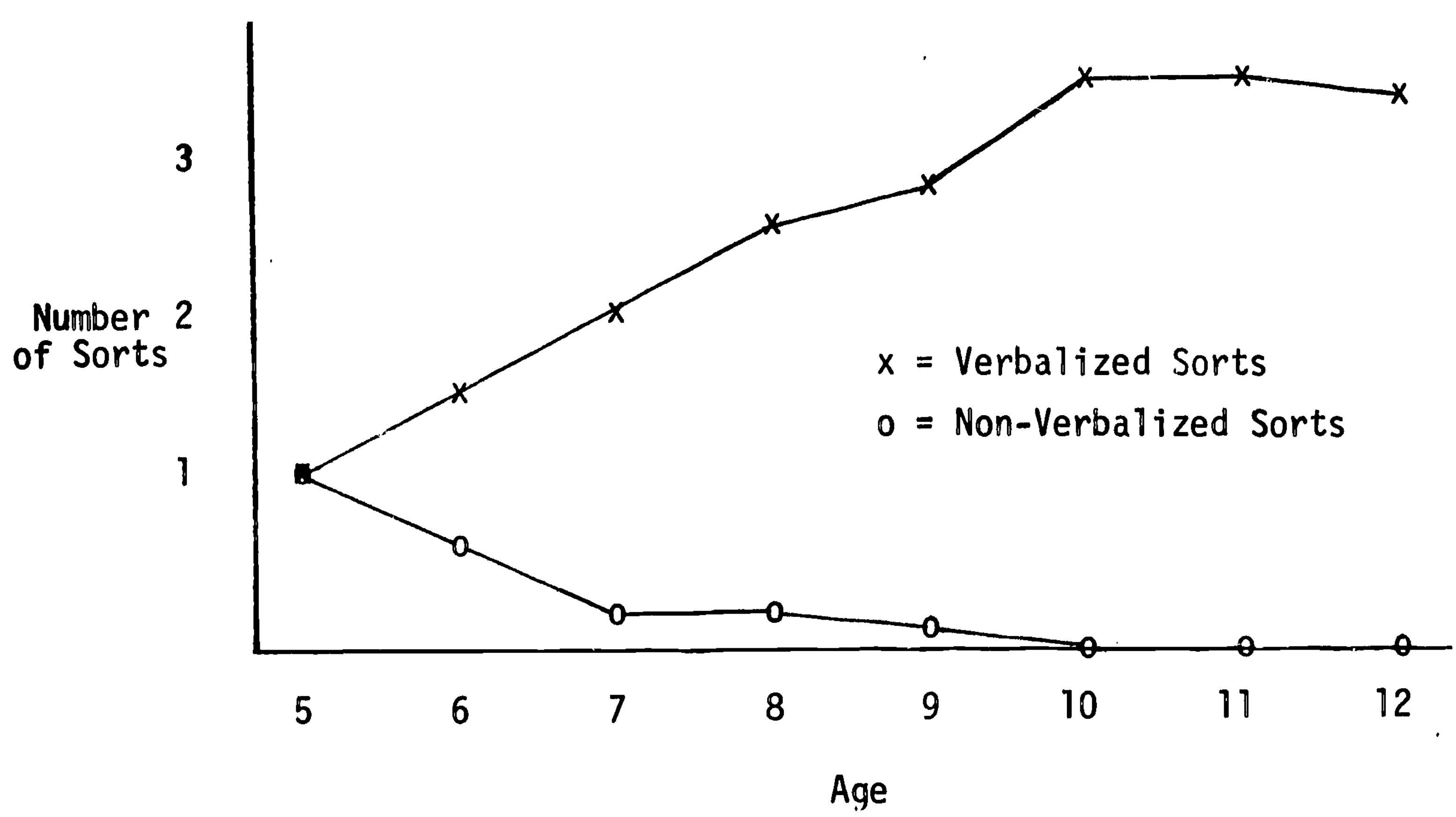


Figure 5

VALID DIVERGENT SORT COMPONENTS



Of special interest are the differences in frequency of occurrence of valid divergent sorts. Verbalized convergent sorts increase significantly with age and non-verbalized sorts decrease significantly with age.

At age 5 the concept repertoire subjects apparently had available for classificatory behavior, were equally balanced between verbalizable and non-verbalizable concepts. With increasing age the number of concepts available to the child not only increased but the proportion of verbal to non-verbal concepts also increased. See Figure 5.

This is not surprising. It is exactly what one would expect. It is however, encouraging to corroborate such expectations with empirical evidence.

It is also interesting to note that although the number of blank and non-verbal sorts correlates .28, the mean frequency of non-verbal divergent sorts is lower than the mean frequency of blank sorts at every age level, even though the mathematical probability of the former is 9 times that of the latter. This tends to add confidence that correct classfication of the blocks, even though verbal justification cannot be offered by the subject, is not just random block grouping.

PREDICTIVE VALIDITIES

Table 4 summarizes the correlations of various OST scores with age, sex, IQ, academic achievement, and Torrance creativity scores. By way of auxiliary information, Table 5 presents IQ, achievement, and Torrance score intercorrelations.

Not surprisingly, chronological age correlates with increasing conceptual sophistication, increasing academic achievement and the like.

Table 4

OST SCORE - CRITERION MEASURE INTERCORRELATIONS

Criterion Variable	OST Variable 1	2	3	4	5	6	7	8
9. Age	44**	53**	-34**	-21**	-40**	13*	56**	10
10. Sex	-00	-02	-06	-09	00	19**	-01	21**
11. CTMM Full Scale IQ	06	10	04	-24**	-15*	13	06	13
12. Language Achievement[1]	22**	28**	-21**	-23**	-19*	28**	27**	19*
13. Arithmetic Achievement[1]	26**	27**	-17*	-25**	-18*	20**	29**	17*
14. Total Achievement[1]	27**	31**	-22**	-27**	-21**	25**	28**	16*
15. Torrance Fluency	06	16	-03	-08	00	-14	16	-13
16. Torrance Flexibility	03	21*	01	-07	-01	-13	13	-03
17. Torrance Originality	04	16	-07	-14	-04	-03	15	-00

* $p \leq .05$; ** $p \leq .01$; [1]Iowa Tests of Basic Skills.

Table 5

IQ, ACHIEVEMENT AND TORRANCE SCORE INTERCORRELATIONS

Variable	9	10	11	12	13	14	15	16
9. Age								
10. Sex	-03							
11. CTMM Full Scale IQ	-21**	16*						
12. Language Achievement	61**	23**	39**					
13. Arithmetic Achievement	67**	04	33**	84**				
14. Total Achievement	68**	14	41**	94**	92**			
15. Torrance Fluency	21*	24*	25*	40**	36**	35**		
16. Torrance Flexibility	26**	30**	26**	44**	39**	41**	80**	
17. Torrance Originality	35**	22**	22**	49**	46**	46**	74**	77**

* p≤05; ** p≤01

The more important finding, from the point of view of the OST, is the significant correlation of OST scores with academic achievement, the non-correlation of OST scores with IQ (except for the "blank" and "false" scores which correlated negatively) and the non-correlation of OST scores with Torrance scores.

Quite clearly the OST predicts variance in academic achievement that is not accounted for by CTMM IQ scores or by Torrance scores.

Similarly it is interesting to note that, unlike the Torrance scales, language achievement, and to a small degree even IQ, which correlate with sex (girls do better on all of these), there is no sex correlation with most OST scores. Whatever sex bias is inherent in CTMM IQ, Torrance, and language achievement, is not present in the divergent thinking scores of the OST.

This singular correlation of the OST scores with academic achievement adds additional credence to the belief that the OST is taping a novel, or at least non-common, sector of the cognitive functioning domain.

Table 6 shows that for all practical purposes OST-Achievement correlations are not attenuated when the effects of IQ are partialled out. This is not surprizing in view of the low OST-IQ intercorrelations.

If a multiple R is computed using the two basic scores, number of verbalized divergent sorts and number of convergent recognitions, the OST-Achievement correlation increases to .44, a correlation even greater than that between CTMM IQ and achievement.

Table 6

OST PREDICTION OF ACADEMIC ACHIEVEMENT

OST Variable	OST-Academic Achievement Correlations	
	OST Only	IQ Partialled Out
1. Pencil Sort Comprehension	27	27
2. No. Verbalized Divergent Sorts	31	30
3. No. Non-Verbal Divergent Sorts	-22	-26
4. No. False Divergent Sorts	-27	-20
5. No. Blank Sorts	-21	-17
6. No. Convergent Recognitions	25	22
7. Mean Divergent Verbalization	28	28
8. Mean Convergent Verbalization	16	12

CONCLUSION

In summary it would seem the OST is a technique quite different from anything reported elsewhere in the literature. It has reasonable face and construct validity; and yields scores that make a unique contribution to achievement score variance. The OST scores also reveal clear cut developmental differences that are at the same time both understandable and compatible with current theories of cognitive development. Because of its simplicity, and its inherent and non-threatening appeal to children, there is the distinct possibility it may find its way into applied diagnostic use in clinical assessment areas. Its heuristic character has already been demonstrated and its research possibilities are being reported in this symposium.

By way of conclusion I should like to mention that the work my colleagues and I report today is but a portion of the effort involving the OST procedure. Barbara Landau has completed a major replication study and has extended the OST norms upward to high school students, and Zona Hansen is involved in an experimental study of the trainability of reversability and multiple classification skills with pre-school age children.

APPENDIX A

OST PROCEDURAL GUIDE

PART I--ADMINISTRATION

The administration mat is on the table. Place stopwatch on table top. Pencils and OST blocks are not yet in subject's sight. Proceed as follows, say:

I want to see how many ways you can think of to put things into two groups. All of the things in a group must be the same in some way. I will show you what I mean.

(Place pencils between the circles on the administration board with the eraser ends toward the examiner.)

See these pencils? They are all mixed up. *(mix up)*

Now suppose I asked you to put three pencils in this circle *(point to examiner's right)* and three pencils over here in this circle *(point left)* so that all of the pencils in this circle *(point right)* are all the same in some way and all of the pencils over here *(point left)* are the same in some way. You could do it like this. *Sort pencils--new pencils to circle on examiner's right.*

See these pencils? *(point right)* They are all the same because they are all new. They have never been sharpened. *(point toward unsharpened end of pencils)*

These pencils are the same *(Point left)* because they are not new. They have been sharpened *(point to pencil points)*.

Now, I'll mix them up again *(mix up and put between circles again)*.

Suppose I asked you to sort the pencils <u>again</u>, except this time in a <u>different</u> way. You could do it like this. *(Sort pencils; pencils with erasers to circle on examiner's right)*.

See, these are the same because they all have erasers *(point right to erasers)* and these are the same because they don't have erasers *(point left to ends of pencils where erasers are missing)*.

Always sort them into two groups. Always three things here *(point right)* and three things here *(point left)*.

Now, can you see still <u>another</u> way to divide up the pencils into two groups? A way that we haven't tried yet?

(If successful, say:) Fine. Why did you put them that way? That's right. These are the same because they all have pocket clips *(point to clips)*, and these are all the same because they don't have pocket clips *(point to area of pencils where clips are missing)*.

If the child is unsuccessful, sort the pencils with the points toward him, clips in circle to examiner's right, and say: See, here is another way. These are the same because they all have pocket clips *(point right to clips)* and these are all the same because they don't have pocket clips *(point left).*

Do you get the idea? Do you understand what it is we are going to do? *If the child says, "No," paraphrase the instructions again. If the child says "Yes" say:*

O.K. then we can start, except this time, instead of pencils, *(remove pencils from sight),* we will use these blocks.

(Take blocks from storage area and place on the X's on the administration board.)

Can you see a way to divide up these blocks so that the three you put here *(point right)* are all alike and the three you put here *(point left)* are all alike? Go ahead and try it.

Why did you put them like that?

(Write response, then return blocks to the X's and say:)

O.K. Now let's see if you can find <u>another</u> way to sort the blocks.

(Indicate for the child to try again.)

Why did you sort them that way?

(Write response, then return blocks to X's).

Can you see still another way to do it?

(Indicate child should sort the blocks again) etc.

(If the child asks, "is this a test?" reply, No. It is not a test but it is kind of like one. It isn't part of the regular school work though, and you won't be graded on it. We are seeing lots of different children because we are interested in learning how well they can do some special things for us.

Timing should be from the behavorial indication for the child to go ahead and the behavioral indication when he has finished. If he looks like he is finished and then decides to work on it some more, the time continues to run until he is definitely through.

The test is terminated in the following ways:

a. *After 9 duplicate sorts.*
b. *After the subject has had 120 seconds but can produce no sort, or*
c. *After the subject has indicated he can do no more. If he says, "that's all," tell him* You have lots of time--can you see any more? *This is only indicated to him once, however.*
d. *After five perseverative duplicate sorts.*
e. *After three misses on the convergent recognition phase.*

PART II --VARIABLE DEFINITIONS

Variable

1. *Pencil Sort Comprehension:*
   1 = child got clip sort on pencils.
   0 = child missed clip sort when first administered.

2. *Number of Positive Sorts-With Verbalization:*
   This is the sum of positive arrangements of blocks for which verbal explanations have been offered. The child may not have been able to verbalize his sorting rationale entirely adequately, however.

3. *Number of Positive Sorts-Without Verbalization:*
   For this index the child has an appropriate arrangement of blocks, but does not have the capacity to mount even weak verbal justification for his arrangement. If the child is asked why he arranged the block in the way he did, he may often respond with "I don't know" or give some vague, ambiguous, seemingly random, or unintelligible verbalization. It is presumed that the positive sort has been achieved via preverbal or nonverbal representation systems.

4. *Number of False Positive Sorts:*
   A false positive sort is a positive block arrangement not previously used, given with a well delineated but unacceptable verbalization so that the examiner may safely conclude the subject based his sorting behavior on inappropriate conceptulization. The most common false justification is a 2-1, 1-2 basis for sorting. But external relations and imputed meaning may also be involved.

5. *Number of Times a Blank Sort Appears:*
   Code 0 if blank sort never given. Code 1 if given once, etc.

6. *Number of Convergent Positive Sorts:*
   Number of convergent positive sorts adequately explained.

7. *Mean Verbalization Score for Positive Verbal Sorts:*
   This is arrived at by adding the verbal score for the two poles comprising a sort. See OST Procedural Guide--PART III.

8. *Mean Verbalization Score--Convergent Sorts:*
   See verbalization scoring procedures. This is based on the sum of both poles.

9. *Mean Verbalization Discrepancy Score for Verbal Positive Sorts:*
   This score is arrived at by taking the absolute difference between the scores of the two poles of the classification dimension, summing, and dividing by the number of scored sorts.

10. *Mean Speed of Positive Sorts:*
    This is the mean time, computed in seconds, for a child to complete a positive sort, either verbal or nonverbal.

Variable

11. <u>*Flexibility Score on the OST:*</u>
The different sorts can be grouped into several common categories: form, color, and relational sorts. The flexibility score is the number of shifts from one category to another. Only positive sorts are taken into consideration. The presence of any number of duplicate sorts between two positive sorts of two different classes does not nullify the shift.

12. <u>*Number of Duplicate Sorts:*</u>
This index is the number of times a child repeats a <u>valid</u>, i.e., positive, sort. The verbalization for the duplicate sort need not be similar to that previously given, however, nor even of the same class. A duplicate sort may even involve erroneous verbalizations.

13. <u>*Number of Internal Relational Sorts:*</u>
Here the composition of the groups rather than the characteristics of each block form the basis of the sort. For example, "This group has two red and one blue and this group has two blue and one red."

14. <u>*Number of External Relational Sorts:*</u>
The blocks take on meaning by their relationship to some external reference. For example, "These are on my left and these are on my right."

15. <u>*Number of "Sign" Sorts:*</u>
The number of times where meaning is imputed to the blocks. Examples: "These look like lamp shades," "These are bathtubs," etc. (Divergent sorts only.)

16. <u>*Number of Primitive Functional Sorts:*</u>
The number of times where there is an animistic quality in the verbalizations which implicitly attributes intentionality to the object. This should not be confused with a sophisticated description of function where the locus of intentionality lies clearly within the individual (e.g., "You can see through these.") Examples of primitive functional sorting verbalizations are: "These can cut," "These can hurt you," etc. (Divergent sorts only.)

PART III--VERBALIZATION SCORING

All positive sorts, both divergent and convergent are scored for verbalization. Duplicates, blank and false positive sorts are not scored. Both "poles" of the classification dimension are scored. That is, both the explanation of how all the blocks in the left set, as well as the counterpart explanation of how the blocks in the right set are alike, are scored. The scoring rationale is as follows:

"0" is given for motoric gestures only; for an "I don't know" response, and for verbalizations that appear to be given just to satisfy the adult demand for an explanation. (This type of acquisition is usually found only with very young children.)

"1" is recorded if the explanation can be justified, but only by means of considerable extrapolation on the part of the examiner.

"2" is given when the subject demonstrates an implicit recognition of the correct attribute. In this case he apparently has the correct idea with regard to the attribute; however, the conceptualization of that attribute is so poorly organized that the explanation often includes erroneous and/or inaccurate statements. That is, his verbal justification of his sort is arrived at by attempting to force on the object certain attributes that they don't, in fact, have.

"3" is recorded when the subject explicitly states the correct attribute. Here the verbalization consists essentially of detailed descriptions of the correct attribute and/or qualifications of "1" point responses, such as changing an adjective like "round" to a modifier such as "roundish'" this eliminates erroneous assertions about the blocks. The responses must be correct but not necessarily comprehensive.

"4" is recorded for a precise, and concise, statement of the significant attribute. A "4" response is a high order abstraction whereas the "3" point response is essentially descriptive and concrete in nature. The availability of such a precise statement suggests that the concept is clearly defined in the child's mind and is available for use on demand.

Examples:

<u>Sort #1--Curvilinearity</u>

0--no answer, motoric gesture, "I don't know," "these match with each other," "go together."

1--"not round," "not square."

2--correct idea but very poorly stated, often including an erroneous or inaccurate statement, such as "round corners," "these are all round," "these have sharp edges," "these are circles," "these are squares," "these have pointed ends," "these are like cubes," "these have round edges," "these are ellipses and circles."

3--*an attempt at a qualified statement. Example: roundish, circular, rounded off, rounded lines, oval shaped, round in some places, squarish. Sometimes a child might also focus on some subset of attributes associated with the more general concept being sought. For example, square edges, straight lines, all sides are flat, curved sides, 8 points, almost round, almost square.*

4--*a 4 point response is a precise statement of the significant attribute. Example: angular, rectalinear, curvilinear, these are straight and these are curved-lined objects.*

Sort #2--(Red-Blue)

0--*no answer, "I don't know," "these match with each other, go together."*

1--*"not blue," "not red."*

2--*red, pink, blue; these are red and pink and orange and these are blue; light colored--dark colors; these are almost red; these are almost blue.*

3--*these are all colors of red, they're reddish, pinkish, bluish, blue colored, they're all different kinds of blue, they're reddish in tint, these are a blue hue.*

4--*these are shades of red, blue hue(s) of blue, these are in the red family.*

Sort #3--Transparency

0--*no answer, "I don't know," "these match with each other, " "go together," "these are pretty."*

1--*"these are not clear, these are blocks."*

2--*implied optical properties: "you can see scratches on the other side," "it sparkles," "it's shiny," "translucent," "it's made out of plastic," "glass," these are painted and these are not, these are wood, light colors, dark colors, solid colors.*

3--*a functional optical property: "you can see through these," "it transmits light;" "it magnifies things;" "it catches or filters light (clear objects); "it reflects light" (opaque objects).*

4--*clear, transpartent, opaque.*

Sort #4--Height

0--*"round shapes" (when it occurs in conjunction with "talls"); no answer, "I don't know," "these match with each other, go together."*

1--*"not tall, "not short."*

2--*"big," "little;" "small," "these are level," "these are all the same size."*

3--"these have long lines;" "these have short lines;" "these are high and these are low;" "these are flat."

4--tall, short.

Sort #5--Perpendicularity of Side

0--no answer, etc.

1--these are fatter than those; these are narrower than those; these have flat edges; all the bottoms are even, these are triangles, these are not straight.

2--these curve out and these go straight; these get width as they go down; these are shaped like a pyramid; these go out; these go out at the bottom; these tilt down; these are cone shaped; these are bigger at the bottom than at the top; these are bigger at the top than at the bottom; these come up in a triangular shape.

3--the top edges are the same size as the bottom edges; these are shaped like a triangle with the top cut off; these are shaped like a pyramid with the top cut off; sort of like a pyramid; these are straight on two sides; these go straight up; these have slanting lines; these have slanting edges.

4--a 4 response indicates a recognition that it is the sides themselves that are slanting, not some particular aspect of the side such as an edge. These have slanting sides; sloping sides, beveled sides; these sides are slanting; these sides are not perpendicular; these sides go straight up.

Sort #6--Radial Equality of the Base

0--no answer, etc.

1--these are thinner; these are fatter.

2--the implicit awareness of elongation but a very awkward or erroneous way of stating it. For example, these are long and these are not; all the tops are even (referring to the regular objects); all the tops are equal (implying length and width of the tops are equal for each top); these are all like cubes (referring to the regular objects).

3--the explicit notion of proportionality, i.e., comparing length to width, but stated in a roundabout way. These are longer than they are wide. These are skinny compared with their size. These are longish; these figures are all like circles, they approach a circle in shape, (all referring to regulars).

4--the verbalization should indicate the presence of a well and clearly defined verbal label for the criterial attributes elongated, oblong, regular.

Sort #7--Saturation:

0--no answer, etc.

1--"baby colors and big boy colors," "these are soft and these are hard."

2--"light colors, " "dark colors," "bright and dark colors."

3--"these have light shades of red and blue and these have dark shades of red and blue."

4--a statement with regard to the fact that these are pastels and these are not; or these are pastels and these are intense colors.

Sort #8--Volume:

0--these are taller

1--"these are more than those"

2--big, little, small, large, heavier, lighter, thin, thick, these are bigger than these.

3--larger in size, smaller in size; these would hold more water than these if they were empty; it would take more material to make these than these, these wiegh more.

4--the volume of this group is larger than the volume of this group; these have large volumes, etc.

Sort #9--Area of the Base:

0--no answer, etc.

1--"these would be good pounders."

2--an implied recognition of area differences but very poor specifications, such as, these are fatter than these; these are skinnier than these; these are thinner than these; these are big and these are little; these are bigger than these.

3--these cover up more of the table top than these; it would take more paint to paint the bottoms of these; these bottoms are bigger than these; these have big bottoms and these have little bottoms.

4--a statement to the effect that this group has a larger basal area than the other group. The base of these are larger than the base area of these.

REFERENCES

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Bruner, J. S. The course of cognitive growth. *American Psychologist,* 1964, 19, 1-15.

Bruner, J. S. *Toward a theory of instruction.* Cambridge: Harvard University Press, 1966.

Cahill, H. E., & Hovland, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology,* 1960, 59, 137-144.

Guilford, J. P. Three faces of intellect. *The American Psychologist,* 1959, 14, 469-479.

Guilford, J. P. *The nature of human intelligence.* New York: McGraw-Hill, 1967.

Heidbreder, E. The attainment of concepts: I. Terminology and methodology. *Journal of General Psychology,* 1946, 35, 173-89.

Hilgard, E. R. *Introduction to psychology.* (4th ed.) New York: Harcourt, Brace & World, 1967.Hull, C. L.

Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychological Monographs,* 1920, No. 123.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.

Kerlinger, F. N. *Foundations of behavioral research.* New York: Holt, Rinehart & Winston, 1964.

Lingoes, J. C. The rationale of the Guttman-Lingoes nonmetric series: A letter to Doctor Philip Runkel. *Multivariate Behavioral Research,* 1968, 3, 495-507.

McQuitty, L. L. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational Psychological Measurement,* 1957, 17, 207-229.

Morgan, A. B. Sex differences in adults on a test of logical reasoning. *Psychological Reports,* 1956, 2, 227-230.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford University Press, 1953.

Piaget, J. *The psychology of intelligence.* New York: Harcourt, Brace & World, 1950.

Piaget, J. *Origins of intelligence.* New York: International Universities Press, 1964.

Riess, B. F. Semantic conditioning involving the galvanic skin reflex. *Journal of Experimental Psychology*. 1940, 26, 238-40.

Riess, B. F. Genetic changes in semantic conditioning. *Journal of Experimental Psychology*, 1946, 36, 143-52.

Ruch, F. L. *Psychology and life*. New York: Scott Foresman, 1948.

Safford, P. L. The prediction of academic performance from an object-sorting concept formation-recognition task. *American Psychologist*, 1967, 22, 546.

Safford, P. L., & Dunn, J. A. Age, sex, and IQ differences in performance on a perceptual-cognitive task. *American Psychologist*, 1967, 22, 726.

Scott, R. O. Decrement in children's cognitive performance as a consequence of information overload. In J. A. Dunn (Ed.), *School anxiety and cognitive functioning: Exploratory Studies*. Report 4, IRCOPPS Midwest Research Center for Pupil Personnel Services. Ann Arbor: University of Michigan, 1967.

Simon, J. C. & Newell, A. S. Models: Their uses and limitations. In L. D. White (Ed.), *The state of the social sciences*. Chicago: University of Chicago Press, 1956. Pp. 67-83