

AN OBJECTIVE TECHNIQUE FOR THE MEASUREMENT
OF FLEXIBILITY IN THINKING

by
Esther Berg

A Thesis Submitted for the Degree of
MASTER OF ARTS
(Psychology)

UNIVERSITY OF WISCONSIN

1 9 4 6

AWM
B452
E385

608513
APR 2 1957

INTRODUCTION

The phenomena of "abstract behavior" and "shift of set" has recently been the subject for much discussion. The impairment of the ability to respond selectively to one quality of a situation or object and the lack of the ability to shift from one aspect of a situation to another is well known in psychopathology.

A number of investigators have made attempts to measure these phenomena. Most of the tests have been sorting tests, where the subject had to classify for color, form or some category inherent in the materials. (1, 2, 3, 9, 10) The extreme difficulty in obtaining quantitative measurements have forced these experimenters to make qualitative analyses of their results. Goldstein and Scheerer are quite right in emphasizing the fact that when dealing with an individual who is mentally ill it must be remembered that he is living in an environment which is no longer adequate for him and that he cannot respond to the same stimuli as the normal. They assert that the problem may appear to be solved quantitatively but the patient has not solved the task in the same way that it has normally been solved. "The tests of common usage apply a quantitative scoring in which success and failure are recorded in an arithmetical way as plus and minus. In our tests, this method is applicable only to a very limited extent." (1, p.14)

This presents an interesting task for the psychologist. Can some method be found that will lend itself to both quantitative and qualitative analyses?

Hanfmann and Kasanin attempted to solve this problem by devising a scoring system for the Vigotsky test. Their scoring, however, is tinged with subjective analysis. They say: "If in any of the items enumerated the subject at first showed an inferior performance, but later improved it - either spontaneously or with the examiner's help - it was his highest performance that was scored." (3, p.54) Oftimes, the subject's system of classification depends upon the subjective analysis of the observer. What may seem to be classification for color may not be that at all, even among normal subjects and their verbal report is not always enlightening. It then remains in the hands of the observer to determine the subject's system of classification.

The problem of "shift of set" is intimately tied with abstraction. It seems that people in whom categorical behavior is impaired, lack facility in shifting, while children, who have not developed the abstract attitude, react in a changing lively manner. (10) Goldstein and Scheerer assume that the inability to shift is one of the criteria for concreteness. "We have attempted to present some criteria for concreteness. All of these can be manifest in the normal as well as the abnormal individual. These are criteria for pathological concreteness only insofar as the subject encounters insurmountable difficult-

ies in volitionally performing a shift either within the concrete realm or transcending from it to the conceptual." (1, p.101) Lack of flexibility is an indicator of rigidity and concreteness.

The purpose of this experiment was to devise a test which included both "abstraction" and "shift of set" that could be quantitatively and objectively scored.

Materials

A pack of sixty (60) cards were devised so that each included three categories: color, form and number. The colors were red, green, yellow and blue. The forms were triangles, five point stars, crosses and circles. The numbers were 1, 2, 3, and 4. A single card could then have four green triangles, or two yellow stars or any combination of the three categories. In addition to the pack of sixty cards there were four stimulus cards: first, one red triangle, second, two green stars, third, three yellow crosses, fourth, four blue circles. Reproductions of these stimulus cards appeared in the pack of sixty. There were also three duplications in the pack. (Reasons for this will be discussed later.)

Procedure

Nine combinations of color, form and number were worked out for each subject so that each subject received a different key. There were 96 different keys so that no

subject received the same key. For example:¹ The key for one subject would be:

C N F N F C F C N

A second subject could have:

N C F C N F N C F

The four stimulus cards were placed on the table before the subject and he was given the pack of sixty cards. The instructions were as follows: "I want you to put these cards (the pack of sixty) into four groups, underneath the ones lying on the table. I will tell you whether you are right or wrong." If the subject asks if he is to classify according to color, form or number he is merely told, "I will tell you whether you are right or wrong." The experimenter had the key before him, out of sight from the subject. The subject then begins to classify the cards. If the first category according to his key, is for color, the subject is told "wrong" until he sorts for color. As soon as he makes² five correct consecutive responses in one category the experimenter shifts to the next category without telling the subject. If the next category is number, he is told "wrong" until he classifies for number and then after five consecutive correct responses in the

-
1. A table of random numbers was used to assign each subject key.
 2. The criterion of five correct consecutive responses was not chosen arbitrarily. It was found experimentally that three responses were not sufficient to establish the category and ten were more than necessary. Five correct consecutive responses were sufficient to establish the category.

number category he is forced to shift again according to the key. This procedure is followed until the subject completes nine shifts. If he uses the whole pack before the nine shifts are completed it is reshuffled and handed back to him. The experiment is considered completed after the subject has made nine shifts, regardless of the number of cards used. When the experiment was over, the subject was asked: "What do you think is the purpose of this experiment?" "What were you trying to do?"

Subjects

There were fifty-one subjects, nineteen men and thirty-two women. One male was negro, one female was chinese. All were students in an elementary psychology class at the University of Wisconsin. The age range was 17-28. The mean age was 20.4 years.

Results

On the basis of the results the data can be divided into three groups. Group A: Those students who went through nine shifts successfully and who early in the experiment discovered that the experimenter was shifting. Group B: Those students who went through nine shifts successfully and had a vague notion that the experimenter was shifting but could not state precisely what the shift entailed or when it occurred. Group C: Those students who could not successfully complete nine shifts and had no knowledge

whatever that there was a shift. Finally, Total Group: A combination of the data of Group A and Group B, those students who successfully completed nine shifts.

The data is based on the total number of errors, the number of errors for each shift and the total number of cards used to complete the experiment.

Quantitative Results

A glance at table 1 shows statistically significant differences for groups A, B, and C. All these differences are significant below the 1% level of confidence for both the average number of errors and the average number of cards used. These groups are characterized by a difference in understanding of the problem and ability to verbalize the process.

TABLE 1
DIFFERENCES BETWEEN GROUPS

ERRORS					
GROUPS	MEANS	NO.	D.f.	"t"	LEVEL OF CONFIDENCE
	<u>36.33</u>	<u>15</u>			
A and B	72.10	21	34	3.834	less than 1%
	<u>36.33</u>	<u>15</u>			
A and C	120.87	15	28	7.250	less than 1%
	<u>72.10</u>	<u>21</u>			
B and C	120.87	15	34	4.858	less than 1%
CARDS					
	<u>95.80</u>	<u>15</u>			
A and B	156.95	21	34	4.297	less than 1%
	<u>95.80</u>	<u>15</u>			
A and C	223.67	15	28	7.277	less than 1%
	<u>156.95</u>	<u>21</u>			
B and C	223.67	15	34	3.568	less than 1%

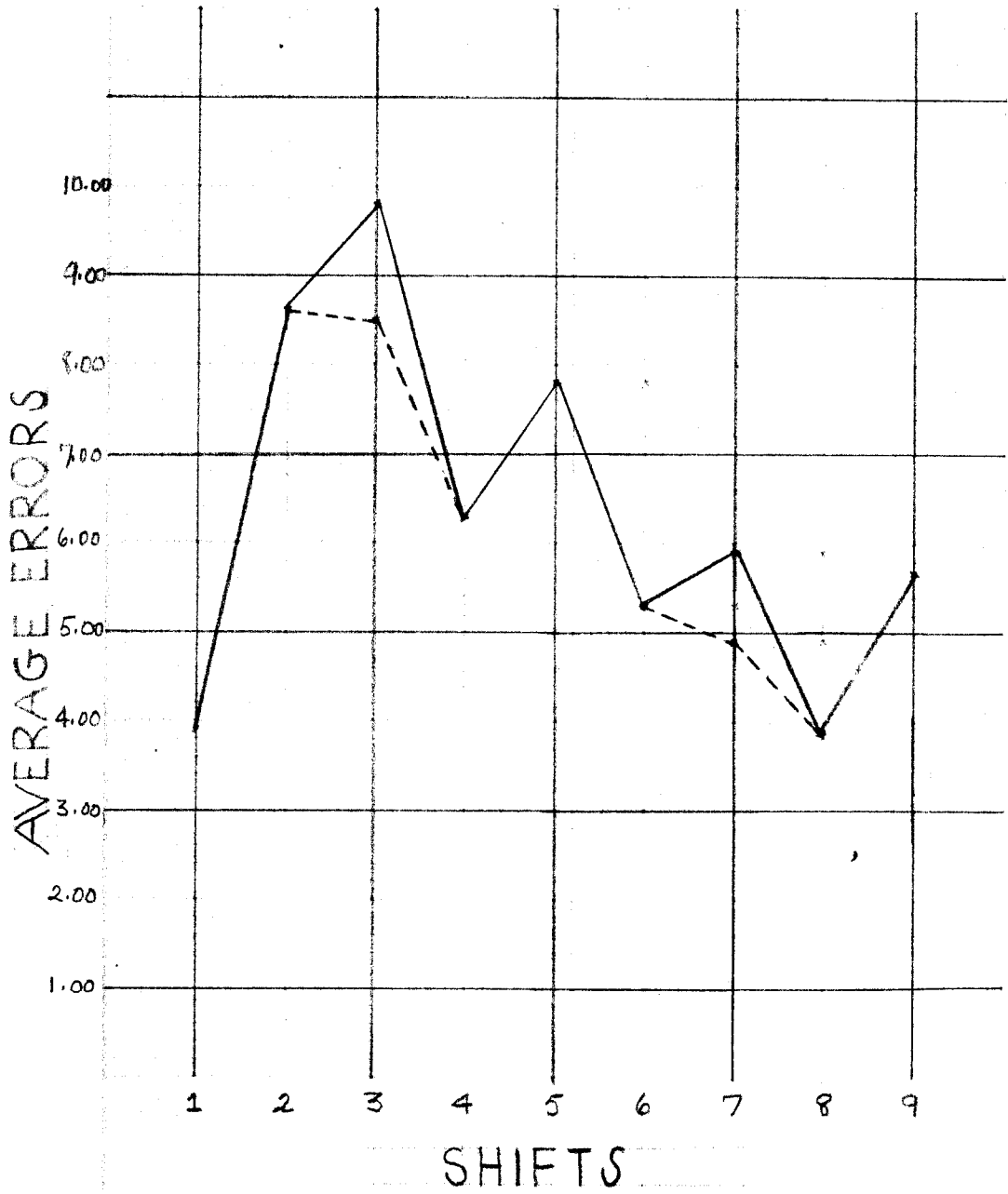
Table 2 shows the differences between each set of four shifts for the Total Group (A and B combined) eliminating shift No. 1. The difference (in average errors) for the first set and the second set is significant below the 1% level of confidence. (It must be remembered that Shift No. 1 is eliminated because the average number of errors is far below the others. This shift is always easy because no shifts have come before it.) This indicates that there was a significant reduction of errors toward the end of test which implies an improvement in shifting.

TABLE 2
DIFFERENCES BETWEEN SHIFTS 2, 3, 4, 5 and 6, 7, 8, 9
FOR TOTAL GROUP

SHIFTS	MEANS	NO.	d.f.	"t"	LEVEL OF CONFIDENCE
2,3,4,5	7.725	4	6	4.37	less than 1%
6,7,8,9	4.945	4			

Figure 3 indicates a marked difference between Groups A and B. The errors in Group A fall sharply after shift number three. After shift number three, the subject is returning to an old category. Group A is characterized by an early understanding that a shift between color, form and number was involved. Although there is a reduction of errors in group B the decline is not as sharp. There is greater irregularity in Group B than there is in Group A. Group B is characterized by a vague understanding that a

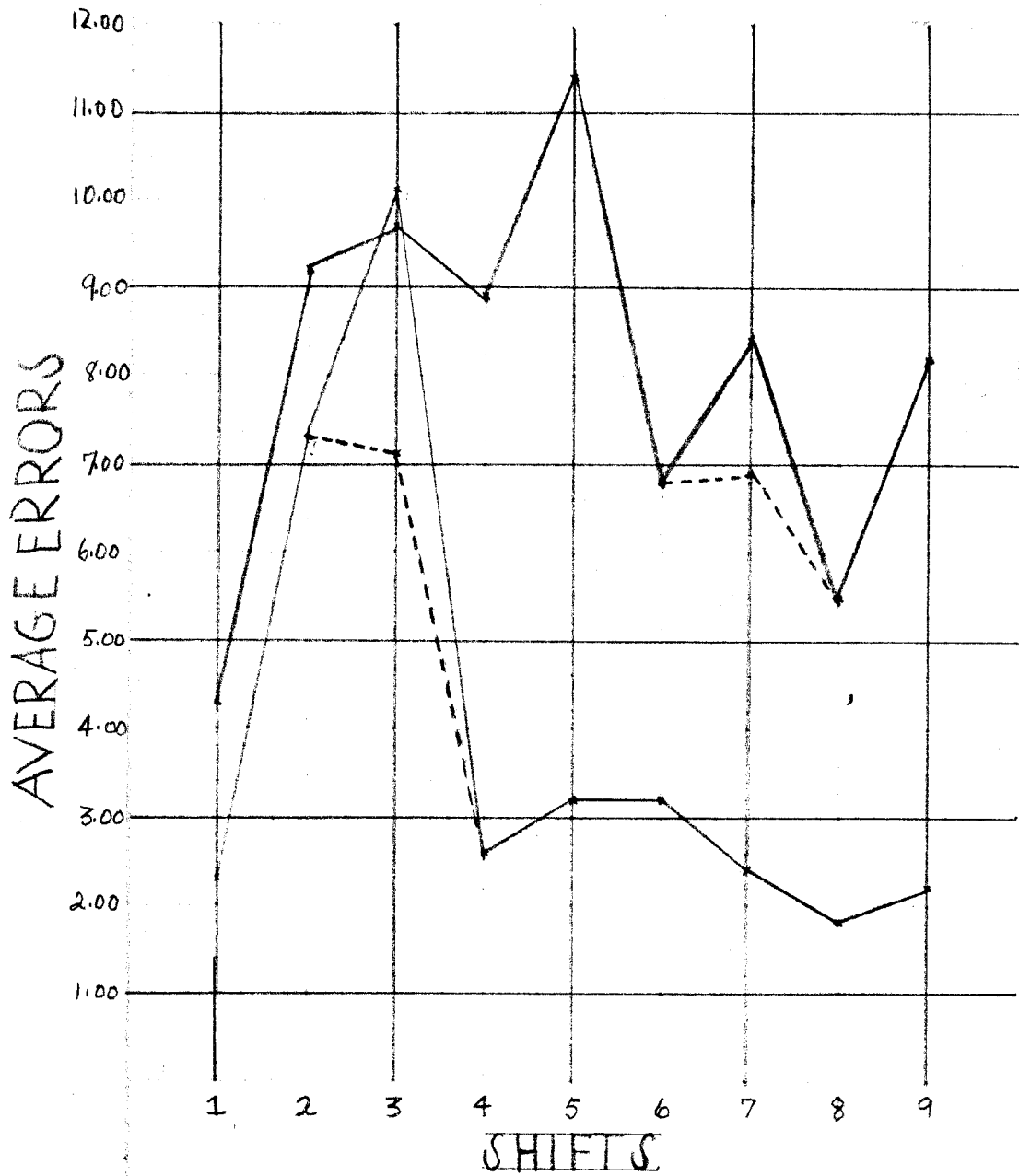
shift was involved, but they weren't quite sure what it entailed or approximately when it would come. The majority of the students in group A could anticipate the shift. Group C (Figure 3) shows marked irregularity and a great increase in errors and number of cards used. This would have still been greater if the experiment was carried through the ninth shift. These experiments were not completed because the subjects showed no knowledge or understanding that a shift was involved and the time consumed to solve the problem was well over an hour. (The whole test usually takes $1/2$ to $3/4$ of an hour.)



— TOTAL - (Groups A and B Combined)
- - - - - Deducting one extreme case from
CATEGORY 3 (51 errors) + CATEGORY 7 (41 errors)

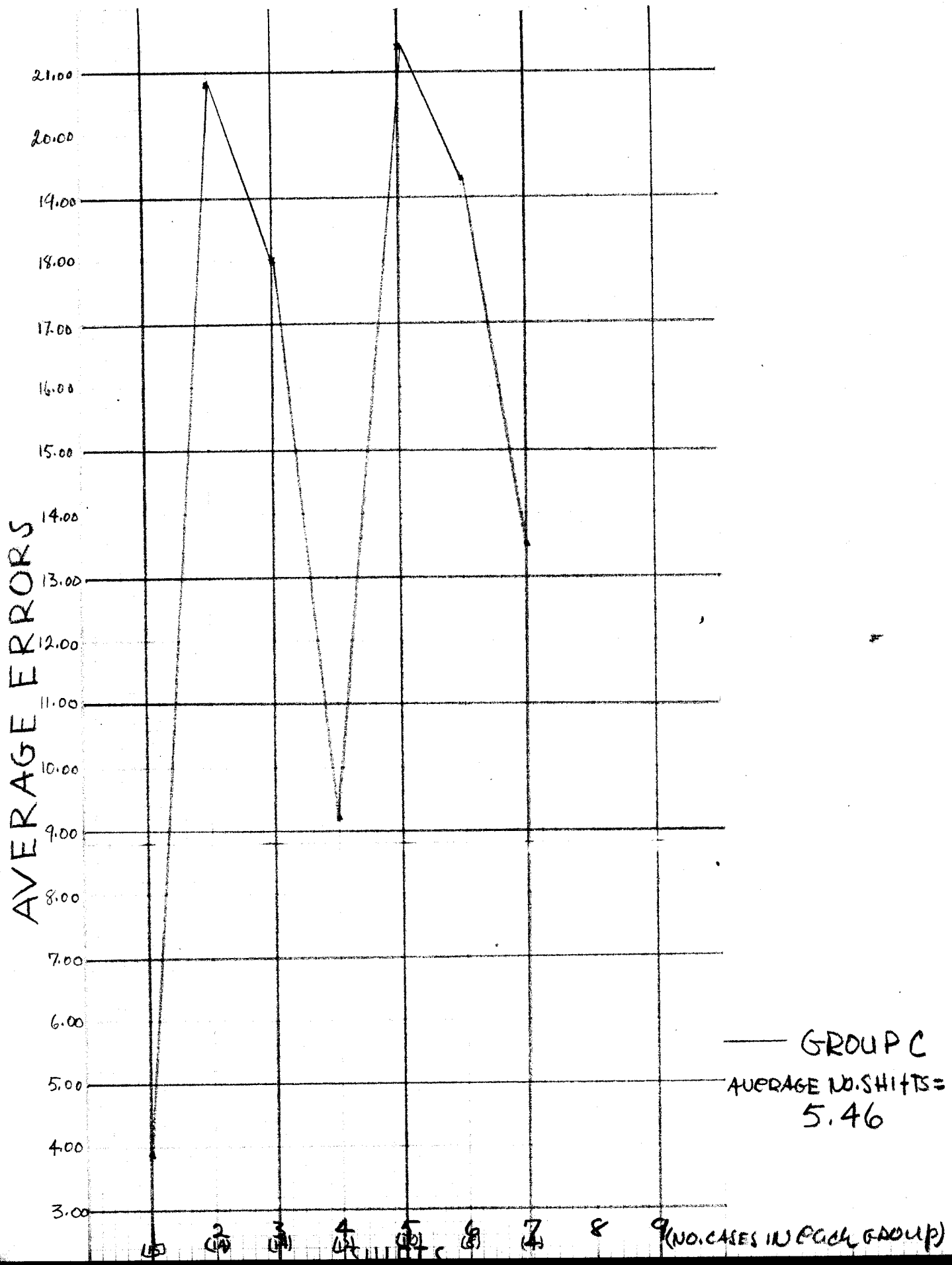
TOTAL GROUP

FIG 1



— GROUP A
- - - - - Deducting one extreme case from category
3 (51 errors)
— GROUP B
- - - - - Deducting one extreme case from category
7 (41 errors)

FIG. 2.



An item in the data that might prove to be important in later research is the ratio of errors of shift number 2 to shift number 1. Table 3 indicates that the Total Group (A and B), Group A, and Group B all have about a 2:1 ratio, while Group C has about a 5:1 ratio. With further experimentation this may turn out to be a valid indicator of rigidity. There is a marked increase in errors for the group that does not solve the problem on the second shift.

TABLE 3
RATIO OF AVERAGE ERRORS IN SHIFTS 1 AND 2

GROUP	RATIO SHIFTS 2:1
TOTAL (A and B)	Ca. 2:1
A	Ca. 2:1
B	Ca. 2:1
C	Ca. 5:1

Another item that may prove to be important is the ratio of shift 3 to shift 4 in Groups A and B. The ratio in Group A is 3:1, in Group B $1 \frac{1}{4}$:1. This may be an indicator of "full insight" versus "partial insight".

TABLE 4
RATIO OF AVERAGE ERRORS IN SHIFTS 3 AND 4

GROUP	RATIO SHIFTS 3:4
A	Ca. 3:1
B	Ca. $1 \frac{1}{4}$:1

TABLE 5
AVERAGE ERRORS IN EACH SHIFT

GROUP	1	2	3	4	5	6	7	8	9
TOTAL	3.88	8.63	9.93	6.25	7.83	5.27	5.90	3.90	5.67
A & B	36	$\sigma = 8.94$	$\sigma = 7.52$	$\sigma = 5.32$	$\sigma = 7.94$	$\sigma = 5.04$	$\sigma = 1.80$	$\sigma = 4.93$	$\sigma = 3.71$
A	15	3.33	7.27	2.60	3.20	3.20	2.40	1.80	2.13
B	21	4.27	9.14	8.86	11.43	6.76	8.38	5.48	8.19
C	15	3.87	20.79	18.00	9.17	21.40	13.50		
"No.	15	14	14	12	10	8	4		

*Eliminating one extreme-case (51 errors) from Shift 3 and (41 errors) from Shift 7.

"Number cases in each shift.

Finally, odd-even correlation on Total Group (A & B) with the Spearman-Brown correction formula turned out to be + .982. This indicates that the score is reliable.

Qualitative Results

The individual records were not only scored according to the number of errors and cards used, but also each response was recorded according to the subject's system of classification. Therefore, the type of errors the subjects made could be qualitatively analyzed. Many subjects had hypothesis that were neither color, form nor number. In this situation the duplications in the pack of sixty of the four stimulus cards proved to be of value. If the subject classified one of the reproductions of the stimulus cards not under its duplicate, but under one of the others, where it could not possibly fit for either color, form or number, the experimenter then knew that the subject had some unknown system of classification and could later question the subject about it. Negative instances also proved to be of value in determining the subjects' system of classification. The subject might classify for color four times in succession and then classify a card that is neither form, number or color. This was a hint to the experimenter that the four previous classifications for color might not have been color at all but some other hypothesis. This was noted in the record beside each response and the subject

was later questioned about his systems of classification. Each group showed differences in classification. These groups will be discussed separately.

Group A

Not one subject in group A classified the cards for anything else but color, number or form. They discovered early in the experiment that the experimenter was shifting and tried to anticipate her shifts. They all reported that they thought the experimenter shifted when they made approximately 6-7 correct responses in any category. When the category was changed they knew that they had two possibilities left and they tested these possibilities analytically. By the process of elimination, they could determine what category they were shifted to. Therefore, there should be a maximum of two errors in any one category. This was borne out by the results. In figure 2, the last three shifts for Group A have average errors between 2 and 3. Group A completed the experiment in 15-20 minutes, 6 subjects completed the experiment in 7-12 minutes.

Group B

This group showed greater variability in their hypotheses than group A. They tried hypotheses other than color, form or number. They attempted to classify for position on the card, complimentary colors, two categories simultaneously and the position of the spaces between the

figures. All the subjects knew that there was a shift, but could not state precisely what the shift was. Most of the questionable responses (hypotheses other than color, form and number) came between the fourth and the sixth shifts. They were confused by the fact that what was right on one trial would be wrong on another. "I'd get a series of 'rights' and then 'wrongs'. I thought I'd stumbled on a plan." "You seemed to change on me." "It seemed like there was changing. One of us was inconsistent." "As soon as I get something right it doesn't seem to go. You're not changing them as soon as I get one right are you?" These protocols were typical of the subjects' remarks during or after the experiment. When asked which hypotheses worked best throughout the experiment, they always stated that color, form and number gave them the best results and they usually discarded other hypotheses by the seventh shift. When asked what they thought was involved in the experiment they could not precisely verbalize either what they were doing or what they thought the experimenter was doing. The tendency to perseverate the responses of the last category was apparent in this group. If they had successfully sorted for number and the shift was made to color they would continue to sort for number in spite of the fact that it would now consistently be called wrong. There was little perseveration in group A.

Group C

This group is characterized by persistency or extreme variability. Some of the subjects would consistently classify for position without changing without regard for the experimenter's "wrongs". One subject completely ignored the stimulus cards and tossed the cards at random. Another was too confused and refused to go on. This group shows the greatest perseveration. Some would consistently classify for the last category, after the shift, for more than one hundred cards, in spite of the fact that they were all wrong. Many of the subjects thought that there was one underlying hypothesis in the test and tried to find it. The test was demonstrated by the experimenter to three subjects. They still did not understand what was involved and only after verbal explanation were they enlightened. As in group B they showed great variability in their hypotheses, but unlike group B, instead of discarding false hypotheses, they persisted in them. Many of these subjects looked at the situation vertically instead of horizontally. Card one was for triangles (form), card two was for green (color), card three was for all three's (number) and card four turned out to be a wastepaper basket for all those cards that did not fit the other categories.

Not only were there quantitative differences in these groups but as can be seen by the above, there were differences in the manner of solution in these groups.

Discussion

All the subjects recognized that the cards could be sorted for color, form or number. Before the experiment started many of them asked, "How shall I match them, for color, form or number?" The difficulty in some subjects (Group C) arose not in isolating the "concept" but in shifting from one "concept" to the other. Did these subjects see color, form and number, or did they see red, green, blue, triangle, star, one, two, etc.? That is, did they see these properties as isolated elements and therefore, they were unable to shift? It is the writer's opinion that they did.

One subject classified red under green and yellow under blue. This is not seeing a "color category", but seeing specific isolated colors. Another subject classified all triangles under one red triangle (the first stimulus card), all greens under two green stars (the second stimulus card), all three's under three yellow crosses (the third stimulus card) and all others under the fourth stimulus card which turned out to be a repository. This again, is seeing things as specific elements, not as categories. All of the subjects in group C were characterized by extreme flexibility or extreme rigidity (10). Either they changed rapidly from one thing to another or they persisted in the same category without changing. This, the writer believes, was

due to the fact that they saw specific elements and not categories.

There was a sharp difference between the three groups in "conscious awareness". Group A knew there was a shift and could verbalize precisely what it was. Group B vaguely knew there was a shift and could not adequately verbalize it. Group C was completely oblivious to the shift and could only not verbalize it, but could also not verbalize clearly what they were doing. Most of the subjects in group C thought there was one underlying hypothesis.

In the light of Goldstein and Scheerer's discussion of abstract and concrete behavior this proves to be interesting. They maintain that "any sorting which involves genuine abstraction is necessarily bound to conscious volitional act of reflecting upon the properties of the object with reference to a concept, a class or category." (1, p.24) By this they do not mean that verbalization need necessarily be synonymous with conceptualization. In order to abstract the subject must attend to categories or classes and not specifics. I would maintain, from these results, that the better an individual could verbalize a concept, the clearer he has it. This is borne out by the differences of Groups A and B.

This was a preliminary experiment. We have scores in terms of average errors and average cards used for a superior normal group. Further experimentation is needed

to make comparisons with other groups, normal and abnormal and at various intelligence levels. When these data are obtained there may be critical differences in the performance of the groups (both quantitative and qualitative), so that the test may turn out to be an adequate quantitative measure of flexibility in thinking.

In a recent experiment (7) we have some indication as to how an old group will perform on this test. (Ages 58-73, mean age 66) There were twenty-two subjects, sixteen sorted spontaneously for color, four for form and two for number. If by chance, the experimenter's category was not the same as their spontaneous sorting they could not get beyond the first shift and continued to sort for their spontaneous category, without changing, until the experiment had to be stopped. If by chance, the experimenter's category was the same as their spontaneous sorting, they got five consecutive correct responses in the first shift, but continued to sort in the same manner after it was called wrong. These people, then, did not get beyond the second shift. They rarely sorted for anything but color, form and number. Sometimes they would sort for position on the card. One subject said, "Blue goes with red because it is the color in the flag; yellow goes with blue because it is the moon in the sky." Another subject said, "Star goes with star, because they look alike." The tendency to perseverate was greater in this group than in any of the

normal groups. Here, again, the subjects are classifying for specifics, not for categories. At the close of the experiment they were asked if the cards could be sorted another way. They saw no other way until it was explained to them by the experimenter, and this was a revelation. They exclaimed, "Oh, yes, it could be done that way, too." In this group there is greater rigidity than any of the normal groups.

Summary

An attempt was made to devise a test for "abstract behavior" and "shift of set" that could be objectively scored. A group of college students age 17-28 were used for the preliminary tests. They were divided into three groups. Group A: those who discovered early in the experiment that a shift was taking place and tried to anticipate the experimenter's shifts. Group B: those who completed nine shifts and had a vague notion that the experimenter was shifting. Group C: those who completed an average of 5.46 shifts and had no knowledge that a shift was taking place. Finally, an old group, age 58-73, showed greater rigidity than the young group and never got beyond the first shift. These people classified predominately for color.

It remains for future experimentation to decide what other groups may do on this test and what kind of objective scoring will be available. So far, the test can be scored

in terms of errors, number of cards used, number of shifts, the ratio of shift number two to shift number one and the ratio of shift number three to shift number four.

BIBLIOGRAPHY

1. Goldstein, G. and Scheerer, M. Abstract and concrete behavior: An experimental study with special tests. Psychol. Monog., 53, No. 2, 1941, p. 1-151.
2. Hanfmann, E. and Kasanin, J. A method for the study of concept formation. J. Psych., 1936, 3, p. 521-540.
3. Hanfmann, E. and Kasanin, J. Conceptual thinking in schizophrenia. Ner. and Ment. Dis. Monog., No. 67, 1942.
4. Heidbreder, E. An experimental study of thinking. Arch. Psych., 1924, 11, No. 73.
5. Hull, C. L. Quantitative aspects of the evolution of concepts. Psych. Monog., 1920, 28, No. 1.
6. Kuo, Z. Y. A behavioristic experiment on inductive inference. J. Exper. Psych., 1923, 6, p. 247-293.
7. Morrow, M. A. Performance on the Hunt test with differential vocabulary scoring compared with performance on a test of rigidity. Unpublished Master's thesis, University of Wisconsin.
8. Smoke, K. L. An objective study of concept formation. Psych. Monog., 1932, 42, No. 4.

9. Vigotsky, L. Thought in schizophrenia. Arch. Neur. and Psych., 1934, 31, p. 1063-1077.
10. Weigl, E. On the psychology of so-called processes of abstraction. J. Abnorm. and Soc. Psych., 1941, 36, p. 3-33.

Approved:

David A. Grant

David A. Grant
Associate Professor of Psychology

Date:

January 15, 1947
