THE SQUARES TEST AND LEVELING-SHARPENING: A STUDY OF INSTRUCTIONAL SET AND SEX DIFFERENCES

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY

Albert R. Gilgen

1965

This is to certify that the

thesis entitled

THE SQUARES TEST AND LEVELING-SHARPENING:

A STUDY OF INSTRUCTIONAL SET AND SEX DIFFERENCES

presented by

Albert R. Gilgen

has been accepted towards fulfillment
of the requirements for

PhD degree in Psychology

Paul Bakan
Major professor

Date July 12, 1965

O-169

ABSTRACT

THE SQUARES TEST AND LEVELING-SHARPENING: A STUDY OF INSTRUCTIONAL SET AND SEX DIFFERENCES

by Albert R. Gilgen

According to the theory of cognitive controls (Klein, 1951), leveling-sharpening represents an important dimension of cognitive structure. It is believed that memory traces of previous stimuli fuse or assimilate, and that they do so more for some individuals (levelers) than others (sharpeners). This supposedly leads the former to differentiate less among successive stimuli than the latter.

The Squares Test, designed by Holzman and Klein (1951), generally serves as the criterion task for the concept leveling-sharpening. The test consists of a series of 150 squares of light projected successively onto a black screen in an almost completely darkened room. The squares range in size from 1.2 to 13.7 inches and the series is made up of 10 overlapping subseries. Subseries 1 involves the 5 smallest sizes (presented in 3 different orders), and the series progresses in stepwise fashion from the smallest to the largest squares so that Subseries 10 includes only the 5 largest sized squares. Subjects (Ss) are required to estimate the size of each square.

Numerous scores, all purporting to reflect differences in memory trace assimilation (leveling-sharpening) have been used, some measuring across-series changes in judgment and others based on differences in intrasubseries square size differentiation. For this reason, the operational definition of leveling-sharpening is rather unclear. Definitions of the concept are further obscured by the procedural modifications which characterize much of the research. Furthermore, investigators have had little success in demonstrating that leveling-sharpening is a major factor in cognitive functioning.

A study by the author (Gilgen and Diehm, 1964), in conjunction with a review of the literature, led to a general dissatisfaction both with the concept leveling-sharpening as defined in terms of assimilation theory and with the Squares Test as a measure of any single dimension of cognitive functioning.

The present study consists of two parts: Experiment I, which was exploratory and had as its purpose the determination of those aspects of the Squares Test which are used by Ss as judgmental referents; and Experiment II which systematically investigated the hypotheses suggested by Experiment I findings.

Thirty-three Ss (14 men and 19 women) were individually tested in Experiment I, each S taking the Squares Test twice during the same session. A short form of the test (the first 90 presentations) was used in order to reduce boredom.

The findings were as follows: (a) the Squares Test has good short-term reliability (.94); (b) some Ss tended to overestimate extremely; (c) women tended to overestimate more than men; (d) intrasubseries discrimination (as measured by Positional Ranking Accuracy) was positively related to the range of size estimates (as measured by the largest size estimate given) for women (.66), but not men (.21); (e) the instructions fail to make it clear whether Ss are to use whole numbers or fractions in making their size estimates; (f) explicitly telling Ss that none of the squares are exact whole numbers of inches in size increased intrasubseries square-size discrimination; (g) Ss tended to discriminate less well among the square sizes the second time they took the Squares Test.

Experiment II (N = 350; 175 men, 175 women) was designed to investigate the roles of instructional set, instructional ambiguity, sex, and the motivation of Ss on Squares Test performance. Since the smallest judgment made by most Ss was 1 inch, the range of size estimates was measured by the largest size estimate made (LSE); in addition, performances were scored for Positional Ranking Accuracy (PRA). Three different sets of instructions concerning the range of square sizes were used (1-7 inches, 1-12 inches, and 1-16 inches) and Ss were either given the regular ambiguous instructions concerning response scale refinement or explicitly told that fractional estimates were permissible. One group which was given no range information was tested later.

The findings were as follows: (a) the ranges of size estimates were functions of the information given Ss concerning the range of square sizes to be judged (size expectancy was operative); (b) most Ss who were given no range information underestimated; (c) although sex was not significantly related to the range of size estimates, women who overestimated tended to overestimate more than men who overestimated; (d) PRA and LSE scores were significantly related for both men and women; however, the relationship obtained under more conditions for the women than the men; (e) the ambiguity of the instructions concerning the inch scale refinement (whole numbers or fractions) which Ss are permitted to use was reflected in PRA scores (PRA increased when Ss were explicitly told that they could use fractions); and (f) the motivation of Ss on the Squares Test (as measured by questionnaire responses) was positively related to PRA scores.

It was concluded that instructional set, instructional ambiguity, and the sex and motivation of Ss are involved in Squares Test performance, and that the test fails to provide a meaningful operational definition of leveling-sharpening. The concept obscures rather than clarifies Squares Test performance and should be dissociated from the test. It was suggested that the test might, however, be fruitfully employed in the investigation of instructional set.

THE SQUARES TEST AND LEVELING-SHARPENING:
A STUDY OF INSTRUCTIONAL SET
AND SEX DIFFERENCES

By

Albert R. Gilgen

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1965

ACKNOWLEDGMENTS

My first acknowledgments go to my wife, Carol, and my three children, De Forest, Beth and Bert, who endured all of the hardships demanded by this undertaking. To my guidance committee: Dr. Paul Bakan, Dr. Hans Toch, Dr. Bertram Karon, and Dr. Howard Bartley, I extend my thanks and appreciation for professional assistance. I wish, in particular, to thank Dr. Bakan, the committee chairman, for encouraging and guiding my interest in this area of research. A special acknowledgment must also go to Dr. Toch, whose constructive criticism played a primary role in the final design of the study. In addition, thanks go to Dr. Donald Johnson, who read the original proposal and offered many helpful suggestions.

I also extend my appreciation to Dr. William Stellwagen, whose cooperation expedited the enrollment of subjects for this study; to Mr. Royal Olson, who helped assemble the equipment; and to Mr. Joseph Zerbolio, who designed the electrical circuit and offered time-saving suggestions with regard to the analysis of the data.

Finally, I am indebted to the U. S. Department of Health, Education and Welfare whose financial assistance, in the form of a pre-doctoral National Institutes of Health Fellowship, enabled me to work full-time on this project.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

The Squares Test was originally designed by Holzman and Klein (1951) as part of a psychoanalytically oriented research effort to identify the dimensions of cognitive organizational structure. It is now the criterion task for the hypothesized cognitive control known as leveling-sharpening.

Although leveling-sharpening is operationally defined in terms of certain characteristics of Squares Test performance, it is theoretically defined in terms of Gestalt assimilation theory. The logical link between these two definitions has, however, not been adequately demonstrated; and clarification of this concept depends in part on a better understanding of the Squares Test.

The exploratory phase of this study (Experiment I) suggested that instructional set and the sex of the subjects (Ss) play important roles in Squares Test performance, and Experiment II was designed to systematically investigate these variables. Since the review of the literature indicated that sex differences on perceptual and cognitive tasks were in part due to such factors as interest and motivation, an effort was made to determine what bearing these variables had on Squares Test scores.

The Theory of Cognitive Controls

Psychoanalytic theory, according to the "ego psychologists" (Hartmann, 1939; Rapaport, 1957), had failed to consider adequately the reality-testing functions of the ego. Klein attempted to remedy this deficiency with his theory of cognitive controls.

The main points of Klein's theory first appeared in a paper concerning the adaptive properties of sensory functioning (Klein, 1949). In agreement with both Freud (1922) and Stern (1938), Klein envisioned the sensory apparatus as serving two adaptive ends: one protective, the other constructive. The sense organs and affective nervous system protect the organism against excessive stimulation and constructively organize that information which is permitted into consciousness.

Klein said that "such constituent properties of sensory organization" as absolute thresholds, differential thresholds, organizing speed and systems of organization are not determined by the physiological limitations of the organism alone but are modifiable and therefore capable of taking various "form-varieties." One "form-variety" of absolute threshold, Klein theorized, might be characterized by a combination of a high lower limit and a low upper limit, resulting in a narrow range of stimulation being admitted into awareness. The defensive function of such a condition is obvious, and Klein proposed that the schizophrenic's sensory system might be characterized by such a "form-variety" of absolute threshold. Other personality types would likewise be distinguished by different "form-varieties" of this perceptual property.

Although Klein did not specifically mention leveling-sharpening in this paper (1949), he did refer to some work being conducted at the Menninger Clinic concerning the "integration of sensory data." He also talked about "schematizing," a term which later became associated with the Squares Test. Schematizing refers to a process noted by Hollingworth (1913) whereby an individual becomes attuned to a series of successively presented stimuli through establishing a reference level against which all subsequent stimuli are compared.[1]

To summarize, Klein (1949) accepted the basic tenents of psychoanalytic theory but emphasized the adaptive capabilities of the sensory apparatus. The cognitive and perceptual functions of the ego, he felt, should be studied in the laboratory; and the individual performance differences found in psychophysical experiments were seen as manifestations of the ego's cognitive organizational strategies.

A more complete version of Klein's theory appeared in a paper delivered at the 1949 meeting of the American Psychological Association (Klein and Schlesinger, 1949). Cognitive structure was here envisioned as a hierachical arrangement

---

[1]This idea was the forerunner of Helson's adaptation level theory.

of modes, syndromes, and Anshauungen, which together comprise total personality. Modes are the different ways ("form-varieties") in which perceptual properties such as thresholds, etc. are used to serve adaptive ends. These modes cluster into syndromes, each characterized by a single perceptual attitude (Anschauung) or integrative principle. A syndrome might have several adaptive properties, the interrelationship of which may provide clues to the Anschauungen involved. Furthermore, Klein and Schlesinger conceptualized syndromes as preferred rather than required styles of expression.

The mode, syndrome, Anschauung scheme, in other words, represented a model of the way individuals typically but not invariably function cognitively. Klein and Schlesinger, in discussing the approach they intended to employ in their search for broad integrative principles of cognitive functioning, stated: "We may seem to imply it, but we really do not intend to jump the gap between clinical and experimental methods with correlation coefficients." (P. 45) As further review of the literature will indicate, however, most of the work on cognitive controls has, in fact, involved correlational techniques.

Klein's next major theoretical article (1951) did not introduce any basic changes but was devoted primarily to an elaboration of the concept Anschauung. Whereas this concept had been defined as an integrative principle inductively arrived at by the systematic uncovery of modes and syndromes (mode clusters), it was now referred to as "a personal outlook on the world, embodying in perception one of the ego's requirements." (P. 33) Moreover, an Anschauung was described as a solution to an adaptive task and a style of organization which expresses a central or "executive directive" of the ego-control system.

Leveling and sharpening were presented as examples of Anschauungen although there was at that time little empirical basis for giving these tendencies the status of broad integrative principles of cognitive organization. Klein also stated; "The 'sharpening' attitude, which I will discuss more fully later on, says in effect: 'Be alert to all shades and nuances. Let nothing slip by unnoticed." (Klein, 1951, p. 353) However, it does not follow that Ss classified as sharpeners on the basis of their Squares Test performance are necessarily characterized by such a point of view.

The conceptual elaboration in this article, in addition to the rather premature assumptions which were made concerning the status of leveling and sharpening, tended to obscure Klein's original model of cognitive functioning. It was no longer clear exactly what was meant by the concepts mode, syndrome and Anschauung. The emphasis had shifted from the operational to the theoretical definition of terms.

Considerable discussion has appeared in subsequent articles concerning the relationship between Anschauungen[1]

---

[1]The term "cognitive control" was later substituted for Anschauung (Gardner et al., 1959).

and psychoanalytic theory. Although this topic will not be considered in detail, it should be mentioned that much has been said about the similarities and differences between cognitive controls and defense mechanisms. Gardner *et al*. (1959), for example, state:

> It is . . . doubtful that two sets of controls are involved, each invoking different mediating structures. It seems more likely that defenses and cognitive controls involve the same signal and action apparatuses. Both defenses and controls can expedite or complicate drive discharge. The concept of control only draws attention to the extent to which drive discharge is appropriately attuned to reality. (P. 12)

In conclusion, cognitive control theory, while concerned with perception and cognition, remains faithfully within the framework of psychoanalytic theory. As such, it is heavily encumbered with conceptual structure. Research associated with the theory has been mostly correlational in nature. The main points of the theory are: (a) that individuals develop different cognitive organizational styles (cognitive controls) which manifest themselves in perception, memory, learning, etc., and (b) that perceptual experiments are the most convenient way of studying these manifestations.

The Squares Test

The Squares Test consists of a series of squares of light projected singly and successively onto a black screen in an almost completely darkened room. The series progresses gradually but irregularly from the smallest to the largest squares. Subjects (*S*s) are asked to estimate the size of each square. The procedure presented here is that originally used by Holzman and Klein (Holzman, 1952). It is the same as that currently in use at The Menninger Foundation except that the sizes of the reference squares shown during the instructions are now 1 and 18 inches, instead of 1 and 16 inches.

Setting and equipment

The squares of light are projected by means of a 35 mm. slide projector.[1] Room cues are reduced either by conducting the experiment in a black room or by draping the projection area with black cloth. Ss sit about 10 to 15 feet from the screen and are usually tested in groups of 5 or less.[2] Each S is given a small flashlight or provided with a well-shielded table lamp which illuminates the recording sheets.

The procedure

Ss are admitted to the darkened room, given recording sheets and read the following instructions (or, sometimes the

---

[1]A film strip projector may also be used. The Squares Test film strip can be ordered from The Menninger Foundation.

[2]The test is sometimes administered to large groups in an auditorium or other large room (Holzman, P. S., personal communication, 1962), although it is difficult to understand how such a radical change in setting can be employed without changing the nature of the test.

instructions are projected on the screen):

> We wish to see how well you can judge the sizes of squares. We are going to show you a number of squares on the screen and we want you to tell us how big they are.
>
> The squares may range anywhere between one and sixteen inches. This does not mean that you will necessarily get a square which is one inch or sixteen inches, though you may. The squares, however, will always be somewhere within this range.
>
> To help you judge the sizes of the squares, we shall show you what a one inch square looks like--the smaller end of the range and what a sixteen inch square looks like--the larger end of the range.
>
> (The one and sixteen inch squares are then presented for five seconds each, following which the instructions continue.)
>
> You will see 150 squares during the course of the experiment, and you have answer sheets with a total of 150 spaces provided for your estimates. Write your estimation of the size of each square in its own numbered space. Thus for square number one, record its size in inches next to number one, etc.
>
> Don't go back over your judgments to change them. In changing them you are more likely to be inaccurate. Please don't compare your estimates with anyone or make any comments during the experiment. Make your judgments independently.
>
> Now to remind you of the range in which the squares will fall, we shall again show you the smaller and the larger ends of the range.
>
> (The one and sixteen inch squares are then presented twice for periods of five seconds each.)
>
> Now we are ready to begin. You will see each of the following squares for only a few seconds. Look at it all of the time it is on the screen and make your estimation when it disappears. The next square you will see will be number one.

The 150 squares of light, ranging in size from 1.2 through 13.7 inches, are then projected one after another.

Each square is visible for 3 seconds and is followed by an 8 second dark interval during which time the Ss record their size judgments.

The series of squares

The series is comprised of the following sized squares (inches): 1.2, 1.6, 2.0, 2.4, 2.8, 3.2, 3.8, 4.6, 5.5, 6.6, 7.8, 9.5, 11.4, and 13.7, the differences between sizes being equal to j.n.d. plus 20 per cent (Holzman, 1952). The order of presentation is as follows: (a) The first 15 presentations (Subseries 1) involve the 5 smallest squares only, first in ascending order, followed by two random orders; (b) the smallest square (1.2 inches) is then discontinued and the next largest square (3.2 inches) introduced, each of these sizes then appearing three times (15 presentations) in the same relative order as the squares in Subseries 1. This stepwise procedure continues until only the five largest squares appear during the last 15 presentation (Subseries 10). Thus the series is characterized by a gradual but fluctuating increase from the smallest to the largest squares. Figure 1 illustrates the construction of the series (see Appendix A for order of presentation).

Scoring procedures

Squares Test performance has been scored in several different ways, each technique purporting to provide an

Figure 1

Square sizes involved in each subseries

(inches)

Subseries	1.2	1.6	2.0	2.4	2.8	3.2	3.8	4.6	5.5	6.6	7.8	9.5	11.4	13.7
1	x	x	x	x	x									
2		x	x	x	x	x								
3			x	x	x	x	x							
4				x	x	x	x	x						
5					x	x	x	x	x					
6						x	x	x	x	x				
7							x	x	x	x	x			
8								x	x	x	x	x		
9									x	x	x	x	x	
10										x	x	x	x	x

index of leveling-sharpening tendencies. Only those scores most commonly used will be described.

Positional Ranking Accuracy (PRA)

This measure, which has also been called Percentage Ranking Accuracy (Holzman, 1962[1]), Ranking Accuracy (Gardner *et al*., 1959), and Accuracy (Krathwohl and Cronbach, 1956), has been the most commonly employed index of leveling-sharpening. It is computed as follows:

> Divide the test protocol (150 judgments) into ten series of 15 judgments each. Divide each of these series of 15 into three sub-series of 5 each. Count the number of times the smallest square is correctly ranked within each of the first three sub-series (minimum, 0; maximum, 3). Then, the number of times the next smallest square is correctly ranked; the number of times the third, or middle square, is correctly ranked; likewise the fourth, or second largest; and the fifth, or largest. Enter these values on the score sheet in the first *diagonal* column. Do the same for the next 9 series. Sum all the entries on the score sheet. Divide this sum by 150 (the maximum number of correct within-sub-series rankings) to get the percentage of correct rankings) to get the percentage of correct rankings. A high score is assumed to represent the sharpening end of the score dimension. A low score is assumed to represent the leveling end of the dimension.
> (Holzman, 1962[2])

Although the original scoring system allowed no credit when two or more adjacent squares were judged the same size within a subseries, the most recent technique gives ½ credit to each square estimated as being the same size as another.

---

[1]Personal communication.

[2]*Ibid*.

For example, if a S's estimates for the first five squares rank 1,2,3,3,5 whereas the actual sizes of the squares rank 1,2,3,4,5, the accuracy credits would be assigned as follows:

S's judgments (rankings)		Accuracy Credits
1		1
2		1
3		½
3		½
5		1
	Total	4

The percentage accuracy of this set of judgments is 80 (4 out of 5). The overall PRA score is the mean of the ten subseries scores.

While PRA is a function of the number of times a S assigns the same estimate to two adjacent but different-sized squares, it also reflects the number of ranking reversals which occur. A reversal takes place when a larger square is judged as being smaller than a smaller square. For example, if the 1.2 inch square is estimated as 2 inches in size while the 1.6 inch square is judged to be 1 inch, then a reversal in the ranking of the judgments occurs. In such cases, zero credit is given for both estimates involved in the reversal.

Accuracy Loss (AL)

This score, also known as "Loss of Positional Accuracy" (Holzman, 1952), is derived as follows:

> The 'accuracy loss' score is obtained by examining separately the accuracy on the largest stimulus in each set of five and the accuracy on the other four. We shall symbolize accuracy on the largest stimulus

> (rank 5) as A5 and the others as A1, A2, A3, and A4. 'Per cent accuracy loss' examines whether accuracy on the squares 'embedded' in the series is as great as on the large square. (Experience shows that accuracy on square 5 is always greatest.)
>
> (Krathwohl and Cronbach, 1956, p. 311)

The following formula is used to calculate Accuracy Loss (AL):

$$AL = \frac{(A5-A4) + (A5-A3) + (A5-A2) + (A5-A1)}{4}$$

PRA and AL scores usually correlate highly. Krathwohl and Cronbach (1956) found that the two measures correlated .82. This, they pointed out, indicated that the AL score, which is based on differences between highly correlated values, probably has low reliability. The AL score is not used much currently. It has been described because it, in conjunction with PRA, comprised the original leveling-sharpening index (Holzman and Klein, 1954).

<u>Leveling-Sharpening Score (L-S)</u>

This measure is a composite of the PRA and AL scores and is computed as follows:

$$L\text{-}S = \frac{PRA}{\sigma_{PRA}} + \frac{AL}{\sigma_{AL}}$$

(Jeffreys, 1953)

It should be noted that although this score is referred to as the leveling-sharpening score (L-S), PRA and the other measures to be discussed have all been used as indices of leveling-sharpening.

Average Shift (AS)

Jeffreys (1953) devised a score referred to as Average Shift (AS) which is based on the trend of the entire series rather than the subseries. This score is calculated by subtracting the mean of Subseries 1 estimates from the mean of the Subseries 10 estimates and dividing by ten. It yields a two-point slope measure. This performance index is relatively independent of PRA, Jeffreys reporting a correlation of .08, and Krathwohl and Cronbach (1956) a correlation of .26 (neither relationship reaching significance).

Early Slope (ESL) and Late Slope (LSL)

Krathwohl and Cronbach (1956) found that the slope of the first five subseries correlated only .24 with the slope of the last five subseries indicating that the two measures reflect different information. ESL correlated .68 and LSL .84 with AS; therefore, LSL is "the major determiner of the average slope for the whole response curve." (P. 313) Of even more interest is the finding that PRA correlated only .06 with ESL while correlating .40 with LSL. Krathwohl and Cronbach concluded that PRA and the two slope measures reflect relatively independent aspects of Squares Test performance and for this reason are the best indices of the various response patterns which occur.

Increment Error (IE)

This score is computed as follows:

1. The percentage increase of each of the last 9 series means over the mean of the first series is computed thus, using the actual sizes of the stimuli [a 'series mean' here is the average size of 5 stimuli (each presented three times in one series)]. Series 1 contains the smallest figures, Series 10, the largest:

$$\frac{M_{\text{series 2}} - M_{\text{series 1}}}{M_{\text{series 1}}}, \quad \frac{M_{\text{series 3}} - M_{\text{series 1}}}{M_{\text{series 1}}}, \quad \text{etc.}$$

This yields 9 percentage increment values describing the objective shift in series means.

2. The same procedure is followed with the judgments of stimulus sizes by each S. (It should be noted that a 'series mean' for a S is the average of 15 judgments, 3 for each of the 5 figures in the series.) This yields 9 percentage increment values for each S that can be compared with percentage increment values of the actual stimuli.

3. Each S's percentage increment for each series is then subtracted from the actual stimulus percentage increment for that series (signs ignored). This yields a set of 9 increment error scores for each S.

4. These 9 increment error scores are averaged for each S. This yields a Mean Percentage Increment Error Score for each S.

(Holzman, 1962, personal communication)

A small IE score is considered an index of sharpening, while a large score represents leveling.

Lag or Increment Error (Regression) (IE-reg)

The Lag or Increment Error (Regression) score (IE-reg) was devised to eliminate the dependence on Subseries 1 estimates that characterizes the IE score. This score, more than any of the other measures discussed here, is sensitive to the absolute dimensions of the size estimates. All of the other scores are based either on intrasubseries comparisons or increment changes relative to Subseries 1 estimates.

It is computed as follows:

> For each S, the 150 judgments that comprise the test are divided into 10 series of 15 judgments each. . . . Ignore the first series. Compute the mean of the judged sizes for each of the last 9 series. Convert these means to logarithms. Compute the regression of the logarithms of these 9 series means on the series numbers (2-10). In the resulting regression equation, $Y = b_{yx} - a_{yx}$, only slope (b_{yx}) is used. The Lag Score is the difference (signs observed) between the slope for the actual stimuli (.0767) and the slope for the logarithms of S's means (i.e., .0767 - S's slope). A small value (or a minus value) is assumed to represent the sharpening end of the score dimension. A large value is assumed to represent the leveling end of the score dimension.
>
> (Holzman, 1962, personal communication)

In summary, while the Squares Test (or modifications of it) has consistently been used as the criterion task for leveling-sharpening, so many different scoring procedures have been employed that the actual operational definition of this hypothesized cognitive control is extremely unclear. The score most often selected, either alone or in combination with other scores, as the index of leveling-sharpening has been PRA.

Reliability

The long-term stability or reliability of cognitive controls was investigated by Gardner and Long (1960b).

Criterion tasks were administered to a group of 38 women three years after the original testing. The test-retest correlations were: PRA .52 ($P < .001$), IE .36 ($P < .05$), and IE-reg. .45 ($P < .01$) indicating that PRA was the most stable score.

Leveling and Sharpening

This section of the review is somewhat detailed not only because the Squares Test serves as the criterion task for leveling-sharpening[1] but because the findings reported are meaningless without an understanding of the procedural modifications and theoretical changes which characterize much of the research. The review is arranged in chronological order.

The first study which involved the Squares Test as an index of leveling-sharpening was conducted by Klein in collaboration with Holzman. This research was not discussed in a separate article but rather was described in some detail by Klein (1951). In referring to "one slice of the most relevant findings," it was observed that some Ss kept pace with the gradual increase in square size and were very accurate, while others lagged progressively behind. Lagging was ascribed to Ss' tendencies to "ignore, deny or suppress differences, to 'level' them to some simpler uniformity."

---

[1]The concept leveling-sharpening was first used by Wulf (1922) with reference to the recall of geometric figures. Allport & Postman (1947) applied the terms to certain phenomena connected with the transmission of rumors. Our concern is only with the concepts as defined by Klein and his associates.

Levelers were described as individuals who are unable to break away from an expectancy which develops early in the series that all the squares will be small. Sharpeners, on the other hand, were described as individuals who are better able to "consider each stimulus in its own right," being alert to change rather than suppressing it.

Klein reported that levelers found squares embedded in the middle of a subseries particularly difficult to estimate. Sharpeners discriminated better among the sizes, regardless of their position within a subseries. This observation led Klein to surmise that leveling and sharpening had something to do with Ss' abilities to extract stimuli embedded within a complex context.

Although the scoring procedure was not presented, it appears that Ss were classified as levelers and sharpeners on the basis of some kind of lag score.

In order to investigate the generality of leveling and sharpening tendencies, the extreme levelers and sharpeners were administered the Gottschaldt Figures (Thurstone version), a hidden faces test, and a task wherein the effects of contour on figure-ground contrast were determined. Although no data were presented, Klein reported that the levelers, to a greater extent than the sharpeners: (a) found the Gottschaldt Figures difficult to extract, (b) had trouble finding the hidden faces, and (c) were less influenced in the direction of greater contrast by contours placed between a figure and its ground.

An attempt was also made to determine the personality correlates of leveling and sharpening. This was done by asking the Ss' therapists[1] to describe their patients by means of a Q-sort of statements concerning personality characteristics.[2]

Levelers were found to be characterized by "self-inwardness" (retreat from objects, avoidance of competition, exaggerated dependency, self abasement, and passive drifting), while sharpeners were about equally divided between "self-inwardness" and "self-outwardness." This finding was the first clue that leveling-sharpening might not be related to a undimensional characteristic of personality.

No quantitative information or data analysis was presented, and therefore the findings and conclusions of the study cannot be evaluated. This is unfortunate because Klein, on the basis of this research, felt that leveling-sharpening had been validated as a broad integrative principle of cognitive structure.

Leveling-sharpening and the time-error phenomenon

Time-error refers to the constant error in judgment which occurs when a S compares the intensities of successively presented stimuli. If a circular area of ground glass is briefly illuminated twice in succession with equal intensity,

---

[1]The Ss in this study were mental patients.

[2]Based on the thirteen traits proposed by Murray (1938).

for example, Ss tend to judge the second stimulus as being brighter than the first. This is called negative time error because the intensity of the first stimulus is underestimated. Under some conditions, a positive time-error occurs--that is, the second stimulus is judged as less intense than the first.

In time-error studies, Ss are shown a series of successive stimulus pairs and required to determine whether or not the first member of each pair is heavier or lighter, louder or softer, etc., than the second. Although purely relative judgments are involved, this procedure has much in common with the Squares Test. The latter requires estimates using an absolute scale (the inch scale) but Ss to some extent make their judgments, as in time-error tasks, by comparing each stimulus with the previous one.

Between 1951 and 1954, two studies concerning leveling-sharpening and time-error were conducted. The first (Holzman and Klein, 1951[1]; Holzman and Klein, 1954) investigated only visual time-error while the second (Holzman, 1954) concerned auditory and kinesthetic time-error as well. As was mentioned previously, this research was suggested not only because the

---

[1]According to Holzman & Klein (1954), the Holzman & Klein (1951) article (a convention paper) is an abbreviated version of the Holzman & Klein (1954) article. In Holzman & Klein (1951), however, 17 male Psychiatric Aide students acted as Ss, whereas in the Klein & Holzman (1954) study, 43 Ss "about equally divided between men and women and consisting of students in a midwestern university and a psychiatric aide school" were involved.

tasks involved are similar, but because Gestalt assimilation theory provided a theoretical link between leveling-sharpening and time-error. Since assimilation theory underlies much of the theorizing about leveling-sharpening, a brief history of the theory follows.

Kohler (1923) proposed that negative time-error occurs because the second stimulus is compared with the faded trace of the first, thus leading to a tendency to perceive the second as more intense. Lauenstein (1933), however, investigating auditory time-error, found that when a tone louder than the first comparison stimulus was interpolated, positive time-error occurs, while an interpolated tone of less intensity than the first stimulus results in a negative time-error. These phenomena, Lauenstein believed, occur because the interpolated and first stimulus assimilate and comprise the internal referent against which the second stimulus is compared. According to this theory, negative time-error occurs because the first stimulus and the silent (or dark) period assimilate causing the first stimulus to appear less intense than the second.

Lauenstein's theory ran into difficulty when Pratt (1933a, 1933b) observed that more negative time-error occurred when a soft tone rather than silence was interpolated between the two comparison stimuli. Woodworth and Schlosberg (1954) reviewed the research on time-error and concluded that "time-error and assimilation effects are two separate phenomena."

Klein and Holzman believed that assimilation effects occurred both in Squares Test and time-error phenomena. They investigated the hypothesis that levelers are characterized by particularly strong assimilation tendencies and therefore demonstrate more visual time-error than sharpeners (1954). Leveling and sharpening were defined as follows:

> . . . a leveling tendency implies a dedifferentiation of the stimulus field, which may manifest itself either (a) through the reduction, within limits, of figure-ground distinctions, leading to difficulty in extracting embedded stimuli from larger, more cohesive stimulus organizations; or, (b) by assimilating new stimuli to a dominating organization, this leading to failure to recognize gradual changes in the stimulus field. In general, maximum simplicity of the cognitive field seems to be the consequence of leveling. In the _sharpening_ tendency, on the other hand, maximum complexity and differentiation are sought. Figure-ground distinctions are heightened, and this is expressed in sensitivity to embedded stimuli and in tendencies to exploit differentiation to the fullest extent.
>
> (Holzman and Klein, 1954, p. 106)

As was stated earlier, 43 _S_s, consisting of about an equal number of males and females (university undergraduates and students from a psychiatric aide school) took part in this study (see p. 20n.). The Squares Test was administered to groups of 3 _S_s seated so as to be equally distant and within a narrow range of angular displacement from the screen. The Squares Test procedure described on pages 6-9 was employed.

_S_s were classified as levelers and sharpeners on the basis of their PRA and AL scores (see pp. 11-13). In order to be classified as a leveler, a _S_ had to score in the lower half of the distribution of PRA scores and in the upper half of the distribution of AL scores. Sharpeners had to meet the reverse criteria. Eleven of the 43 Ss failed to meet these qualifications and were not tested further. Of the 32 Ss classified as levelers and sharpeners, only the top 27 per cent of each clasification was administered the time-error test.

Klein and Holzman permitted Ss to use whatever refinement of the inch scale they wished; scale preference, however, was considered "irrelevant to the aims of the study." (P. 108)

Time-error was determined under three conditions: (a) Condition I (dark), wherein no stimulus was interpolated; (b) Condition II (bright) where the interpolated stimulus was brighter than either of the two comparison stimuli; and (c) Condition III (dim) with an interpolated stimulus dimmer than the comparison stimuli. A five inch circular translucent screen under different illuminations served as the stimuli.

Levelers and sharpeners did not differ significantly under any of the three conditions. The difference in time-error between the bright and dim conditions, however, was significantly larger for the levelers than the sharpeners. This finding is somewhat equivocal because, as Klein and Holzman admitted, differences in time-error under the two conditions were obtained by algebraically summating

time-error scores for each S, a procedure which maximized the "distinction between groups." (P. 115)

Klein and Holzman (1954) concluded: "The present experiment tends to support the hypothesis of two system-principles of organization in cognitive functioning. In demonstrating their generality upon the TE phenomenon it further illustrates the relevance of 'neutral' psychophysical situations to personality theory." (P. 121) This conclusion hardly seems merited when one considers that (a) only small groups of extreme levelers and sharpeners were tested for time-error, (b) no relationship was found between leveling-sharpening and time-error within any of the three conditions, and (c) the significantly greater time-error difference between the bright and dim conditions exhibited by the levelers was based on a questionable quantitative technique.

The second time-error study (Holzman, 1954) concerned auditory and kinesthetic in addition to visual stimuli. Referring to the previous study (Klein and Holzman, 1954), Holzman stated that he and Klein "found that interpersonal consistencies in visual TE could be predicted from knowledge that a person's preferred approach to a situation was to level or sharpen differences and figure-ground gradients." (P. 377) However, as was just pointed out, the findings of the earlier study did not warrant such a strong conclusion.[1]

---

[1]The tendency to overemphasize the significance of rather weak and inconclusive findings has characterized much of the leveling-sharpening research. This has unfortunately led to an acceptance of leveling-sharpening by some as an established dimension of cognitive functioning.

In any case, Holzman hypothesized that levelers are characterized by larger time-error effects than sharpeners in all sense modalities. The same procedure was used in administering the Squares Test as in the previous study. One hundred and six Ss (50 men and 56 women) took part, levelers and sharpeners again being selected on the basis of PRA and AL scores. Thirty-one Ss were eliminated because they did not meet the criteria and once again only the extremes of each group (21 levelers and 22 sharpeners) were tested for time-error.[1]

Although the contribution to the total variance made by the two groups (in all sense modalities) was insignificantly larger than the error term, there was a significant interaction of groups x conditions. Levelers showed greater negative time-error under the condition of more intense interpolation and more positive time-error under the condition of the more intense interpolation than did sharpeners. The within groups variance for the kinesthetic data, however, was significant; and Holzman stated: "This finding is consistent with the view that there are other important determinants of variation in TE besides the cognitive controls of leveling and sharpening and the possibility also exists that there are various forms of the leveling-sharpening attitudes themselves." (Holzman, 1954, p. 387)

---

[1]Only 5 levelers and 5 sharpeners were available for the kinesthetic phase of the study.

Comparing the differences in time-error shifts between the bright and dim conditions for the levelers and sharpeners revealed that levelers were distinguished by greater shift than sharpeners in all modalities. Difference scores correlated significantly for the kinesthetic and auditory modes ($P < .05$) but not the auditory.

In summary, this study provided some evidence that: (a) extreme levelers are more prone to time-error shifts under conditions involving interpolated stimuli of different intensities than are extreme sharpeners, and (b) these tendencies demonstrate some consistency across certain sense modalities. It must be emphasized, however, that no relationship was found between leveling-sharpening and time-error as such and that difference scores have a tendency to inflate intracondition variability.

Considering the findings of both time-error studies, it must be concluded that very little relationship between Squares Test performance and time-error phenomena was evidenced. Even if the findings concerning leveling-sharpening and time-error difference scores are accepted as valid, no real support for the assimilation hypothesis was provided by this research.

Performance on the Squares Test and other serial tasks

One of the most interesting studies involving the Squares Test (Smith and Klein, 1953) did not directly concern leveling-sharpening. It is reviewed in detail because it represents the most thorough report of Squares Test response patterns available. The criterion task in this study was the Color-Word Test (Thurstone version of the Stroop Test). Ss were classified as cumulators, dissociatives, stabilizers, and non-consistent on the basis of their response time patterns. These groups were then tested on the following serial tasks: the Squares Test, a camoflaged faces test, an anticipation learning test, and the Gottschaldt Figures Test.

The Color-Word Test consists of 100 color names printed on a large card in different colored inks. Each color name is printed in an ink of different color than the color name. The task of the Ss is to ignore the color-names (the irrelevant stimuli) and report the colors of the inks. In this study, Ss' reading times for every 20 colors were recorded yielding a 5-point response time plot. Ss went through the series of colors five times (trials) and were classified on the basis of their most typical response time pattern if that pattern appeared on the first two trials.

Forty-eight undergraduates (an almost equal number of males and females) served as Ss. On the basis of their Color-Word Test performances, 9 were classified as cumulators (Cs), 8 as stabilizers (Ss), 12 as dissociatives (Ds), and 19 as non-consistent. Only the first three groups were tested further. See Table 1 for a summary of the findings.

Table 1. Summary of findings (Smith & Klein, 1953)

	Color-Word Test	Squares Test	Camoflaged Faces	Anticipation-learning Test	Gottschaldt Figures
CUMULATORS (Cs)	Increasingly slower reading times within trials	Progr. underest. until Subs. 5; then large shift in right direction. Used few size categories, large units.	Reported smallest no. of faces	Needed most trials; made most errors on last 2 of 4 series of pseudo-words.	Required most time (on complex figures).
STABILIZERS (Ss)	Relatively stable reading times within trials.	Overestimated until Subs. 3, then progr. underestimated. Used many size categories.	Reported more faces than Cs, less than Ds.	Made fewer errors on last 2 trials & had fewer trials than Cs.	Large decrease in solution time between 1st & last part of test; more learning effect than other groups.
DISSOCIATIVES (Ds)	Irregular resp. times (increas. & decreases)	Made most approp. size shifts. Used many categories, but large units	Reported most faces, but many were fabulated.	Same as Ss, but made fewer errors than Cs.	Solution times same as Ss but lower than Cs.

Smith and Klein (1953) summarized the characteristics of the three groups as follows:

(a) Cumulators--had difficulty discarding anticipatory set, became easily fatigued and appeared to be susceptible to subjective anchors. They adjusted late and in an exaggerated manner to changing situations and chose strategies which had seemed to work best in the past.

(b) Stabilizers--were able to maintain attention and were capable of discarding inappropriate sets, being generally attuned to the nature of the task.

(c) Dissociatives--were able to adopt appropriate sets over the short run but lacked the ability to sustain attention.

The possible relationships between leveling-sharpening and cumulating-stabilizing-dissociating were not discussed in this article. This is unfortunate since leveling and cumulating as well as sharpening and stabilizing would seem to be related categories of behavior. It should also be mentioned that dichotomous classifications of response patterns on the Squares Test can only be of limited value in view of the complex patterns which are reported in this study.[1]

---

[1]The patterns reported by Smith & Klein have been observed personally (Gilgen & Diehm, 1964).

Leveling-sharpening, tolerance vs. resistance to the unstable, and selected personality variables

Jeffreys (1953) investigated the relationship between leveling-sharpening and another hypothesized cognitive control, "Tolerance vs. resistance to the unstable." An apparent movement task served as the criterion for the latter. The Squares Test used was modified: Squares ranged from 2 through 14 inches and were whole numbers of inches in size. In addition, Ss were shown reference squares of 2 and 8 inches rather than 1 and 16 inches, and told to use whole number estimates. Four scores were derived: PRA, AL, L-S, and AS (see pp. 9-16 for scoring procedures).

The other tests administered were: (a) the Rorschach, (b) the Gottschaldt Figures Test, (c) a critical flicker fusion task, (d) an object sorting task, (e) a motor behavior test (writing "q"s forward and backward), and (f) a self-concept questionnaire.[1]

Sixty Ss (mostly males) were tested, and the following significant relationships between Squares Test scores and other measures reported:

(a) PRA with Gottschaldt Figures	.32
(b) PRA with self-concept subscores:	
(1) Social adaptability	.33
(2) Conformity	.25

---

[1]Developed by D. Sheer and based on D. W. Fiske's five concepts (1949).

(3) Confident self expression .30

(c) L-S with self-concept subscores:

(1) Social adaptability .44

(2) Emotional control .39

(3) Conformity .37

(4) Confident self expression .53

(d) AS with self-concept subscores:

(1) Inquiring intellect -.29

"Tolerance vs. resistance to the unstable" and leveling-sharpening were found to be unrelated concepts. The correlations between personality variables (as measured by the self-concept questionnaire) and leveling-sharpening, though, are the highest reported in the literature. In general, these findings are in agreement with those reported by Klein (1951), there being a tendency for sharpeners to be self-outward, socially adaptable individuals. Because of the procedural differences involved, comparisons between the two studies must be made guardedly.

<u>Leveling-sharpening, performance on other perceptual tasks and selected personality measures</u>

The search for the perceptual and personality correlates of leveling-sharpening was continued by Murney (1955). One hundred women (aged 17-20) were administered the unmodified Squares Test and the following other tasks: (a) Picture Completion and KOH Blocks subtests of the WAIS,

(b) gestalt completion, (c) concealed figures, (d) hidden pictures, (e) copying, (f) identical forms, (g) hidden words, (h) scrambled words, (i) weight discrimination, and (j) a Q-sort of personality items.[1] In addition, ACE scores were obtained.

Both PRA and AL scores, the indices of leveling-sharpening used in this investigation, failed to correlate significantly with any of the personality measures.[2] Leveling-sharpening, however, was significantly related to (a) hidden pictures ($P < .05$), and (b) copying ($P < .03$), though it must be noted that only the 20 most extreme Ss in each category were involved in the analysis. The most important finding was the failure of Squares Test scores to correlate significantly with the weight discrimination task which had been designed so as to provide a kinesthetic index of leveling-sharpening. In other words, Murney unlike Holzman (1954) did not find any evidence for the intermodal consistency of assimilation tendencies. A possible reason for the difference in findings might lie in the fact that Holzman tested both men and women while Murney used only female Ss.

---

[1]Based on the needs enumerated by Murray (1938).

[2]Lichtenstein (1962) also failed to find any significant relationships between leveling-sharpening and personality variables.

<u>Leveling-sharpening and cooperativeness</u>

Tear and Guthrie (1955) investigated the relationship between leveling-sharpening and cooperativeness. So many modifications of the Squares Test were introduced, however, that no meaningful comparison of the findings of their study and other research is possible. The squares they used ranged from 3 through 16 inches and <u>S</u>s were told to use whole number estimates exclusively (see Jeffreys, 1953). The index of leveling-sharpening was a new across-series score which was computed by taking the algebraic sum of the differences of estimate from the absolute size of the first square in each subseries. This score was unique in the sense that it reflected the absolute rather than the relative accuracy of a sample of square-size estimates; to use it as another indicator of leveling-sharpening, however, simply obscured the operational definition of an already nebulous concept.

The Squares Test was administered to 156 males (in small groups) with the <u>S</u>s in each group being asked to write down the names of those members in their group which they would classify as cooperative and those they could classify as uncooperative. Votes per <u>S</u> were summed and a simple majority determined the category into which each <u>S</u> was placed. Significantly more levelers than sharpeners ($P < .02$) were classified as cooperative, sharpeners tending to be described as competitive, non-conforming and self-centered.

Putting it another way, those Ss who were most accurate on the Squares Test were seen by their group members as being competitive, etc. Since high accuracy requires sustained attention, it is easy to see why individuals who were motivated to attend to the task might appear to others as being competitive.

A critical review of leveling-sharpening research and another investigation of perceptual and personality correlates

Krathwohl and Cronbach (1956) presented a critical review of Squares Test research and concluded that the variant techniques which had been employed concerned "initial set, the nature of the stimulus presentation, and the response unit used." (P. 308) No effort was made, however, to determine what role any of the above factors might play in test performance. Instead a study was conducted which concerned the relationships between various Squares Test scores and performance on the following tasks: (a) Gough's Rigidity Scale, (b) the ACE Psychological Examination, (c) the Minnesota Clerical Test, (d) an Object Aperture Test, and (e) the Gottschaldt Figures Test. In addition, the relationships among the various Squares Test scores derived were determined in order to evaluate their usefulness.

Ninety-nine freshmen enrolled in the architectural program (sex not given but presumed to be mostly males) served as Ss. Three scores were obtained: PRA, ESL and LSL. The following relationships were significant: (a) ESL and Gough's Rigidity score (.21, $P < .05$); (b) LSL and Gough's Rigidity score (.21, $P < .05$); (c) ESL and Object Aperture measures (.21, $P < .05$); (d) PRA and Object Aperture scores (-.34, $P < .01$); and (e) PRA and the Gottschaldt Figures (corrected) (.26, $P < .05$).

The negative relationship between PRA and the Object Aperture Test was unexpected since it meant that levelers tended to do better on this spatial relations test than sharpeners. Each item of this test "consists of a picture of a three dimensional object and of five two-dimensional apertures. The subject indicates the one aperture of the five through which the three-dimensional object could pass. The test appears to require spatial judgment, attention to detail, and accurate estimation of sizes." (P. 315) On the basis of assimilation theory and the usual descriptions of levelers and sharpeners, sharpeners would be expected to do better than levelers on this test. What happened possibly was that the sharpeners in trying to take into account all of the small differences among the three-dimensional figure and the five apertures, became somewhat anxious and tended to lose the overview which was necessary to match the correct aperture with the figure.[1]

---

[1]See Easterbrook (1959).

According to Krathwohl and Cronbach's evaluation of Squares Test scores, PRA and the two slope measures (ESL and LSL) are the preferred measures because they tend to reflect different aspects of the response patterns. On the other hand, AL, which was found to be based on highly correlated values, was judged to be unreliable.

A factor analytic study of cognitive controls

One of the most extensive reports concerning cognitive control research is a monograph which appeared in 1959 (Gardner, Holzman, Klein, Linton and Spence, 1959). This paper includes a thorough review of theory and presents the findings of a factor analytic study of the following cognitive controls: leveling-sharpening, tolerance for unrealistic experiences, equivalence range, scanning, constricted-flexible control, and field dependence-independence. In addition, some findings concerning cognitive controls and defense mechanisms were discussed.

Since cognitive control theory has already been reviewed, only the experimental aspects of this monograph are dealt with here. A rather atypical group of Ss was tested. The women,who were older than the men (mean ages--men 23.7, women 27.4) were mostly friends or acquaintances of the experimenters (wives of staff members, skilled secretaries, nurses, etc.). The men, on the other hand, were strangers to the experimenters, having been recruited

from a university and the staffs of local psychiatric hospitals. The experimenters "shared the impression that the women subjects were generally more sophisticated and had a broader range of cultural interests than the men." (Gardner _et al_., 1959, p. 19).

Several testing sessions per _S_ were involved. The Squares Test was administered to groups of five _S_s seated about 15 feet from the screen. The usual Squares Test procedure was employed with the exception that _S_s were given 2 minute rest periods after trials 40 and 75, and 1 and 18 inch reference squares were presented during the instructions. PRA, IE, and IE-reg. scores were derived.

Factor analysis isolated a factor for female but not male performance which was entitled "Leveling-Sharpening." Table 2 is an abbreviated and modified version of a table which appears in Gardner _et al_. (1959), p. 101. The authors state: "Although solution time for Cluster II of the Embedded Figures Test had a moderate loading on Factor II (.38), the actual scores of extreme levelers and sharpeners on this measure did not differ significantly. The factor loading was apparently due to the correlation between Factors I and II and would disappear from Factor II if an oblique rotation were made. This measure was therefore not interpreted in the discussion of Factor II." (P. 101)

Table 2. Factor II (Females) (Gardner *et al.*, 1959) Leveling-Sharpening

Loading	Score	Performance of *Ss* with high factor scores
.71	Size Estimation: Cons. error, Perc.	Overestimation
-.68	Squares Test: PRA	Inaccuracy
.67	Free Assoc.: Mean unit length	Long units
.65	Aniseikonic Lenses: Log Recog. time	Long Recog. times
.65	Kinesthetic Time Error: Assimilation	Much assimilation
-.50	Free Assoc.: Percent. Distant "Dry"	Close associations
.47	Squares Test: IE-reg.	Extreme lagging
.45	Color-Word: Reading time, colors alone	Slow reading
.44	Size Estimation: Variability, Memory	High variability
-.38	Free Assoc.: Percent. "Home" responses	Few Home responses
.38	Embedded Figures: Mean log time, Cl. II	Long solution time

An inspection of the intercorrelation matrix makes it rather difficult to understand this reasoning because PRA and the solution time of Cluster II of the Embedded Figures Test correlated -.28, while PRA correlated as follows with the other scores loading on Factor II: (a) Size Estimation, Constant Error, Percept. -.29; (b) Free Assoc.: Mean length of unit -.19; (c) Ansiseikonic lenses, log recognition time -.58; (d) Kinesthetic time-error, assimilation -.09; (e) Free Assoc., percentage distant, "Dry" .29; (f) Squares Test, increment error-Regr. -.41; (g) Color-Word, reading time, colors alone -.12; (h) Size estimation, variability, memory -.16; (i) Free Assoc., percentage "Home" responses .37.

Only three of the above relationships [(c), (f), and (i)] are substantially greater than the -.28 correlation existing between PRA and the Embedded Figures (Cluster II) solution times, while four of the correlation [(b), (d), (g), and (h)] are considerably less than .28 and two of them [(a), and (e)] are about the same. If the Embedded Figures score is not interpreted in the discussion of Factor II for the reason given by the authors, then six other measures should likewise be eliminated from the interpretation.

According to Gardner *et al.* (1959), this factor is "relevant to situations involving temporal patterning of stimuli. It does *not* represent discrimination of differences per se. Rather, it pertains to differentiation in memory organization as a function of the extent to which successive stimuli assimilate to each other." (P. 100) This definition of leveling-sharpening remains within the framework of assimilation theory but limits the concept to characteristics of *serial* response patterns. This is somewhat of a departure from the earlier conceptualization which extended the definition to response dispositions which manifest themselves in any situation wherein *S*s have to identify or judge stimuli embedded with complex configurations.

A "person cluster" analysis was also performed, but the clusters were so small (3 to 6 *S*s per cluster) that the findings are of little value.

Leveling-sharpening and repression

As was mentioned before, the Gardner et al. (1959) monograph also reported findings concerning the relationships between certain cognitive controls and defense mechanisms. Included were the results of a study dealing with leveling-sharpening and repression. The following aspects of Rorschach protocols served as indicators of repressive tendencies: (a) certain expressive and phobic verbalizations, (b) notable lack of specificity, (c) failures, (d) poor integrative efforts, (e) childlike materials, (f) symbolic content without awareness of the significance of these responses, (g) little content variety, (h) less than 25 responses, (i) fewer than two human movement responses, and (j) emphasis on color (unusual predominance of CF and C). It was pointed out that "in every instance the evaluation process went beyond the use of these indicators." (P. 131) No details were given.

Repression was considered to be the preferred mode of defense of 8 of the 30 men. Six of these were levelers (scores in the upper half of the leveling-sharpening index scores[1]; this finding, however, was not significant. Eleven of the 30 women were judged to prefer repression as their primary method of defense, 9 of the 11 being relatively

---

[1]The leveling-sharpening index score consisted of the average of the standard scores for PRA and IE.

extreme levelers ($P < .03$).[1] As Spivak (1964) has pointed out, some of the Rorschach indicators of repressive tendencies such as the non-specificity of a response and the number of responses may just as easily be indices of leveling, and therefore it is not possible to conclude with much assurance that a relationship between leveling (the cognitive control) and repression (the defense mechanism) has been demonstrated.

Lichtenstein (1962), using the difference between the Hysteria and Psychasthenia subtest scores (MMPI), as the measure of repressive tendencies, found no relationship between such tendencies and leveling-sharpening. Lachman, Lapkin and Handelman (1962) in another study did find that Ss who recall numerous dreams (and are therefore presumed not to employ a primarily repressive defense system), tend to score toward the sharpening end of the continuum. The correlation between leveling-sharpening and frequency of dream recall, however, was not significant. Spivak (1964) found that levelers tended to report seeing a Blacky picture to which they had been sensitized[2] less frequently than

---

[1]Holzman and Gardner (1959) replicated this study using 10 extreme levelers and 10 extreme sharpeners (all women); all extreme repressors were also levelers, but all levelers were not repressors.

[2]Ss were first given the picture to look at and told that it showed "Blacky" masturbating. They were then instructed to think of themselves in similar situations. Ss were tested tachistoscopically, with this picture being flashed along with 3 other neutral "Blacky" pictures, and told to indicate which one they saw.

sharpeners. This finding was interpreted as meaning that levelers, more often than sharpeners, repress guilt-producing stimuli. Not seeing a threatening stimulus, however, seems to be more an instance of denial than repression.

In summary, although no strong relationship has been established between leveling and repression, there is some evidence that leveling may have something to do with the preference for a certain style of communication. Ss who use the same size estimate for two or more squares on the Squares Test and therefore score low on PRA, and who also give few responses on the Rorschach and recall few dreams, may be characterized by a general tendency to limit the details of their communication. If this is the case, leveling may have very little to do with assimilation tendencies but may instead result from S's preferences for answering psychologists' probings with rather minimal, undetailed responses.

Leveling-sharpening, intellectual abilities and certain aspects of personality

Gardner, Jackson and Messick (1960) administered a large battery of tests to a group of 63 women in order to investigate the relationships among cognitive controls, intellectual abilities and certain aspects of personality. PRA scores failed to correlate significantly with any of the ability test scores, but low significant correlations were found between PRA and two of the Edwards Personal Preference scores (Deference -.27, and Order -.28).

Since the relationships are extremely minimal, no interpretation was given.

Leveling-sharpening and the learning or memory of verbal materials

Gardner and Lohrenz (1960) reported that a group of five female levelers (PRA scores), when passing a story along from one to another, used fewer total and correct themes than a comparable group of sharpeners. In another study, Gardner and Long (1960a) had groups of 10 extreme levelers and sharpeners (women) learn two lists of words which were presented via a memory drum. Ss were given eight learning trials per list and then told to write down in proper order all of the words that they could remember from List 1. After this they were instructed to do the same for List 2. Although the two groups did not differ significantly in amount recalled, levelers made significantly more backward errors (offering a word which would have been correct at an earlier point in the list) than the sharpeners. In addition, the levelers made more intrusion errors that were displaced more than one place from the correct position.[1]

---

[1]Another similar study (Nicks, 1960) concerned leveling-sharpening and retroactive inhibition. It was found that: (a) RI increases as the tendency toward leveling increases, and (b) there is no relationship between leveling-sharpening and learning ability. These findings generally support those reported by Gardner & Long (1960).

The third study (Holzman and Gardner, 1960) dealt with leveling-sharpening and memory organization. Forty-one women (16 levelers and 25 sharpeners classified on the basis of PRA scores) were told to relate the story "The Pied Piper," a tale with which all were familiar. Ss' stories were analyzed for the number of correct thematic elements included (mention of the town of Hamlin, etc.). It was found that the sharpeners' stories contained more total and correct themes than the levelers' stories.[1] This finding is compatible with the hypothesis that leveling may be one manifestation of a general tendency to communicate using a minimum of detail, which was discussed with reference to the findings of the leveling-sharpening-repression studies.

In conclusion, there appears to be some relationship between leveling-sharpening and the amount of detail individuals either attend to or prefer to communicate with. Again the assimilation hypothesis need not be invoked to account for this finding and whatever theoretical speculation is ventured must be limited to female performance since only women were used as Ss in all of these studies.

---

[1]Livant (1964) inspected the protocols of the Ss involved in the Gardner & Lohrenz (1960) and Holzman & Gardner (1960) studies and found that sharpeners used all of the following grammatical connectives more than levelers: "but, so, if, since, because, or, nor, therefore, when."

Leveling-sharpening and word-rating styles

Gilgen and Diehm (1964), using a word rating task similarly structured to the Squares Test, found no significant relationships between leveling-sharpening (PRA and AL scores) and word-rating styles. It was this research which led to a general dissatisfaction with the theoretical definition of leveling-sharpening in terms of assimilation theory. This dissatisfaction inspired the present research.

Summary

The review of literature concerning leveling-sharpening indicates that:

(a) There is no clear operational definition of leveling-sharpening. The numerous modifications of Squares Test structure, administration and scoring make it impossible to determine just what aspect of performance is supposed to be a manifestation of this hypothesized cognitive control principle.

(b) The theoretical definition of leveling-sharpening has tended to change to conform with whatever scores correlated most highly with Squares Test measures. Whenever, for example, a significant correlation has been obtained between PRA scores and performance on the Gottschaldt Figures Test, leveling-sharpening was supposed to have something to do with the ability of _S_s to extract stimuli embedded in complex configurations (Klein, 1951; Holzman and Klein, 1954). At other times when significant relationships only occurred between Squares Test measures and other _serial_ task scores, the concept leveling-sharpening was said to concern only assimilation effects occurring among the memory traces of _successive_ stimuli (Gardner _et al._, 1959).

(c) There appear to be relationships between leveling-sharpening (operationally defined in terms of PRA or combination PRA-AL scores) and: (1) the amount of retroactive inhibition exhibited in serial learning tasks, (2) the tendency to use detailed communication, and (3) the preference for repressive tendencies. These relationships, however, have only been demonstrated for female _S_s.

(d) There is some evidence that high PRA scores are related to self-outwardness and competitiveness. Individuals who really assert themselves appear to obtain higher PRA scores than those who are less ego-involved in the Squares Test.

(e) Most research has been unsystematic. In spite of Klein and Schlesinger's assurance (1949) that the bridge between cognitive controls and personality variables was not to be built of correlation coefficients, almost every study has been correlational. Moreover: (1) small groups have generally been used; (2) replications have been rare; (3) most positive findings have occurred when only female Ss were tested; and (4) the roles of various aspects of the Squares Test such as the instructions, the response units, the sizes of the squares, etc. have not been investigated.

Current status of the leveling-sharpening concept

In spite of the theoretical lack of clarity and unconvincing empirical support, the concept leveling-sharpening is unfortunately being presented by some current sources as a well-substantiated dimension of cognitive functioning. Levine and Spivack (1964) state:

> Leveling-sharpening is a reliable dimension of individual differences, in which the leveler, in contrast to the sharpener, diminishes the distinctiveness between a sequence of perceptual stimuli, and between elements in memory. The memories assimilate to each other, and in the end are more vague. . . . The leveling-sharpening dimension refers to a pervasive set of individual differences, and may even have its analogues in individual differences in characteristics of the organization of the nervous system. . . . The leveling-sharpening dimension may have many response indicators, but its basic property is the tendency of elements of the memory to assimilate. (P. 8)

Likewise Silverman (1964) writes:

> Recent studies of individual consistencies in cognitive behavior have emphasized the importance of enduring response dispositions which are relatively invariant over broad ranges of laboratory tasks and

> environmental conditions (Gardner *et al.*, 1959; Klein, 1958). These response dispositions, termed cognitive control principles have been conceptualized as more or less automatic adjustments to the environment which emerge in the course of development and which are shaped by both constitutional and experimental factors. (P. 354)

Referring to sharpening tendencies Silverman states "For example, some people have an underlying preference for making fine discriminations about nearly everything." (P. 354)

Gardner (1964) after citing the findings of some of the research concerning cognitive controls concluded that:

> Once again it is apparent that cognitive controls are enduring patterns of cognitive functioning that mediate the expression of particular intentions when the individual is confronted with particular stimulus conditions.
>
> (Gardner, R. W., in C. Scheerer (1964), p. 156)

Although other examples of this point of view could be presented, suffice it to say that leveling-sharpening has over the past fifteen years or so gained the status of a well-substantiated dimension of cognitive structure. The review of the literature does not support this assertion and systematic investigations of the Squares Test are clearly called for. The present study represents a step in this direction.

EXPERIMENT I

Experiment 1 of this study was exploratory. No specific hypotheses were tested and the primary purpose was to determine what aspects of the Squares Test enter into the frames of reference Ss develop in estimating the sizes of the squares in the series. It was decided that each S would take two forms of the test during one testing session. The use of a single session eliminated the possibility of Ss discussing their performances between testing sessions. Since the length and repetitiveness of the test created a boredom problem, a short form of the test (the first six subseries) was used.[1]

Method

Subjects (Ss)

Thirty-three undergraduate volunteers (14 men and 19 women) were tested. Class credit was given for volunteering.

Procedure

Ss were individually tested in a room completely darkened except for a small lamp which illuminated the recording sheet. A 35 mm. slide projector (500 watt bulb),

[1]According to Gardner (1964) this is the best short form of the test. (Personal communication)

automatically timed by two Hunter Timers,[1] projected the series of squares onto a black cloth screen located 10 feet from the S. Black cloth was draped from the walls and laid in front of the screen in order to reduce size cues.

The Squares Test was administered in the usual manner to one S at a time (see pp. 6-9). One group of 15 Ss (Group I) was given the same form of the test twice so that the short term reliability of the test could be determined. Interviews with the Ss in this group indicated that the instructions were ambiguous because some thought size estimates were to be made to the nearest whole number while others used fractions. A number of Ss asked if they were to use only whole number estimates after the instructions had been read to them. Since Ss who used fractions tended to score higher on PRA than those using all whole numbers, it was decided that this score might be a function, in part, of the way Ss interpret the instructions.

The Ss in Group II (N = 10), therefore, were explicitly told (the second time they took the test) that none of the squares were exact whole numbers of inches in size. The instructions used were the same as before except that the following phrase was added: "None of the squares that you will see are exact whole numbers of inches in size."

---

[1]See Appendix D for a schematic drawing of the wiring arrangement.

The other 8 Ss who participated in the study took the Squares Test once and then were administered a variety of other exploratory tasks. The Squares Test data obtained from this group was included in the analysis of the relationships between PRA and LSE scores.

Results and Discussion

Short-term reliability

The ranking of the PRA scores changed little from test to retest as indicated by a rank-order correlation of .94. This short form of the test, therefore, appears to have good reliability provided that no feedback occurs between administrations. Despite the high reliability, PRA dropped an average of 5.7 points when the test was given the second time with all Ss experiencing a decrease. This probably indicates that Ss did not try quite as hard during the retest. This assertion receives some support from the questionnaire data which revealed that many Ss became rather impatient when told that they were to estimate the sizes of another series of squares.

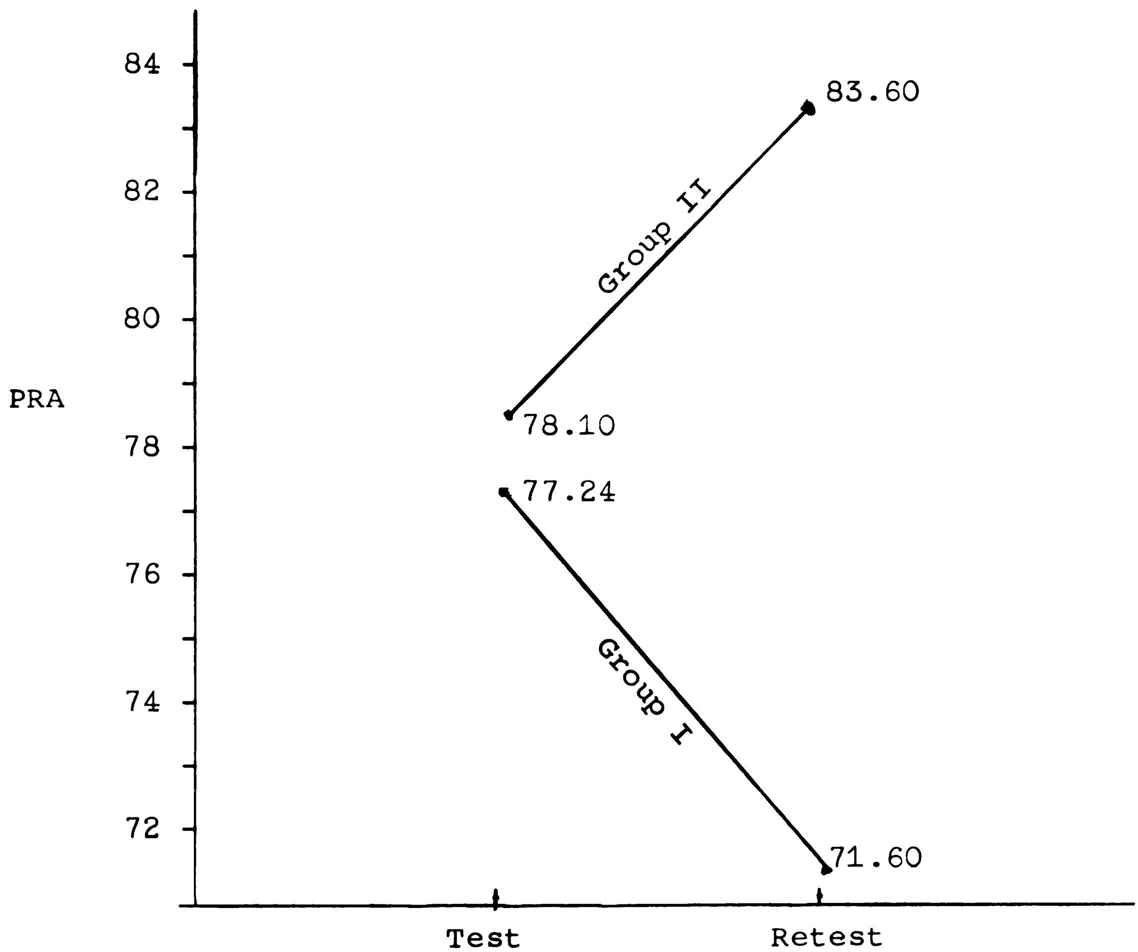
The effects of the Fractional Instructions

Nine of the 11 Ss who took part in this part of the study (Group II) achieved higher PRA scores when told that none of the squares were whole numbers of inches in size.

This was quite a dramatic effect in view of the fact that PRA decreased when the test was readministered using the regular instructions (Group I). Figure 2 illustrates the performances of the two groups.

Figure 2

PRA scores for test and retest



Additional findings

An inspection of Ss' response curves on the regular short form of the Squares Test indicated that considerably more men than women overestimated increasingly as the series progressed. Using each S's largest size estimate (LSE) as an index of slope, it was found that the mean LSE-females (9.3 inches) was significantly higher than the mean LSE-males (7.5 inches), [$P < .05$, $df = 31$, ($t = 2.22$)]. Furthermore, it was found that LSE and PRA scores correlated significantly for the women (.658, $P < .002$), but not for the men (.21). These findings are particularly interesting because sex differences in Squares Test performance have not previously been reported. Gardner et al. (1959) isolated a factor for female performance only which was entitled "Leveling-Sharpening," but no sex differences as such were found on the Squares Test.

When the performances of the extreme[1] levelers (low PRA) and sharpeners (high PRA) were analyzed, the findings were even more dramatic. As Table 3 indicates, the female sharpeners' mean LSE (12.5 inches) was significantly higher than the mean LSE values of the other three groups. Most female sharpeners, in other words, were overestimators as well, while the other Ss were quite accurate.

The leveling-sharpening literature says nothing about overestimation tendencies; yet in a previous study (Gilgen & Diehm, 1964), involving the full-length Squares Test, overestimation was observed. Such effects cannot be attributed solely to the fact that the short-form of the test was used.

---

[1]Ss falling in approximately the top and bottom 27 per cent of the distribution were compared.

Table 3. Performances of Extreme Levelers and Sharpeners (means)

Squares Test Score	Males (N = 6)		Females (N = 8)	
	Levelers	Sharpeners	Levelers	Sharpeners
PRA	69.6	86.6	69.4	84.8
LSE (inches)	6.7	7.7	6.0	12.5

If assimilation theory is invoked to account for these findings, then it would be proposed that the estimates of _S_s who overestimate are characterized by particularly weak assimilation effects between the stimulus being judged and the memories of the previous and smaller stimuli in the series. In the light of the information obtained from the _S_s during the interviews following the Squares Test, however, it seems more appropriate to attribute overestimation to instructional set or expectancy.[1]

It will be recalled that _S_s are told that the squares range somewhere between 1 and 16 inches, whereas the actual range of sizes only extends from 1.2 through 6.6 inches. It is not surprising, therefore, that under the cueless condition of a dark room, some _S_s estimate the 6.6 inch square as being 15 or 16 inches.

Operating in conjunction with a range expectancy is the ambiguity concerning the scale refinement _S_s are to use.

---

[1]The role of instructional set in judgment has been well established (Vernon, 1937; Gibson, 1941).

As was mentioned before, some Ss assume that estimates are to be made in whole numbers, while others find the instructions unclear and ask if they may use fractions. If a person expects to see squares ranging from 1 through 16 inches, and at the same time believes that estimates are to be made using only integers, the chances are good that overestimation will result. This is particularly true if the S has difficulty remembering what the 1 and 16 inch reference squares looked like.

If the above assumptions are correct, then it appears that female sharpeners (high PRA scores) are particularly susceptible to instructional set. The literature review concerning sex differences which is presented in the next section was conducted in order to determine whether or not analogous findings had been reported in other similar research.

Summary

The purpose of this study was to identify some aspects of the Squares Test which affect PRA scores. The short form of the test used, was found to have a short-term reliability of .94, although PRA dropped an average of 5.7 percentage points between test and retest. It was found that there is ambiguity in the instructions concerning the refinement of the inch scale which Ss can use in making judgments. Many Ss believe that only whole number estimates may be used.

When a group of Ss was told that none of the squares are exact whole numbers of inches in size, 9 of the 11 Ss in the group achieved higher PRA scores than when they had previously taken the test with the regular instructions. It was also found that some Ss, particularly female sharpeners, overestimated quite markedly. This was attributed to the discrepancy between the instructional information given Ss concerning the range of the squares to be judged and the actual sizes of the squares, in conjunction with the ambiguity of the instructions with regard to the inch scale refinement permitted in the judgments. Because of the small group of Ss involved, a replication was called for; this, in part, was the purpose of Experiment II.

THE REVIEW OF THE LITERATURE ON SEX DIFFERENCES

On the Squares Test

As was mentioned earlier, sex differences have not been reported for performance on the Squares Test. Gardner *et al*., (1959) isolated a factor for female but not male performance which they called "Leveling-Sharpening," but sex differences in Squares Test scores per se were not obtained.

Sex Differences on Other Perceptual and Cognitive Tasks

The review of the literature indicated that sex differences occur in the following types of experimental tasks or situations:

(a) Size and length estimations

(b) Evaluations made under disorienting conditions

(c) Tests of scientific and mathematical achievement

(d) Tests involving a spatial factor

(e) Certain subtests of the Wechsler Adult Intelligence Scale (WAIS)

(f) Tests of persuasibility and suggestibility

Some of these classifications overlap. For convenience, the discussion of sex differences and their possible relevance to the Squares Test has been organized around them.

Before discussing the literature, it is important to bear in mind that most of the research in this area is based on tests of performance which, like the Squares Test, are characterized by repeated modifications both in procedure and scoring techniques. As Cronbach (1960) has stated:

> We cannot point to any one 'official' version of a performance test, and we have no manual systematically reporting its technical qualities. Each investigator modifies a test for his own purposes and reports such findings about reliability and validity as emerge from studies of his own theories. (P. 551)

Estimations of size and length

Thompson (1903) reported that males were somewhat more accurate than females in discriminating the size differences among a set of white squares individually mounted on black cardboard. Comparing squares 19.5, 19.0, 18.5, and 18.0 mm. on a side with a 20 mm. standard, the males, more often than the females, were able to discriminate the squares closest in size to the standard.

Thompson (1903) also cited a study by MacDonald (1897-1898) in which it was reported that boys were more accurate than girls in estimating the lengths of lines; males, therefore, appeared to be more proficient than females in the use of the inch scale.

On the basis of these findings one would expect females to be somewhat less accurate in an absolute sense on the Squares Test.

Evaluations under disorienting conditions

Sandstrom (1953) found that women were less accurate than men when asked to point at a luminous spot of light in a completely darkened room. The women appeared generally less able to orient themselves in such a cueless situation than the men. Witkin's field dependence-independence research, which is discussed next, also indicates that women, on the average, seem to be more disoriented in dark-room experiments than men.

Witkin et al. (1954) conducted a series of studies involving the Rod and Frame Test, the Tilted Chair-Tilted Room Test and a modified version of the Embedded Figures Test which concern the effects of disorienting or distracting stimuli upon judgment. The concept field dependence-independence was coined in order to account for the differences in the degree to which Ss rely on visual cues, whether relevant or irrelevant, in making their judgments. In the Rod and Frame Test, for example, Ss whose verticality judgments are influenced by the tilted frame are classified as field-dependent, whereas Ss who are better able to disregard the irrelevant (or disorienting) frame, are categorized as field independent.

Results of these studies indicate that women, on the average, are more influenced by the frame in the Rod and Frame Test and more disoriented in the Tilted Chair-Tilted Room Test than men. An interesting additional finding was that women do as well as men on the tilted chair situation when they close their eyes. Women seem to be able to make use of kinesthetic cues as well as men, but have more difficulty in disregarding misleading visual cues. For this same reason, possibly, some females are misled by the visual information (the 1 and 16 inch squares) which is given during the instructional period of the Squares Test.

As was mentioned previously, a modified form of the Embedded Figures Test (EFT) has also been used as a test of field dependence-independence. Scores on this test usually correlate significantly with performance on the Rod and Frame and the Tilted Chair-Tilted Room tests, those Ss having the most difficulty in disregarding the irrelevant visual cues in the two verticality tasks, also finding it difficult to locate the simple figures embedded within the more complex ones.

Although Gardner et al. (1959) found that leveling-sharpening and field dependence-independence appeared to be independent cognitive control principles, several studies (Klein and Holzman, 1950; Jeffreys, 1953; Murney, 1955; and Krathwohl and Cronbach, 1956) have found that levelers (low on PRA) tend to have more difficulty in extracting the simple from the complex figures in various versions of the Embedded Figures Test (EFT).

Other investigators have found that women tend to do less well on this test than men (Thurstone, 1944; Witkin et al., 1954; and Fink, 1959). In addition, there is some evidence that good performance on the EFT is a function of motivation. Wertheim and Mednick (1958) reported that time to locate the embedded figures correlated -.40 ($P < .01$) with n.Ach. Fink (1959) found a correlation of -.72 ($P < .01$) between performance on the EFT and Psychasthenia (MMPI). Furthermore, sex differences on the EFT also appear to involve motivational factors with Fink reporting that Interest (Masculinity-Femininity) scores (MMPI) correlated .303 ($P < .01$) with EFT measures--Ss scoring on the masculine side tending to do better than those scoring on the feminine side.

In summary, therefore, women tend to do less well than men on tasks involving misleading or irrelevant visual cues. This sex difference in performance appears to be due in part to the fact that women have less need to achieve than men.

Tests of scientific and mathematical achievement

"One other type of sex difference which shows up with great clarity at the higher age and education levels has to do with achievement in science." (Tyler, 1956, pp. 254-255) Without going into all of the specific studies cited by Tyler, it seems quite evident that at the high school and college levels, males, on the average, do better on tests of science (and mathematics) achievement or aptitude than females. The consistency of this finding is well substantiated by the almost universal superiority of male entering freshmen on the SAT Mathematical aptitude scores (Cass and Birnbaum, 1964).

According to McAndrew (1943) male proficiency in this area seems to be a function of our educational system because a sex difference does not appear until about adolescence. Furthermore, it appears that this difference is again in large part due to motivational factors. Wechsler (1950), for example, cites a study by Alexander (1935) in which "interest or concern" loaded .74 on a "success in science" factor while "g"[1] only loaded .36 on this factor.

A study by Pemberton (1952) found that Ss who performed well on the EFT tended to agree with the following items of a personality questionnaire: (a) "I am interested in physical sciences," and (b) "I have high theoretical interests and values." This finding is, of course, compatible with the reports of the research previously discussed, interest or motivation being related to performance on the EFT.

Since some researchers have found a relationship between EFT and Squares Test performance, with levelers tending to do less well on the EFT than sharpeners, and EFT performance is in part a function of the motivation of the Ss, it seems reasonable to infer that levelers are less motivated to perform maximally on the Squares Test than sharpeners. This prediction is tested in Experiment II of this study.

---

[1]Spearman's "g" or general factor (educative).

Tests involving a spatial factor

Tyler (1956) states: "in judgment and manipulation of spatial relationships, a consistent (male) superiority has been demonstrated." (Author's parenthesis) P. 253. This superiority usually manifests itself in form board tests, maze and puzzle tests, and tests involving mechanical ability. Bennett and Cruikshank (1942), for example, found that males score higher than females on all 60 items of the Mechanical Comprehension Test by Bennett. This is an interesting finding since this test requires Ss to observe pictures in order to answer questions about mechanical relationships. As in the research concerning evaluations made under disorienting conditions, women in this task appear to have more difficulty than men in making judgments based on visual cues.

Tyler (1956) also cites a study by Sweeney (1953) wherein males were found to be superior to females in situations requiring "restructuring." Some tasks, in other words, involve problems which can be solved only if previous strategies are discarded and women appear to do more poorly on these problems than men. It was reported that sex differences appeared even in groups that "had been equated for general intelligence, verbal ability, mathematical ability, relevant knowledge, and various backgrounds." Taylor (1954) presents evidence that the difference is primarily due to a matter of attitude toward such tasks and is susceptible to training (Tyler, 1956). Luchins and Luchins (1959) also reported that on every item of the Einstellung Test, women showed more inability to restructure their frames of reference (E effects) than men although the differences did not reach statistical significance. Luchins and Luchins concluded that the teacher-student relationship builds a tendency for females "to have blind confidence in what the teacher says, etc."

Progressive overestimation on the Squares Test can be viewed as a manifestation of the inability to restructure or discard a frame of reference which is no longer appropriate. If this is the case, it would be expected that women would have more difficulty than men in disregarding the information concerning the range of square sizes provided them during the instructions. This would lead them to overestimate.

A study of Krathwohl and Cronbach (1956) provides further evidence that performance on the Squares Test is related to performance on tasks involving a spatial factor. Using the Object Aperture Test (OAT), an unpublished spatial relations test, it was found that OAT performance correlated -.34 ($P < .01$) with PRA. The authors concluded that the negative correlation was "unexpected and uninterpretable." If the PRA score is considered an index of motivation rather than ability, this finding might be accounted for by assuming that <u>S</u>s who attempted to pay attention to all of the small details in the OAT drawings tended to lose the overview necessary for the correct solutions to the problems (see pp. 34-35 for further discussion of this study).

To recapitulate, women, on the average, tend to be slightly inferior to men on tasks involving a spatial factor or the restructuring of frames of reference. This may account, in part, for the finding in Experiment I that women overestimate more frequently and to a greater extent than men. Furthermore, PRA seems to be an index of motivation rather than a measure of ability.

Subtests of the Wechsler Adult Intelligence Scale (WAIS)

Although standardization data on the WAIS show small positive sex differences on the Full Scale scores in favor of females, the differences are too small to warrant separate sex norms (Wechsler, 1958). When the individual subtests are analyzed, however, significant sex differences are apparent. Females, on the average, do better on (a) Similarities, (b) Vocabulary, and (c) Digit Symbol, whereas males tend to do better on: (a) Information, (b) Comprehension, (c) Arithmetic, (d) Picture Completion, and (e) Block Design. Wechsler shows how a Masculinity-Femininity score (MF), analagous to such measures as the Interest score of the MMPI can be derived from WAIS subtest scores. It will be recalled that Fink (1959) found that Ss high on masculine interests (MMPI Interest scores) tended to do better on the EFT than Ss with more feminine interests. Since performance on the EFT and Squares Test have in some studies been found to be related, it might be expected that WAIS subtest scores might also correlate significantly with certain aspects of Squares Test performance.

Further evidence for the link between the concepts of intelligence, interest and ability in spatial relations tasks is provided by the fact that performance on the EFT correlates from .35 to .60 with such tests as Block Design, Number Series, and Thurstone's tests of spatial factor (Cronbach, 1960).

Wechsler (1950) points out that when matrices of intelligence tests are factored, less than 60 per cent of the total variance is accounted for, and that the greater the number of tests included in the analysis, the smaller the per cent of the variance accounted for. Furthermore, the unaccounted for variance is believed to be due to such non-intellectual factors as curiousity, interest and the need for achievement. The review of the research in this area suggests that such factors also play a role in Squares Test performance.

Tests of persuasibility and suggestibility

There is good evidence that female high school and college students are more persuasible than their male contemporaries (Janis and Field, 1959). It has also been found that individuals who change their opinions under the pressure of authority tend to do less well on the EFT than those who are able to resist authoritative influence (Linton and Graham, 1959). Since levelers have been found to perform more poorly on the EFT than sharpeners, it might be inferred that the former are more persuasible than the latter. If being persuasible is interpreted as being non-assertive or passive, then such an inference is compatible with the hypothesis that S who assert themselves or try on the Squares Test score higher on PRA than those who are less motivated.

Studies of suggestibility also indicate that women tend to be more suggestible than men (Eysenck and Furneau, 1945; Stukat, 1958). Likewise, Tuddenham (1958) reported that on an Asch type study, women were more susceptible to the influence of a simulated norm than men. In addition, as was mentioned earlier, Luchins and Luchins (1959) found that on every item of the Einstellung Test, women were more influenced by sets or frames of reference adopted early in the test.

These findings show that young women are both susceptible to suggestion and yet slow to give up a frame of reference once it is adopted. If this is the case then it would be expected that female Ss are more susceptible to instructional set on the Squares Test than men, and would, as was found in Experiment I, tend to overestimate more.

Summary

Sex differences have been found in many different perceptual and cognitive tasks. Such differences do not appear to be caused by differences in the anatomical or physiological characteristics of the nervous systems of the two sexes, but rather seem to be manifestations of differential training and cultural emphasis. Such factors as interest, motivation, and suggestibility seem to play major roles in male-female performance differences although it is not to be denied that training and experience are also involved. Certain of the findings in this area suggest that interest, motivation and suggestibility might also play a part in Squares Test performance, with low interest or motivation resulting in low PRA scores and susceptibility to suggestibility leading to the inability to discard or readjust a set or expectancy induced by the instructions.

EXPERIMENT II

The findings of Experiment I suggested that instructional set and sex differences are involved in Squares Test performance. The primary purpose of Experiment II was to investigate systematically the roles these variables play. Since the review of the literature indicated that sex differences on perceptual and cognitive tasks are to some extent a function of such factors as interest and motivation, these factors were also investigated.

The specific hypotheses tested were as follows:

(a) Hypothesis I--Size estimates are a function of the information concerning the size range of the squares to be judged, given _S_s. _S_s who are told that the range of square sizes lies somewhere between 1 and 16 inches make larger estimates than _S_s who are told that the range of sizes lies between 1 and 12 inches. Likewise, _S_s who are informed that the squares range somewhere between 1 and 7 inches, make the smallest size estimates of all.

(b) Hypothesis II--Positional Ranking Accuracy (PRA) is a function of the information given _S_s concerning the inch scale refinement they may use in making their judgments. _S_s who are explicitly told that they may use fractions, score higher on PRA, on the average, than those given the regular Squares Test instructions which are ambiguous.

(c) Hypothesis III--Women are influenced more by the instructional range information than men.

(d) Hypothesis IV--PRA and the largest size estimate (LSE) made are positively related for female, but not male, performances under conditions 1-12 (Regular) and 1-16 (Regular), see page 75.

(e) Hypothesis V--High interest and motivation on the Squares Test are positively related to PRA scores.

Method

Subjects

Three hundred undergraduate volunteers (150 men and 150 women) took part in the main part of the study. Another group of 50 Ss (25 men and 25 women) was subsequently tested under a "No-information" control condition.

Setting and equipment

The experiment was conducted in a (12.5' x 18.5') windowless room. Two booths were constructed to permit the testing of 2 Ss at a time. The experimenter (E) sat in a projection booth located between the two Ss' booths. This arrangement eliminated communication between the Ss during testing (see Figures 3 and 4).

Figure 3

Top view of experimental room.

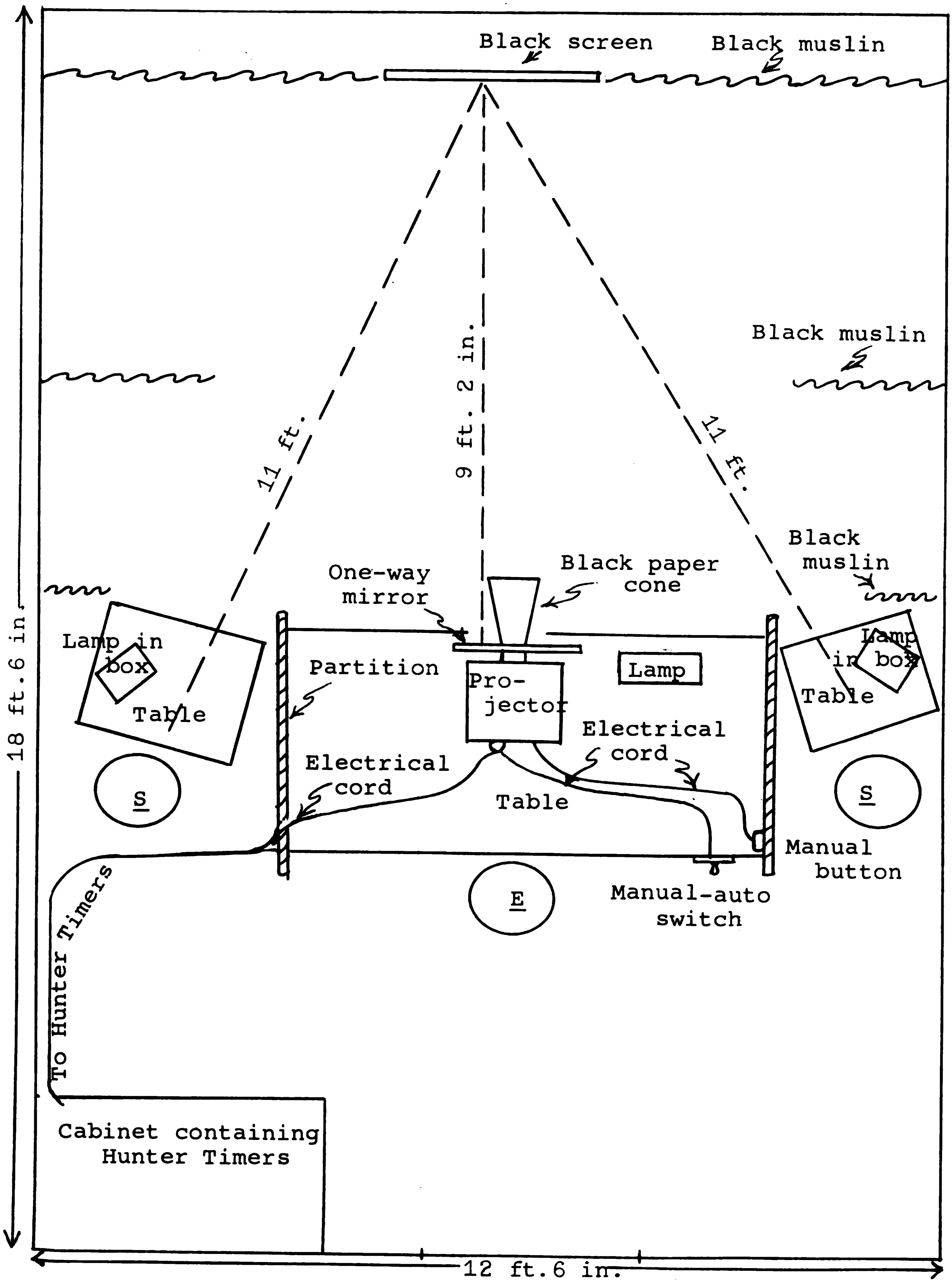


Figure 4

Pictures of the experimental room.

S in booth

E's booth and one S booth

E's projection booth

The projection screen was made of black cloth stretched over a (4' x 7') wooden frame. Black muslin was liberally draped around the projection area and laid on the floor in front of the screen. The insides of the Ss' booths and the tops of the small tables and lamps in the booths were also covered with black muslin. In this way virtually all size cues were eliminated. Except for the very low illumination provided by the recording lamps in each booth, the room was dark.

A Kodak "Carousel 800" projector, equipped with a 500 watt bulb, projected the squares onto the screen. A piece of one-way mirror was placed in front of the projector lens in order to reduce the noticeable increase in room illumination which accompanied the appearance of the large squares; this also eliminated the light beam which extended from the projector lens to the screen providing an extremely obvious distance cue when the squares were shown under full illumination. The projector bulb was set on "high."

The total effect of the dark room, the cue-reducing black cloth, and the projection of the squares at reduced illumination was quite dramatic, the squares seeming to appear in a black void unattached to any apparent surface.[1]

---

[1]This procedure was similar to that originally used by Holzman (1952) wherein small groups were tested in a dark room with the squares being projected by a film strip projector equipped with a 200 watt bulb.

The projector was automatically operated by two Hunter Timers. A table-mounted switch enabled E to change from automatic to manual control. See Appendix D for a schematic drawing of the wiring arrangement.

The series of squares

The first 6 subseries (90 presentations) of the full-length test were employed. This was the same short form of the test used in Experiment I. See Appendix A for the sizes and order of presentation of the squares.

The scoring procedure

As in Experiment I, PRA and LSE were used as the indices of Squares Test performances. See pages 11, 12 and 53 for the details of the scoring procedure.

The design of the study

Three variables were investigated in the main part of the study: (a) the instructions concerning the range of the squares to be judged, (b) the instructions concerning the scale refinement Ss were permitted to use in making their judgments, and (c) the sex of the Ss. Twenty-five Ss from each sex were assigned at random to each of the 6 conditions, resulting in 12 separate groups. Table 4 presents the conditions and variables characterizing each group. The study was designed so as to permit a statistical analysis using a 2 x 3 x 2 factorial analysis of variance.

Table 4. Groups and Conditions (Cell entries are group numbers)

Instructional information given Ss: (a) The squares range from: (b) Concerning the scale refinement permitted:	1"-7"		1"-12"		1"-16"	
	Reg.	Fract.	Reg.	Fract.	Reg.	Fract.
Males	1	2	3	4	5	6
Females	7	8	9	10	11	12

As can be seen from Table 4, three different range information conditions were used, with 4 groups of Ss taking part in each condition. In the 1"-7" Condition Ss were told that the squares range somewhere between 1 and 7 inches; in the 1"-12" Condition, Ss were informed that the squares range somewhere between 1 and 12 inches; while in the 1"-16" Condition, the range was given as no greater than 1 through 16 inches. In all conditions, except 1"-7" fractional, the wording of the instructions concerning range were exactly the same as given on page 8. In the 1"-7" Condition the instructions read: "The squares may range anywhere between approximately 1 and 7 inches, rather than the usual:"The squares may range between 1 and 7 inches." This was done in order to have one precise set of instructions.

Two sets of instructions concerning the scale refinement that could be used in making judgments were used; one consisted of the regular Squares Test wording (Regular) while the other explicitly told Ss that they could use fractional judgments (Fractional). The Regular groups were told: "Write your estimation of the size of each square in its own numbered space. Thus for square number one, record its size in inches next to number one, etc." The Fractional groups were told the same thing except right after the statement "Thus for square number one, record its size in inches next to number one, etc." they were told: "You are not restricted to whole numbers of inches in making your judgments, but may use fractions. If for example, a square looks $4\frac{1}{2}$ or $4\frac{1}{4}$ inches in size, record your estimate as such." Immediately before the series was shown they were again reminded: "Remember you may use fractions in making your judgments." The Fractional instructions removed the ambiguity concerning the scale refinement which Ss could use in making their judgments which characterizes the regular Squares Test instructions.

The "No-information" condition

The "No-information" condition was included in the study in order to determine how Ss perform on the Squares Test when they are given no information concerning the range of the square sizes and are not explicitly told that they may use fractions. The instructions used were the same as the regular Squares Test instructions except that nothing was said about the range of sizes and no reference squares were shown. The instructions read as follows:

> We wish to see how well you can judge the sizes of squares. We are going to show you a number of squares on the screen and we want you to tell us how big they are.
>
> You will see 90 squares during the course of the experiment, and you have answer sheets with a total of 90 spaces provided for your estimates. Write your estimation of the size of each square in its own numbered space. Thus for square number one, record its size in inches next to number one, etc.
>
> Do not go back over your judgments to change them. In changing them you are more likely to be inaccurate.
>
> Now we are ready to begin. You will see each of the following squares for only a few seconds. Look at each square all of the time it is on the screen and make your estimation when it disappears. The first square that you will see is number one.

Questionnaire information

The main purpose of administering a short questionnaire after the completion of the Squares Test was to investigate the relationship between interest and motivation, and test performance (see Appendix C).

Results

The Relationships Between the Main Variables and LSE

In order to determine what effects the three main variables (range information, scale refinement information, and the sex of the Ss) had on Ss' ranges of size estimates, as measured by LSE, a 2 x 3 x 2 factorial analysis of variance was conducted. Since the variances were heterogeneous and standard deviations tended to be proportional to the means, LSE values were transformed to common logarithms.[1] The transformation, however, failed to make the variances homogeneous. A non-parametric analysis was not performed because the findings of Norton (1952)[2] indicated that the analysis of variance is a legitimate test in spite of the heterogeneity of the variances. The summaries of the analyses of both the untransformed and the transformed data appear in Table 5 (see Appendix E for a table of LSE values).

Range information

As can be seen from Table 5 (on the following page) range information (B) had a significant effect on LSE distributions ($P < .01$). The mean LSE values are presented in Table 6, and as can be seen from Figure 5, LSE tends to increase as the range information goes from 1-7, to 1-12, to 1-16 inches. Hypothesis I, therefore, is supported.

Since scale refinement information and sex did not have significant effects on LSE, individual comparisons

---

[1]The F(max.) test was used to determine the homogeneity of the variances. Since the standard deviations tended to be proportional to the means, a log transformation was used (Edwards, 1962, p. 130).

[2]See Lindquist (1953), pp. 78-90.

Table 5. Summary of Analysis of Variance LSE Scores

Source of variation	SS	DF	MS	F
	(untransformed)			
A (scale Refine.)	2.44	1	2.44	.35
B (range Info.)	203.88	2	101.94	14.43**
C (sex)	.51	1	.51	.07
AB	.71	2	.36	.05
AC	.07	1	.07	.01
BC	.73	2	.37	.05
ABC	.56	2	.28	.04
Within Cell (experimental error)	2032.93	288	7.06	
Total	2241.83	299		
	(transformed)			
A (scale Refine.)	.01	1	.01	.50
B (range Info.)	.38	2	.19	9.50**
C (sex)	.00	1	.00	.00
AB	.01	2	.01	.50
AC	.00	1	.00	.00
BC	.01	2	.01	.50
ABC	.00	2	.00	.00
Within cell (experimental error)	5.91	288	.02	
Total	6.32	299		

** P < .01

Table 6. LSE Means (inches)

Range Information	Males		Females		All Ss		All Ss
	Reg.	Fract.	Reg.	Fract.	Reg.	Fract.	Combined
1-7 inches	6.72	6.47	6.76	6.61	6.74	6.54	6.63
1-12 inches	7.96	7.83	7.86	7.88	7.91	7.85	7.88
1-16 inches	8.62	8.46	8.95	8.53	8.78	8.49	8.64

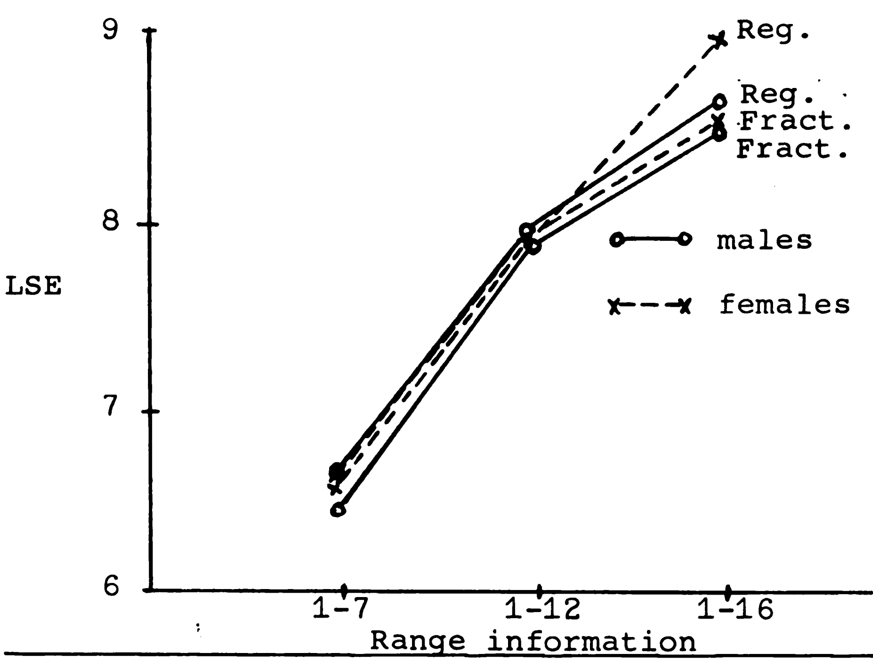
were limited to the means of the groups combined for each range information condition. This analysis[1] indicated that the differences between combined means were significant between all range information conditions (see Table 7).

Table 7. Analysis of the Differences Among Combined LSE Means

Range Information					
1-16 inches	1-12 inches	1-7 inches	d	t	P <
8.64	7.88		.76	2.03	.025
8.64		6.63	2.01	5.36	.010
	7.88	6.63	1.25	3.33	.010

Figure 5

Relationships between Range Information and mean LSE values



[1]The procedure for the a priori comparison of individual means in a factorial design described by Winer (1962), pp. 207-210, was used. Since fixed factors were involved, t tests were conducted using as the error term: $\sqrt{2MS(w\ cell)/N}$, where N = the number of Ss in each group being compared. df = pq (n-1), where p and q equal the number of levels per factor and n is the number of Ss per group in each cell.

Scale refinement information

Whether Ss were explicitly told that they could use fractional judgments or given the regular ambiguous Squares Test instructions concerning scale refinement, made little difference as far as LSE values were concerned. As can be seen from Table 5, no significant relationship obtained between scale refinement information and LSE. The Fractional instructions, however, did reduce mean LSE values in five out of six instances indicating that Ss tend to overestimate somewhat less when encouraged to use a refined response scale (see Table 6).

The sex of the Ss

According to the analysis of variance, sex was not significantly related to LSE (Table 5). An inspection of the distributions of this measure, however, shows that women tend to use the entire response scale more often than men under condition 1-12 Regular (see Appendix E). As can be seen, 7 women but only 1 man used size estimates as large as 12 inches. An analysis of the male and female Ss who overestimated under this range information condition (used estimates of 8 inches or more) showed that significantly more women than men ($P < .02$) tended to use the entire response scale (12 inches)(see Table 8). In Condition 1-16 Regular, women also tended to overestimate more extremely than men. As can be seen from Appendix E, the female mode

Table 8. The frequencies of male and female overestimaters who did and did not use the entire response scale (Condition 1-12 Regular)

	Did not use entire response scale	Did use the entire response scale	
Males	14	1	$\chi^2 = 5.55$
Females	6	7	$\chi^2_{.02} = 5.40$

was 10 inches and the male 8 inches. It appears that both men and women overestimate under range information conditions 1-12 and 1-16, but that women tend to overestimate somewhat more than men.

Summary

There is a significant relationship between the range information given _S_s during the instructions and the range of size estimates on the Squares Test. As range information goes from 1-7, to 1-12, to 1-16 inches, _S_s' ranges of size estimates tend to increase. No significant relationship was found between LSE and scale refinement information, although the Fractional instructions did tend to reduce mean LSE. Neither was sex found to be a significant variable. A comparison of those _S_s who overestimated under condition 1-12 Regular, however, indicated that significantly more women than men used the entire response scale (made estimates as large as 12 inches). Furthermore, under

Condition 1-16 Regular, the female LSE mode was 10 inches and the male 8 inches, indicating that women have a slightly greater tendency to overestimate more than men.

The Relationship between the Main Variables and PRA Scores

The variances of the PRA scores were homogeneous [F(max.) test]. An analysis of variance (2 x 3 x 2 factorial) was conducted on the untransformed data.

Scale refinement information

As predicted (Hypothesis II), scale refinement information had a significant effect on PRA scores, with the Fractional instructions tending to increase PRA (see Table 9 and Figure 6).

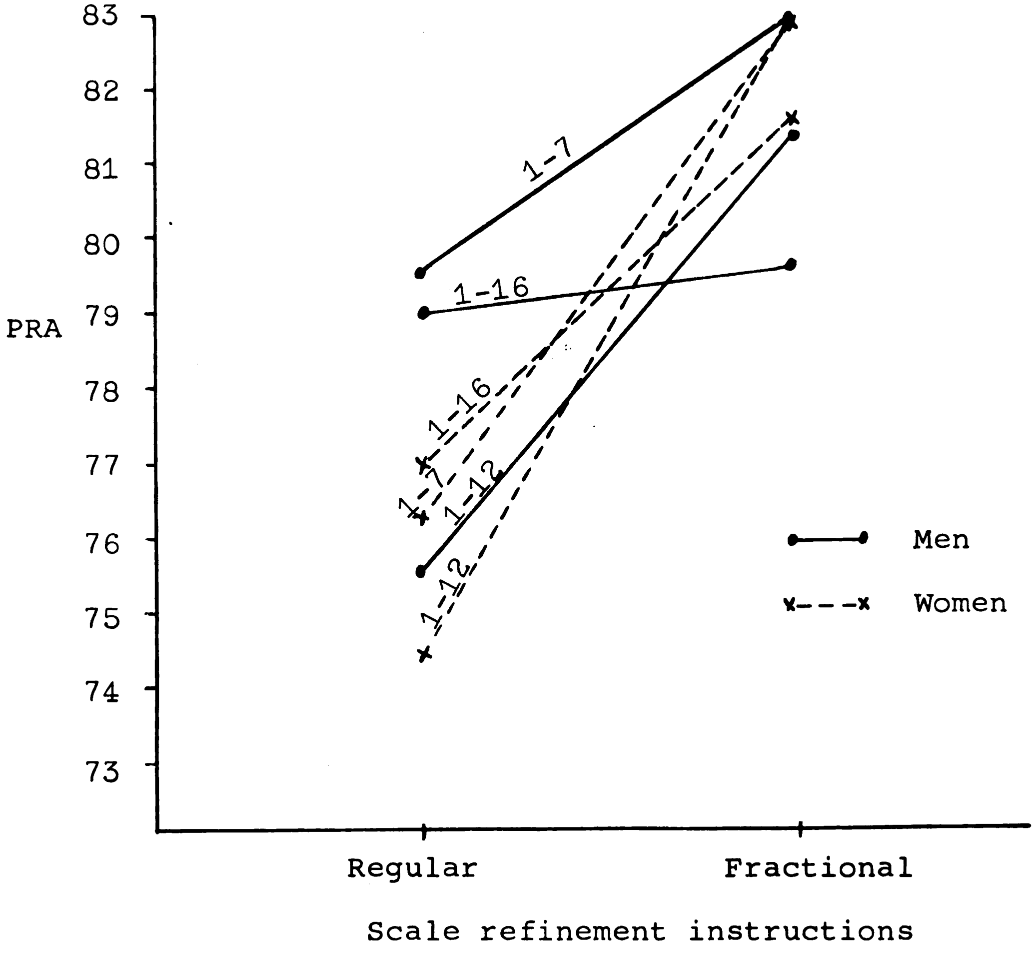
Table 9. PRA scores

Source of variation	SS	df	MS	F
A (scale Refine.)	1842.12	1	1842.12	34.70**
B (range Info.)	173.76	2	86.88	1.64
C (sex)	17.99	1	17.99	.34
AB	270.11	2	135.06	2.54
AC	194.91	1	194.91	3.67*
BC	54.29	2	27.15	.51
ABC	5.13	2	2.57	.05
Within cell (Exper. error)	15,291.09	288	53.09	
Total	17,848.86	299		

* $P < .10$.
** $P < .01$.

Figure 6

PRA scores under Regular and Fractional conditions



The differences between means were significant under all range information conditions for the women; for the men, however, only the difference on condition 1-12 reached significance[1] (see Table 10). The scale refinement x sex interaction approached significance; and as can be seen from Figure 6, female PRA scores were lower than male scores

---

[1]Winer (1962), pp. 207-210.

Table 10. PRA means and differences

	Males					Females				
Inches	Reg.	Fract.	d	t	P	Reg.	Fract.	d	t	P
1-7	79.4	82.9	3.5	1.70	n.s.	76.2	82.8	6.6	3.20	.005
1-12	75.4	81.3	5.9	2.86	.005	74.3	82.9	8.6	4.17	.005
1-16	79.0	79.6	0.6	-	n.s.	76.9	81.5	4.6	2.23	.005

when the Regular instructions were given (under all range information conditions). Furthermore, the PRA increase when Fractional instructions were used was greater for the women than the men under every condition. This would seem to indicate that women, unless explicitly told to attend to fractional differences, tend to discriminate somewhat less among the different sized squares than men. Appendix F presents the distributions of PRA scores.

Range information

Range information had no significant effect on PRA scores (see Table 9).

The sex of the Ss

There was no significant relationship between the sex of the Ss and PRA scores. As had already been discussed, however, the scale refinement x sex interaction approached significance;and women not only scored somewhat lower than men on PRA but got the most benefit from the Fractional instructions.

Summary

Telling *S*s explicitly that they may use fractional judgments significantly increased PRA scores for both men and women. Neither range information nor the sex of the *S*s, however, had much effect on PRA scores, although women tended to have somewhat lower scores than men when the Regular instructions were employed. Furthermore, the increase of PRA was greater for women than men when Fractional, rather than the Regular, instructions were used.

The Relationships between LSE and PRA Scores

Hypothesis IV predicted that LSE and PRA scores correlate positively for female but not male performance under range information conditions 1-12 (Regular) and 1-16 (Regular). This prediction, which was based on the findings of Experiment I, was confirmed in part; the two measures did correlate significantly for the females under the two conditions specified. However, significant relationships also occurred for the females under condition 1-12 (Fractional) and for the males under condition 1-12 (Regular). Table 11 presents the correlations for all of the relevant conditions. (See Appendix F for a table of PRA scores.)

The findings for the females are in agreement with the findings in Experiment I; however, it appears that high PRA scores and high LSE tend to go together for both males and

Table 11. Correlations between LSE and PRA scores (Product-moment)

Range Inform.	Males		Females	
	Regular	Fraction	Regular	Fraction
1-7	.179	.217	-.106	.198
1-12	.395**	.310	.358*	.339*
1-16	.304	.175	.391**	.132

* $p < .05$; ** $p < .025$.

females when Ss are told that the range of square sizes lies somewhere between 1 and 12 inches and the instructions concerning response scale refinement are left ambiguous. Furthermore, the relationships approach significance for the males under conditions 1-12 (Fractional) and 1-16 (Regular), indicating that this sex difference is not a very strong one.

The fact that PRA and LSE correlated significantly for the females under condition 1-12 (Fractional) and approached significance for the males under the same condition, indicates that the relationship is not necessarily dependent on certain Ss using whole number estimates. It appears, therefore, that it is chiefly range information which determines whether or not PRA and LSE will be positively related or not. The ambiguous instructions concerning scale refinement merely maximize the probability that such a relationship will result since the use of whole numbers tends to increase the tendency to overestimate (see Table 12).

Table 12. Mean LSE scores of *Ss* who used all whole numbers compared with those who used some fractions

Condition	Used all whole numbers	Used some fractions	d	t
Males (1-12 Reg.)	8.8	7.0	1.8	n.s.
Males (1-16 Reg.)	9.6	8.1	1.7	n.s.
Females (1-12 Reg.)	9.2	6.5	2.7	$P < .05$
Females (1-16 Reg.)	9.6	8.1	1.5	n.s.

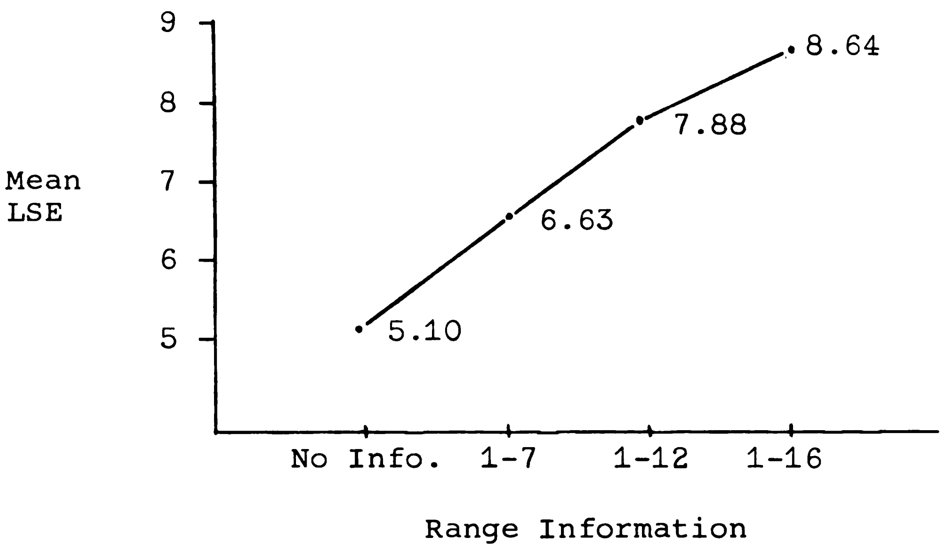
Stated another way, as range information goes from 1-7 to 1-12 to 1-16 inches, the size expectancy induced in *Ss* tends to manifest itself in greater distances between response categories, whether *Ss* use all whole numbers or some fractions; therefore, *Ss* who tend to assign a new response category to each different-sized square in an ascending series have a tendency to overestimate.

Performance under the "No-information" Condition

When *Ss* were given no information concerning the range of the square sizes or the scale refinement permitted, most of them underestimated throughout the series. As Appendix E indicates, only 3 out of 25 men, and 4 out of 25 women used a size estimate larger than 6 inches, with most LSE's ranging from 2 through 5 inches. The mean LSE values were: 5.16 inches (males), and 5.04 inches (females). When compared with the three "information" conditions, the mean LSE (males and females combined) value for the "No-information" group was significantly lower ($P < .01$) than the others[1] (see Figure 7).

Figure 7

Mean LSE scores for the three "information" groups compared with the "No-information" group



The distributions of PRA scores under the "No-information"condition were not significantly different from the distributions of the "information" conditions, the means

---

[1] t test. $t = 2.88$ [$P < .01$ (2-tailed)] between "no-information" mean (5.10) and 1-7 mean (6.63). Since distributions of all groups were similar (Appendix E), the error term used was based on $MS_{w\text{-}cells}$.

being 76.4 for the males and 74.4 for the females (see Table 10, page 85).

Another interesting finding was the fact that only 4 women and 5 men used all whole number estimates in spite of the fact that the Regular instructions were used. This is considerably fewer than the number of Ss who used all whole numbers when the Regular instructions concerning scale refinement were given in combination with range information (see Table 13).

Table 13. Number of Ss using all whole number estimates

Condition	Males	Females
No-information	5	4
1-7 (Regular)	10	15
1-12 (Regular)	16	13
1-16 (Regular)	14	15

The increase in the number of whole number estimates made when range information was provided probably occurred because reference squares which are whole numbers of inches in size are used. This, in the absence of any specific mention of fractional judgments, more than likely, leads some Ss to believe that all of the squares will be whole numbers of inches in size.

Questionnaire Findings

The primary purpose of the short questionnaire which was administered after Ss had completed the Squares Test was to investigate the relationships between Squares Test performance and such factors as interest and motivation. Several items were included in the questionnaire in order to find out how Ss evaluated their own performances. Since few Ss used alternatives (1) and (4) in answering the four-choice Likert scale items, responses were in each case grouped into two categories with alternatives (1) and (2) comprising one category, and alternatives (3) and (4) the other. A high percentage of the Ss gave the same response (alternative 2) on each item; therefore, the analysis was restricted to the 6 Ss having the highest and the 6 Ss having the lowest PRA scores under each condition. (See Appendix C for complete questionnaire.)

Questionnaire responses and PRA scores

Interest and motivation--Hypothesis V states that PRA scores are positively related to Ss' interest in and motivation toward the Squares Test. It was predicted that Ss who find the task interesting and are motivated to do well are more likely to attain high PRA scores than those who are less interested and motivated. Items 1, 4, and 6 pertain to this issue.

Item 1 states: "The task was: (1) very interesting___, (2) interesting___, (3) uninteresting___, and (4) very uninteresting___." Grouping the responses into two categories, it was found that there is no significant relationship between PRA scores and interest. As can be seen from Table 14, however, more Ss with low than high PRA scores, indicated that they found the task uninteresting.

Table 14. Extreme PRA Scores and Interest

		(Item 1) The task was interesting	The task was uninteresting	
MALES	Low PRA	30	6	(n.s.)
	High PRA	35	1	
FEMALES	Low PRA	27	9	χ^2_Y = .80 (n.s.)
	High PRA	31	5	
ALL Ss	Low PRA	57	15	
	High PRA	66	6	χ^2_Y = 3.57 ($\chi^2_{.05}$=3.8 df= 1) (n.s.)

Item 4 reads: "During the task, I was: (1) highly motivated___, (2) somewhat motivated___, (3) very little motivated___, (4) really not motivated at all___." Responses were grouped into a "Motivated" and an "Unmotivated" category.

The relationships between PRA and motivation did not reach significance for the males and females considered separately, but did so when the two groups were combined ($P < .025$) (see Table 15).

Table 15. Extreme PRA Scores and Motivation

	(Item 4) Motivated	Unmotivated	
Low PRA	58	14	$\chi^2 = 5.14$
High PRA	68	4	($\chi^2_{.025} = 5.0$, df = 1)

Item 6 states: "Briefly describe your general mood when you were taking the test." Responses were classified into two categories: Category I (positive orientation), and Category II (negative orientation). Category I responses included "tried, alert, attentive, motivated to do well, interested, very interested, challenged, concentrated, curious, anxious to do well, willing, enthusiastic, good mood, eager, testing myself, cooperative." Category II responses were: "tired, depressed, gloomy, bored, apprehensive, confused, impatient, undecided, mind wandered, nervous, thought it kind of dumb, difficult, apathetic, tense, automatic, worried, lazy, frustrated, futile, disgusted with, restless, uneasy, no mood."

Both male and female high PRA Ss tended to give more positive than negative responses than low PRA Ss although the relationship only reached significance for the males ($P < .01$) and for all Ss combined ($P < .005$) (see Table 16).

Table 16. Extreme PRA scores and Mood

		(Item 6) Category I (positive)	Category II (negative)	
MALES	Low PRA	12	24	$\chi^2_Y = 6.72$
	High PRA	24	12	($\chi^2_{.01} = 6.60, df=1$)
FEMALES	Low PRA	13	23	$\chi^2_Y = 2.01$
	High PRA	20	16	($\chi^2_{.05} = 3.80, df=1$) (n.s.)
ALL Ss	Low PRA	25	47	$\chi^2_Y = 9.02$
	High PRA	44	28	($\chi^2_{.005} = 7.90, df=1$)

In the light of these findings, Hypothesis V is definitely supported. Although a significant relationship was not obtained between PRA and Item 1 responses (Interest), it must be remembered that this item does not ask Ss how interested they were *in* the task, but instead asks how interesting a task they consider the test to be. There is a subtle but real distinction here, and possibly the item should have been worded, "I was: (1) very interested in the task___, (2) somewhat interested in the task___," etc. In this way it would have been possible to find out more about the degree of involvement of Ss in the test. Items 4 and 6 did concern S involvement in the Squares Test and responses to both items indicated that High PRA Ss, more so than Low PRA Ss, tend to be motivated and have a positive orientation toward the task.

Responses to the other questionnaire items--The only other significant relationships between PRA and questionnaire responses occurred on Item 5 which reads: "Who do you think do better (on the average) on this task? (1) men___, (2) women___, (3) both the same___." As can be seen from Table 17, only 1 Low PRA female believed that women do better on the task, while 12 High PRA females felt that women do better ($P < .001$).

Table 17. Extreme PRA scores (females) and the belief that men or women do better on the test

	Men do better	Women do better	Both the same
Low PRA	25	1	10
High PRA	9	12	15
	$\chi^2 = 17.84$	($\chi^2_{.001} = 13.8, df=2$)	

This might mean that Low PRA females have little confidence in their performances, and therefore feel that men, whom they envision as being more interested in the Squares Test, probably do better. Likewise, female Ss who score high on PRA because they apparently try to discriminate well among the square sizes, possibly feel that they did as well as or better than other Ss, men included.

Sex differences on questionnaire items

The only item on which female performance differed significantly from male performance was Item 2 which states: "While making my judgments, I felt that I was usually: (1) very accurate___, (2) fairly accurate___, (3) inaccurate___, (4) very inaccurate___." Grouping the responses into two categories, it was found that more men than women felt that they were accurate on the test ($P < .01$)(see Table 18).

Table 18. Sex and Ss' evaluations of their accuracy

	(Item 2) Accurate	Inaccurate	
Men	128	22	$\chi^2 = 7.17$
Women	108	42	($\chi^2_{.01} = 6.6$, df=1)

LSE scores and questionnaire responses

There were no significant relationships between the largest size estimates made (LSE) and responses to the questionnaire items.

Questionnaire responses for all Ss

In order to present an overview of the way in which Ss generally viewed the Squares Test and their performances on it, the responses of all Ss on each item are presented in Table 19.

Table 19. The responses of all Ss

Item		Responses
1	The task was interesting	248
	The task was uninteresting	52
2	I felt accurate	236
	I felt inaccurate	64
3	I felt that I:	
	Overestimated	45
	Underestimated	148
	Neither	107
4	I was motivated	255
	I was not motivated	45
5	Men do better	134
	Women do better	47
	Both do the same	119
6	Had a positive mood	184
	Had a negative mood	116
7	I used all whole numbers because:	
	That's what the instructions said to do	30
	It was easier	18
	They were my best estimates	26

Note: Not all Ss were given Item 7; for that reason the number of Ss does not tally with Table 13, page 90.

On the basis of self report, it appears that most Ss are fairly interested and motivated when taking the Squares Test. The majority feel that they tended to overestimate but probably not by too much because relatively few believe that they were very inaccurate. In addition most men and women think that men generally do better on the Squares Test than women.

Summary of the Findings

1. Concerning the range of size estimates

(a) Size estimates are a function of the range information given Ss in the instructions. Ss' ranges of size estimates tended to conform with whatever range information was provided. Hypothesis I is supported.

(b) More women than men overestimated extremely when Ss were informed that the range of square sizes lies within either 1-12 or 1-16 inches. The sex of the Ss, however, was not significantly related to overestimation tendencies, with both men and women tending to overestimate under the 1-12 and 1-16 range information conditions. Hypothesis III is only partially supported.

(c) When no range information was provided during the instructions, most Ss of both sexes tended to underestimate.

(d) The degree of interest in the Squares Test, or the motivation of the Ss was relatively unrelated to the range of size estimates.

(e) Most Ss, regardless of the absolute sizes of their estimates believed that they tended to underestimate although many felt that they neither over- or underestimated consistently. Few, however, thought that they overestimated.

2. Concerning the discrimination of square sizes (PRA scores)

(a) Explicitly telling Ss that they may use fractional estimates,rather than leaving the instructions concerning scale refinement ambiguous, tended to increase PRA scores. (Hypothesis II is supported.)

(b) Although the sex x scale refinement instructions interaction did not reach significance, telling Ss that they may use fractional judgments tended to increase PRA scores more for women than men.

(c) Range information had no significant effect on the distributions of PRA scores.

(d) PRA scores and motivation on the Squares Test were positively related.

(e) PRA was related to women's opinions concerning who do better on the Squares Test, men or women. Women who attained high PRA scores tended to believe either that women do better than men or that they perform about equally, while women who scored low on PRA believed that men do better or both sexes perform about the same.

(f) Ss who used all whole number estimates did so for one of three reasons: (1) because they thought that they had been instructed to do so, (2) because it was easier, (3) because whole numbers represented their best estimates.

3. The relationships between LSE and PRA scores

PRA scores and LSE scores were positively related for the women under conditions: (1) 1-12 (Reg.), (2) 1-12 (Fract.), and (3) 1-16 (Reg.). For the men, the relationship was significant for the 1-12 (Reg.) condition only. Hypothesis IV, therefore, is supported in part.

4. Other findings

(a) More men than women thought they were accurate.

(b) Most Ss found the Squares Test moderately interesting and tried to do their best. A large minority of Ss, however, indicated that they were at times either bored, frustrated or both.

DISCUSSION

The findings of this investigation clearly demonstrate that the instructions, sex, and motivation of the Ss play important roles in Squares Test performance. The relationships between these variables and the two indices of Squares Test performance employed (PRA and LSE) will be discussed in turn.

The Range of Size Estimates (LSE)

Instructional range information

The fact that almost all Ss underestimated when no range information was provided, in addition to the finding that the ranges of size estimate tended to conform with whatever range information was given, indicates that an instruction-induced set or expectancy concerning the square sizes was operative. Evidence for the existence of such a set is also provided by the fact that most Ss had the feeling during much of the test that they were underestimating. This feeling was probably caused by the fact that Ss are first told to expect squares ranging over a fairly wide size dimension and are then shown only small squares for most of the series. In such a situation Ss have the choice of either deciding that the E gave them misleading information or that they are underestimating; the majority appear to have faith in the information given them and therefore begin to feel that their estimates must be on the low side as the series progresses. Since most _S_s, however, believed that they were fairly accurate during most of the task, they evidently did not feel that they were underestimating very much. Some probably compensated for the feeling that they were underestimating by adjusting their estimates upward, resulting in the overestimation which was observed.

This is only part of the picture however. It will be recalled that _S_s are shown two reference squares during the instructional period as well as being given verbal information concerning the range of square sizes. It seems likely, therefore, that _S_s who are able to remember accurately the way the reference squares looked tend to inhibit their urge to adjust their estimates upward and thus are less likely to overestimate than _S_s who quickly forget how big the reference squares looked. In view of this consideration, the range of size estimates (LSE) is probably a function of the interaction between size expectancy and the ability to remember how big the reference squares appeared.

Scale refinement information

The two sets of instructions concerning scale refinement (Regular and Fractional) did not significantly affect the ranges of size estimates. Size expectancy induced by instructional range information, in other words, leads many *S*s to overestimate whether they use all whole numbers or some fractions in making their judgments. It will be remembered, however, that *S*s who used all whole number estimates did tend to overestimate more than those who used some fractions. The response units *S*s decide to use, therefore, cannot be discounted as a factor in the ranges of their size estimates.

The relationship between LSE and PRA scores

The fact that PRA and LSE are positively related under some range information conditions indicates that *S*s who discriminated well among the squares also tended to overestimate. By discriminate is meant that more response categories were accurately used, not that more square size differences were necessarily seen. Stated another way, it is believed that PRA scores reflect *S*s' choices of response systems rather than any perceptual differences based on the assimilation of the memories of previous judgments. The tendency to use many response categories in conjunction with the expectancy that some of the squares will be as large as the large reference square, apparently increases the likelihood that overestimation will occur in a task such as the Squares Test which is characterized by an ascending series of stimuli.

Interest and motivation

Since neither interest, motivation nor mood were significantly related to LSE,it is assumed that Ss can be induced to overestimate whether they are highly motivated or not. This finding strengthens the hypothesis that the main cause of progressive overestimation is instruction-induced size expectancy. Nor is it surprising that instructional set plays such a primary role in Squares Test performance for as Vernon (1937) in his discussion of set stated:

> Thus clearly the 'sets' or attitudes adopted and the processes which actually occurred were functions of the experimental material and of the individual's inherent characteristics, as well as of the experimental instructions. (P. 49)

The sex of the subjects

Hypotheses III predicted that women are more influenced by instructional range information than men. This prediction received some support in that more women overestimated extremely than men, although sex and range information were not significantly related overall. This finding is compatible with those of Witkin (1954) concerning field dependence-independence, Janis and Field (1959) concerning persuasibility, Eysenck and Furneau (1945) concerning suggestibility, and Luchins and Luchins (1959) concerning rigidity. In all of these studies, it was found that women were more likely to adopt and retain a frame of reference which had in one way or another been provided for them, regardless of the appropriateness of the referent. As Luchins and Luchins (1959) pointed out, the teacher-student relationship builds a tendency for females to have blind confidence in what the teacher says. The same can probably be said for the experimenter-subject relationship and it is proposed that the reason extreme overestimation is more characteristic of female than male performance is because women have more faith in what the experimenter tells them about the range of the square sizes, than men. Female college students, in other words, are more suggestible than their male contemporaries. It must be emphasized, however, that both men and women are susceptible to instructional set on the Squares Test and that the sex difference which does exist is not a large one.

General considerations

Looked at from another perspective, this study was concerned with the problem of the role of perception versus the role of response systems in evaluative tasks. Are the size estimates on the Squares Test truly reflections of the way the squares looked, or are they primarily manifestations of attempts to match stimuli with the categories of the response system adopted? Furthermore, when size estimates are made within the context of different range information, should the changes in judgment which occur be ascribed to different perceptual experiences (the squares actually looked different) or are the changes in judgment due to a "redefinition of the categories of the response scale?" (Weiss, 1963). Although it cannot be denied that the squares may in fact look different within the context of different range information, it seems clear that the changes in judgment which occur can more meaningfully be conceptualized in terms of changes in the response categories which size expectancy brings about. Stated very simply, the information presented Ss concerning the range of square sizes determines the response scale most Ss use in making their judgments. If this response scale extends from 1 through 7 inches, then the size estimates tend to go from 1 through 7 inches; if the response scale extends from 1 through 16 inches, then there is the tendency for some Ss to extend their judgments across all or most of the 16 inch range.

The research on anchoring effects is also relevant to this study. Although no attempt will be made to consider this area in detail, it is quite obvious that the reference squares and the information concerning the range of squares are stimulus and response anchors respectively. As Sherif, Taub and Hovland (1958), Torgerson (1958), and others have reported, such anchors can be expected to exert either assimilation effects, contrast effects, or both. Both effects occur on the Squares Test as can be seen from Appendix E. Some Ss under range information conditions 1-12 and 1-16 tend to extend their judgments across the entire range given them (assimilate towards the large reference square), while others underestimate (exhibit contrast effects with regard to the large reference square).

To summarize, regardless of whether one considers the present research and findings within the context of set theory, anchoring effects, or the perception-response system dichotomy, the fact remains that the information given the *S*s in the Squares Test is reflected in the range of size estimates *S*s make. Furthermore, the degree to which the instruction-induced set manifests itself in the size judgments depends upon how well *S*s are able to remember the appearances of the reference squares, and the inch scale refinement they employ in making their estimates. Finally, there is some evidence that size expectancy manifests itself most fully in the performances of some women.

The relevance of the findings for the concept leveling-sharpening

Since across-series leveling-sharpening scores (AS, IE, IE-reg., ESL and LSL)[1] reflect the range of size estimates made on the Squares Test, these measures are, to some extent, indices of the interaction between instructional set and *S*s' abilities to remember how big the reference squares looked. Such scores, therefore, are inappropriate measures of differences in assimilation tendencies among memory

---

[1]See pp. 14-16 for description of these scores.

traces which are theorized to underlie leveling and sharpening. Although it is not suggested that the memories of previous squares (particularly the immediately preceding one) do not play a part in Squares Test performance, across-series scores are too sensitive to instructional set to be of much value as indicators of assimilation effects. For this reason, it is recommended that such scores not be used in the operational definition of leveling-sharpening.

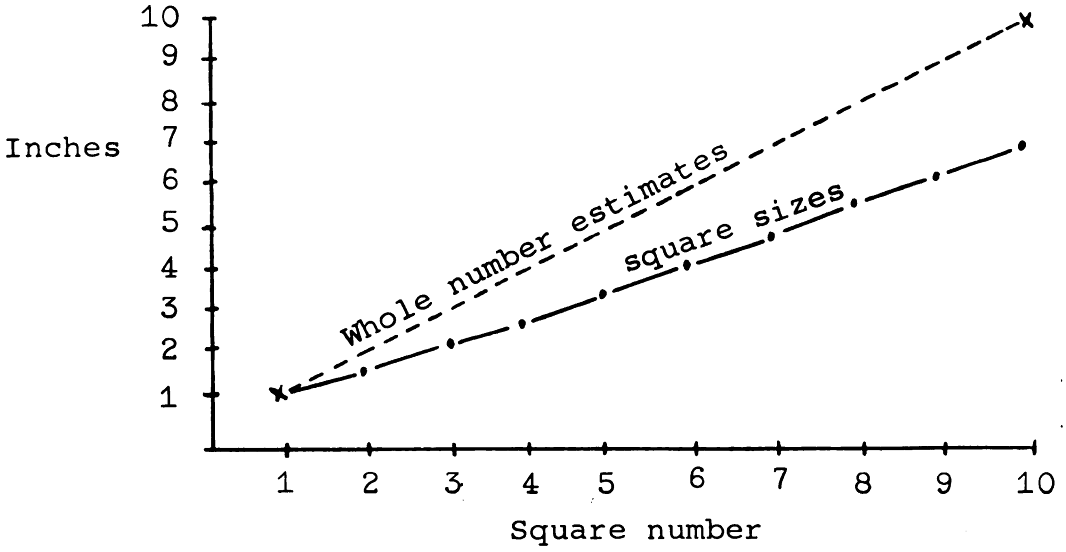
The Discrimination of Square Sizes (PRA)

The PRA score is for the most part an index of the number of times Ss assign the same size estimate to two or more different-sized squares. Although this measure also reflects the degree to which the ranking of size estimates differs from the ranking of the actual square size estimates, ranking reversals are relatively few in number.

The size differences between nine of the thirteen squares which comprise the series are less than 1 inch; therefore perfect discrimination requires that Ss use fractional size estimates if they are to keep from over-estimating. In both phases of the research, it was found that the instructions concerning scale refinement are ambiguous, leading some Ss to assume that all estimates are to be rounded to the nearest whole number of inches. If a S makes this assumption, and he or she has the ability to remember accurately the sizes of the reference squares, it is inevitable that the same size estimate will frequently need to be assigned to two or more different-sized squares. This occurs with the S usually being aware of the fact that the same estimate is being used for squares of different size. If a S, on the other hand, has difficulty remembering the sizes of the reference squares and uses only whole numbers, then progressive overestimation is likely to take place (see Figure 8).

Figure 8

The trend of judgments when a different whole number is used for each square size



Scale refinement instructions

The fact that PRA scores increase when Ss are told that they may use fractions is strong evidence that this score depends, in part, on the way Ss interpret the instructions concerning the inch scale units they may use in making their judgments. The S who uses all whole numbers is faced with a different problem than the one who uses some fractions. The former must in each case decide whether the size-difference between two squares is large enough to merit two different responses (whole numbers) or not. The latter, on the other hand, is not concerned with the problem of deciding whether or not to give the same estimate for two similar squares, but rather attempts to indicate just how large the size difference is.

If a S does use a different whole number every time a new square appears and overestimation occurs, it may also happen that the S becomes aware of the fact that his judgments are out of line with his memories of the reference squares. If this takes place, he or she may then begin to use the same integer as an estimate for several different-sized squares thus compensating for the overestimation which previously occurred. Such changes in judgmental strategy may account for some of the complex patterns which are evidenced on the Squares Test (see the review of the Smith and Klein, 1953, article on pp.26-29). Contrary to Holzman and Klein's (1954) assertion that scale preference is irrelevant to leveling-sharpening studies, the present findings indicate that the response units Ss employ do manifest themselves in PRA scores.

Gardner (1953) made the point that in order to investigate cognitive controls, criterion tasks must be clearly understood by all Ss. There should be no question, in other words, as to what Ss are to do. Only if this is so, it is argued, is it possible to assume that performance differences result from different cognitive strategies, rather than from different interpretations of the instructions. Since the Squares Test instructions are ambiguous, this task is in essence not a suitable criterion task for the concept leveling-sharpening.

Motivation and interest

Low PRA scores, of course, do not necessarily mean that Ss failed to discriminate more carefully among the squares because they thought all judgments were to be made with integers. As the questionnaire data revealed, some Ss used all whole numbers because it was the easy thing to do. This indicates that PRA scores to some degree reflect S involvement or motivation with regard to the Squares Test. Further support for this hypothesis comes from the finding that a positive relationship exists between PRA and the responses to the questionnaire items concerning motivation and mood. A S must obviously attend continuously to the task if the squares are to be correctly discriminated and a high PRA score achieved; it follows, therefore, that a person who for some reason is not inspired to put forth the effort necessary for sustained attention, will score low on PRA. The Experiment I finding that Ss tend to attain lower PRA scores the second time they take the Squares Test also bears out this contention. The interviews with the Ss indicated that they were less motivated to perform maximally during the retest than during the original administration of the test, and it seems reasonable to assume that the drop in PRA resulted from a decrease in motivation.

It will be recalled that both Klein (1951) and Jeffreys (1953) found that high PRA Ss tended to be more self-outward, expressive and conforming than low PRA Ss. If this is the case, it is again possible to assume that S involvement was the factor which brought about the differences in PRA scores, for a S who trys hard on the Squares Test probably is cooperative and therefore is likely to be judged as out-going and expressive. Other research (Gardner *et al*., 1959; Gardner and Lohrenz, 1960; Holzman and Gardner, 1960; Lachman, Lapkin and Handelman, 1962) indicated that PRA seemed to be related to Ss' preferences for detailed or nondetailed responses when answering the queries of psychologists or taking part in psychological studies. It seems logical to assume that almost everyone is capable of making detailed responses if inspired to do so. Ss, therefore, probably give few and undetailed Rorschach responses, recall few dreams and tell stories using few details for the same reason that they use whole numbers on the Squares Test--it's the easy thing to do.

Instructional range information

The particular information given Ss during the instructions concerning the range of the square sizes has little effect on the overall distributions of PRA scores, although, as has already been discussed, PRA and LSE tend to be positively related under some range information conditions.

The sex of the subjects

Although the relationship between the sex of the Ss and PRA scores was not significant, women did, on the average, score lower on this measure than men when the Regular instructions concerning scale refinement were used (see Figure 6, page 84). The male and female groups, with the exception of those in Condition 1-7 (Reg.), did not differ with respect to the number of Ss using all whole number estimates (Table 13, page 90). Therefore the difference in mean PRA scores cannot be attributed to a difference in the tendency to use whole numbers. In view of this, it seems likely that the difference may have been due to a tendency of the women to set lower performance standards for themselves than the men. This conjecture is given some support by the fact that the sex difference disappeared when Fractional instructions were used causing all Ss to use more refined standards of judgment.

The relevance of the findings concerning PRA for the concept leveling-sharpening

The PRA score has been the most commonly used index of leveling-sharpening tendencies. Furthermore, it has generally been hypothesized that this score reflects the degree of assimilation which takes place among Ss' memory traces of previous squares in the series. Ss who score low on PRA (levelers), therefore, are believed to be characterized by a cognitive organizational system wherein a great deal of assimilation or fusion takes place among the memory traces, while Ss who attain high PRA scores (sharpeners) are thought to do so because the traces of previous stimuli, no matter how similar, are kept separate, providing an internal frame of reference which makes possible the accurate discrimination of future stimuli.

The review of the literature revealed that there is little support for the notion that differences in the degree of assimilation effects among memory traces represents an important dimension of cognitive functioning. The results of the present research indicate that even if such general tendencies exist, there is little reason to believe that they manifest themselves in PRA scores. Even if one goes so far as to concede that differences in assimilation effects might be reflected in this measure, it is still a poor index of leveling-sharpening because it is affected by the way Ss interpret the instructions and motivational factors.

For this reason, it is recommended that PRA scores not be used as indices of assimilation tendencies.

Conclusion

The findings of the present study cast serious doubt upon the use of the Squares Test as a criterion task for the concept leveling-sharpening. Both across-series and intrasubseries scores reflect the operation of variables other than the assimilation tendencies among memory traces which are generally hypothesized to underlie leveling-sharpening. Among these are instructional set, interpretations of the instructions concerning scale refinement, sex, and the motivation of the _S_s. With all of these factors involved in Squares Test performance, it is difficult to see how this test can properly be considered the criterion task for any single hypothesized cognitive control. Since previous research has failed to provide much support for the idea that leveling-sharpening is a major aspect of cognitive structure, it is suggested that: (a) Squares Test performance be conceptualized in terms of such factors as were investigated in the present research (particularly instructional set, sex and motivation), and (b) the Squares Test not be used as the criterion task for leveling-sharpening.

This does not mean that further research involving the Squares Test might not be fruitful. The test has high short-term reliability and seems to afford an effective way of investigating instruction-induced set and susceptibility to suggestion. One project of interest, for example, is the determination of the range information which is most effective in inducing overestimation. In the present study, the modal overestimation value tended to be higher under 1-12 rather than 1-16 inch range information condition, indicating that the maximally effective information is less than 1-16 inches.

After the relationship between range information and instructional set is more clearly determined, it would be worthwhile to see if Ss who overestimate most extremely on the Squares Test are also most influenced by the instructions in another similar task. It would be possible, for example, to study instructional set in the auditory mode by the use of a tone duration task. Ss could be asked to estimate the durations of a series of tones which progresses slowly and irregularly from the shortest to the longest. As in the Squares Test, groups could be tested under different information conditions concerning the range of tone durations. If the same Ss who overestimate on the Squares Test also tended to overestimate the tone durations, then there would be some evidence that susceptibility to influence in judgmental situations is more than a task specific trait.

If a positive relationship was found between the manifestation of instructional set on the Squares test and a tone duration test, then the next step might be to determine whether such tendencies are related to performance on Asch-type situations wherein Ss make judgments under the pressure of group influence.

In addition, the attentional and motivational aspects of the test suggest that there might be a relationship between PRA scores and performance on certain vigilance tasks (Mackworth, 1950; Bakan, 1955). It is hypothesized that Ss who score low on PRA do less well on vigilance tasks than those who attain high PRA scores.

Regardless of the problem investigated, however, it is extremely important that the Squares Test be administered to small groups under carefully controlled conditions. The instructions, for example, must be unambiguous and size cues kept to a minimum. Since sex differences appear to be involved, the sex and temperment of the E, as well as the Ss, should be taken into account. Finally, since the test is sensitive to numerous factors, the temptation to interpret performance differences in terms of a single characteristic of cognitive functioning, must be avoided.

SUMMARY

According to the theory of cognitive controls (Klein, 1951), leveling-sharpening represents an important dimension of cognitive structure. It is believed that memory traces of previous stimuli fuse or assimilate, and that they do so more for some individuals (levelers) than others (sharpeners). This supposedly leads the former to differentiate less among successive stimuli than the latter.

The Squares Test, designed by Holzman and Klein (1951), generally serves as the criterion task for the concept leveling-sharpening. The test consists of a series of 150 squares of light projected successively onto a black screen in an almost completely darkened room. The squares range in size from 1.2 to 13.7 inches and the series is made up of 10 overlapping subseries. Subseries 1 involves the 5 smallest sizes (presented in 3 different orders), and the series progresses in stepwise fashion from the smallest to the largest squares so that Subseries 10 includes only the 5 largest sized squares. Subjects (Ss) are required to estimate the size of each square.

Numerous scores, all purporting to reflect differences in memory trace assimilation (leveling-sharpening) have been used, some measuring across-series changes in judgment and

others based on differences in intrasubseries square size differentiation. For this reason, the operational definition of leveling-sharpening is rather unclear. Definitions of the concept are further obscured by the procedural modifications which characterize much of the research. Furthermore, investigators have had little success in demonstrating that leveling-sharpening is a major factor in cognitive functioning.

A study by the author (Gilgen and Diehm, 1964), in conjunction with a review of the literature, led to a general dissatisfaction both with the concept leveling-sharpening as defined in terms of assimilation theory and with the Squares Test as a measure of any single diminsion of cognitive functioning.

The present study consists of two parts: Experiment I, which was exploratory and had as its purpose the determination of those aspects of the Squares Test which are used by Ss as judgmental referents; and Experiment II which systematically investigated the hypotheses suggested by Experiment I findings.

Thirty-three Ss (14 men and 19 women) were individually tested in Experiment I, each S taking the Squares Test twice during the same session. A short form of the test (the first 90 presentations) was used in order to reduce boredom. The findings were as follows: (a) the Squares Test has good short-term reliability (.94); (b) some Ss tend to overestimate extremely; (c) women tend to overestimate more than men; (d) intrasubseries discrimination (as measured by Positional Ranking Accuracy) is positively related to the range of size estimates (as measured by the largest size estimate given) for women (.66), but not men (.21); (e) the instructions fail to make it clear whether Ss are to use whole numbers or fractions in making their size estimates; (f) explicitly telling Ss that none of the squares are exact whole numbers of inches in size increases intrasubseries square-size discrimination; (g) Ss tend to discriminate less well among the square sizes the second time they take the Squares Test.

Experiment II (N = 350; 175 men, 175 women) was designed to investigate the roles of instructional set, instructional ambiguity, sex, and the motivation of Ss on Squares Test performance. Since the smallest judgment made by most Ss was 1 inch, the range of size estimates was measured by the largest size estimate made (LSE); in addition, performances were scored for Positional Ranking Accuracy (PRA). Three different sets of instructions concerning the range of square sizes were used (1-7 inches, 1-12 inches, and 1-16 inches) and Ss were either given the regular ambiguous instructions concerning response scale refinement or explicitly told that fractional estimates were permissible. One group, which was given no range information, was tested later.

The findings were as follows: (a) the ranges of size estimates were functions of the information given Ss concerning the range of square sizes to be judged (size expectancy was operative); (b) most Ss who were given no range information underestimated; (c) although sex was not significantly related to the range of size estimates, women who overestimated tended to overestimate more than men who overestimated; (d) PRA and LSE scores were significantly related for both men and women; however, the relationship obtained under more conditions for the women than the men; (e) the ambiguity of the instructions concerning the inch scale refinement (whole numbers or fractions) which Ss are permitted to use is reflected in PRA scores (PRA increases when Ss are explicitly told that they may use fractions); and (f) the motivation of Ss on the Squares Test (as measured by questionnaire responses) was positively related to PRA scores.

It was concluded that instructional set, instructional ambiguity and the sex and motivation of Ss are involved in Squares Test performance, and that the test fails to provide a meaningful operational definition of leveling-sharpening. The concept obscures rather than clarifies Squares Test performance and should be dissociated from the test. It was suggested that the test might, however, be fruitfully employed in the investigation of instructional set.

REFERENCES

Alexander, W. P. Intelligence, concrete and abstract. Brit. J. Psychol., Monograph Suppl., 1935, 19.

Allport, G. W. & Postman, L. The psychology of rumor. New York: Holt, 1947.

Bakan, P. Discrimination as a function of time in a prolonged vigil. J. exp. Psychol., 1955, 50, 387-390.

Bennett, G. K. & Cruikshank, R. M. Sex differences on the understanding of mechanical problems. J. appl. Psychol., 1942, 26, 121-127.

Cass, J. & Birnbaum, M. Comparative guide to American colleges. New York: Harper & Row, 1964.

Cronbach, L. J. Essentials of psychological testing. (2nd ed.) New York: Harper & Bros., 1960.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. Psychol. Rev., 1959, 66, 183-201.

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, 1962.

Eysenck, H. J. & Furneau, W. D. Primary and secondary suggestibility: an experimental and statistical study. J. exp. Psychol., 1945, 35, 485-503.

Fink, D. M. Sex differences in perceptual tasks in relation to selected personality variables. Unpublished doctoral dissertation, Rutgers Univ., 1959.

Fiske, D. W. Consistency of the factorial structures of personality ratings from different sources. J. abnorm. soc. Psychol., 1949, 44, 329-344.

Freud, S. Beyond the pleasure principle. London: The International Psychoanalytic Press, 1922.

Gardner, R. W. Cognitive styles in categorizing behavior. J. Pers., 1953, 22, 214-233.

Gardner, R. W. The development of cognitive structures. In Constance Scheerer (Ed.), Cognition: Theory, research, promise. New York: Harper & Row, 1964, pp. 147-171.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, Harriet B.,& Spence, D. Cognitive control: A study of individual consistencies in cognitive behavior. Psychol. Issues, 1959, 4.

Gardner, R. W., Jackson, D. N., & Messick, S. J. Personality organization in cognitive controls and intellectual abilities. Psychol. Issues, 1960, 2.

Gardner, R. W., & Lohrenz, L. J. Leveling-sharpening and serial reproduction of a story. Bull. Menninger Clin., 1960, 24, 295-304.

Gardner, R. W., & Long, R. I. Cognitive controls as determinants of learning and remembering. Psychologia, 1960, 3, 165-171.(a)

Gardner, R. W. & Long, R. I. The stability of cognitive controls. J. abnorm. soc. Psychol., 1960, 61, 485-487. (b)

Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. Psychol. Bull., 1941, 38(9), 781-817.

Gilgen, A. R., & Diehm, D. A comparison of leveling and sharpening tendencies on a square-size estimation task and a word-rating task. Paper read at Midwest Psychol. Ass., St. Louis, May, 1964.

Hartmann, H. Ego psychology and the problem of adaptation. New York: International Univer. Press, 1958.

Helson, H. Adaptation level as a frame of reference for prediction of psychophysical data. Amer. J. Psychol., 1947, 60, 1-29.

Hollingworth, H. L. Experimental studies in judgment. Arch. Psychol., 1913, 29, 1-119.

Holzman, P. S. Cognitive attitudes of leveling and sharpening in time-error assimilation tendencies. Unpublished doctoral dissertation, Univer. of Kansas, 1952.

Holzman, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. J. Pers., 1954, 22, 375-394.

Holzman, P. S., & Gardner, R. W. Leveling and repression. J. abnorm. soc. Psychol., 1959, 59, 151-155.

Holzman, P. S., & Gardner, R. W. Leveling-sharpening and memory organization. J. abnorm. soc. Psychol., 1960, 61, 176-180.

Holzman, P. S., & Klein, G. S. The "schematizing process": perceptual attitudes and personality qualities in sensitivity to change. Amer. Psychologist, 1950, 312.

Holzman, P. S., & Klein, G. S. Perceptual attitudes of "leveling" and "sharpening": relation to individual differences in time error. Amer. Psychologist, 1951, 6, 257.

Holzman, P. S., & Klein, G. S. Cognitive system principles of leveling and sharpening: individual differences in assimilation effects in visual time error. J. Psychol., 1954, 37, 105-122.

Janis, I. L., & Field, P. B. Sex differences and personality factors related to persuasibility. In C. I. Hovland & I. L. Janis (Eds.) Personality and persuasibility. New Haven, Conn.: Yale Univer. Press, 1959.

Jeffreys, A. W. An exploratory study of perceptual attitudes. Unpublished doctoral dissertation, Univer. of Houston, 1953.

Klein, G. S. Adaptive properties of sensory functioning: some postulates and hypotheses. Bull. Menninger Clin., 1949, 13(1), 16-23.

Klein, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.) Perception an approach to personality. New York: The Ronald Press, 1951.

Klein, G. S. & Schlesinger, H. Where is the perceiver in perceptual theory? In J. S. Bruner & D. Krech (Eds.) Perception and personality: a symposium. Durham, N. C., Duke Univer. Press, 1949.

Köhler, W. Zur theorie des Sukzessivvergleichs und der Zeitfehler. Psychol. Forsch., 1923, 4, 115-175.

Krathwohl, D. R., & Cronbach, L. J. Suggestions regarding a possible measure of personality: the Squares Test. Educ. psychol. Measmt., 1956, 16, 305-316.

Lachman, F. M., Lapkin, B., & Handelman, N. S. The recall of dreams: its relation to repression and cognitive control. *J. abnorm. soc. Psychol.*, 1962, 64(2), 160-162.

Lauenstein, O. Ansatz zu einer physiologischen Theorie des Vergleichs und des Zeitfehler. *Psychol. Forsch.*, 1933, 17, 130-177.

Levine, M., & Spivack, G. *The Rorschach Index of repressive style*. Springfield, Ill.: Charles C. Thomas, 1964.

Lichenstein, E. The relation of three cognitive controls to some perceptual and personality variables. *Dissert. Abstr.*, 1962, 22(7), 2467.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co., 1953.

Linton, H., & Graham, E. Personality correlates of persuasibility. In C. I. Hovland and I. L. Janis (Eds.) *Personality and persuasibility*. New Haven, Conn.: Yale Univer. Press, 1959.

Livant, W. P. Grammar in the story reproductions of levelers and sharpeners. *Bull. Menninger Clin.*, 1962, 26(6), 283-287.

Luchins, A. S., & Luchins, Edith H. *Rigidity of behavior: a variational approach to the effect of Einstellung*. Eugene, Oregon: Univer. of Oregon Books, 1959.

Mackworth, N. H. *Research on the measurement of human performance*. London: His Majesty's Stationery Office, 1950.

McAndrew, M. B. An experimental investigation of young children's ideas of causality. *Stud. Psychol. Psychiat. Cathol. Univ. Amer.*, 1943, 6(2).

MacDonald, A. Experimental study of children, including anthropometrical and psycho-physical measurements of Washington school children, and a bibliography. *Report of the Commission of Education*, Washington, 1897-98.

Murney, R. J. An investigation of the cognitive system-principles of leveling and sharpening and their relationships to selected personality variables. Unpublished doctoral dissertation, Catholic Univer., 1955.

Murray, H. A. Exploration in personality. New York: Oxford Univer. Press., 1938.

Nicks, T. L. The relationship between the leveling and sharpening principle of cognitive control and retroactive inhibition. Dissert. Abstr., 1961, 22(3), 917-918.

Norton, Dee W. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. Unpublished doctoral dissertation, State Univer. of Iowa, 1952.

Pemberton, Carol. The closure factors related to temperament. J. Pers., 1952, 21, 159-175.

Pratt, C. C. The time-error in psychophysical judgments. Amer. J. Psychol., 1933, 45, 292-297. (a)

Pratt, C. C. Time errors in the method of single stimuli. J. exp. Psychol., 1933, 16, 798-814. (b)

Rapaport, D. Cognitive structures. In Contemporary approaches to cognition. Cambridge: Harvard Univer. Press, 1957.

Sandstrom, C. I., Sex differences in localization and orientation. Acta Psychol., 1953, 9, 82-96.

Sherif, M., Taub, D., & Hovland, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. J. exp. Psychol., 1958, 55(2), 150-155.

Silverman, J. The problem of attention in research and theory in schizophrenia. Psychol. Rev., 1964, 71(5), 352-379.

Smith, G. J. W., & Klein, G. S. Cognitive controls in serial behavior patterns. J. Pers., 1953, 22, 188-213.

Spivak, A. P. Leveling-sharpening as related to repression and projection. Unpublished doctoral dissertation, Mich. State Univer., 1964.

Stern, W. General Psychology. New York: Macmillan, 1938.

Stukat, K. Suggestibility: A factorial and experimental analysis. Stokholm, Sweden: Almqvist & Wiksell, 1958.

Sweeney, E. J. Sex differences in problem solving. Stanford: Department of Psychology, Stanford Univ. Tech. Rep., 1, Dec. 1, 1953.

Taylor, D. S. Paper presented at the symposium of sex differences held at the meeting of the AAAS in Berkley, Calif., Dec. 27, 1954.

Tear, D. G., & Guthrie, G. M. The relationship of cooperation to the sharpening-leveling continuum. J. soc. Psychol., 1955, 42, 203-208.

Thompson, Helen B. The mental traits of sex. Chicago: Univer. of Chicago Press, 1903.

Thurstone, L. L. A factorial study of perception. Chicago: Univer. of Chicago Press, 1944.

Torgenson, W. S. Theory and methods of scaling. New York: Wiley and Sons, 1958.

Tuddenham, R. D. The influence of a distorted group norm upon individual judgment. J. Psychol., 1958, 46, 227-241.

Tyler, Leona E. The psychology of human differences. New York: Appleton-Century-Crofts, 1956.

Vernon, M. D. Visual perception. Cambridge, England: Cambridge Univer. Press, 1937.

Wechsler, D. Cognitive, conative and non-intellective intelligence. Amer. Psychologist, 1950, 5, 78-83.

Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore, Williams & Wilkins, 1958.

Weiss, W. Role of the response system in a judgmental contrast effect. Percept. mot. Skills, 1963, 17(1), 227-241.

Wertheim, J., & Mednick, S. A. The achievement motive and field independence. J. consult. Psychol., 1958, 22(1), 38.

Winer, B. J. Statistical principles in experimental design. McGraw-Hill, 1962.

Witkin, H. A., et al. Personality through perception. New York: Harper Bros., 1954.

Woodworth, R. S. and Schlosberg, H. Experimental Psychology. New York: Henry Holt and Co., 1954.

Wulf, F. Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt) Psychol. Forsch., 1922, 1, 333-373.

APPENDICES

APPENDIX A

Squares Test

Order of presentation of the squares

(Inches)

1.	1.18	31.	1.97	61.	2.83
2.	1.57	32.	2.36	62.	3.19
3.	1.97	33.	2.83	63.	3.83
4.	2.36	34.	3.19	64.	4.59
5.	2.83	35.	3.83	65.	5.52
6.	1.97	36.	2.83	66.	3.83
7.	1.57	37.	2.36	67.	3.19
8.	2.83	38.	3.83	68.	5.52
9.	1.18	39.	1.97	69.	2.83
10.	2.36	40.	3.19	70.	4.59
11.	1.18	41.	1.97	71.	2.83
12.	2.83	42.	3.83	72.	5.52
13.	1.97	43.	2.83	73.	3.83
14.	2.36	44.	3.19	74.	4.59
15.	1.57	45.	2.36	75.	3.19
16.	1.57	46.	2.36	76.	3.19
17.	1.97	47.	2.83	77.	3.83
18.	2.36	48.	3.19	78.	4.59
19.	2.83	49.	3.83	79.	5.52
20.	3.19	50.	4.59	80.	6.62
21.	2.36	51.	3.19	81.	4.59
22.	1.97	52.	2.83	82.	3.83
23.	3.19	53.	4.59	83.	6.62
24.	1.57	54.	2.36	84.	3.19
25.	2.83	55.	3.83	85.	5.52
26.	1.57	56.	2.36	86.	3.19
27.	3.19	57.	4.59	87.	6.62
28.	2.36	58.	3.19	88.	4.59
29.	2.83	59.	3.83	89.	5.52
30.	1.97	60.	2.83	90.	3.83

91.	3.83
92.	4.59
93.	5.52
94.	6.62
95.	7.95
96.	5.52
97.	4.59
98.	7.95
99.	3.83
100.	6.62
101.	3.83
102.	7.95
103.	5.52
104.	6.62
105.	4.59
106.	4.59
107.	5.52
108.	6.62
109.	7.95
110.	9.53
111.	6.62
112.	5.52
113.	9.53
114.	4.59
115.	7.95
116.	4.59
117.	9.53
118.	6.62
119.	7.95
120.	5.52
121.	5.52
122.	6.62
123.	7.95
124.	9.53
125.	11.44
126.	7.95
127.	6.62
128.	11.44
129.	5.52
130.	9.53
131.	5.52
132.	11.44
133.	7.95
134.	9.53
135.	6.62
136.	6.62
137.	7.95
138.	8.53
139.	11.44
140.	13.73
141.	9.53
142.	7.95
143.	13.73
144.	6.62
145.	11.44
146.	6.62
147.	13.73
148.	9.53
149.	11.44
150.	7.95

APPENDIX B

Squares Test Record Sheet

Name and Sex ________________( )

Phone ________________

Instructor ________________

Section ______ Date ____

1.	26.
2.	27.
3.	28.
4.	29.
5.	30.
6.	31.
7.	32.
8.	33.
9.	34.
10.	35.
11.	36.
12.	37.
13.	38.
14.	39.
15.	40.
16.	41.
17.	42.
18.	43.
19.	44.
20.	45.
21.	
22.	
23.	
24.	
25.	

46.
47.
48.
49.
50.
51.
52.
53.
54.
55.
56.
57.
58.
59.
60.
61.
62.
63.
64.
65.
66.
67.
68.
69.
70.
71.
72.
73.
74.
75.
76.
77.
78.
79.
80.
81.
82.
83.
84.
85.
86.
87.
88.
89.
90.

APPENDIX C

Questionnaire

Name____

The following statements concern the square-size estimation task which you just took part in. Read the alternatives for each item carefully and check the one which most accurately expresses your point of view.

1. The task was: (1) very interesting___, (2) interesting___, (3) uninteresting___, (4) very uninteresting___.

2. While making my judgments, I felt that I was usually: (1) very accurate___, (2) fairly accurate___, (3) inaccurate___, (4) very inaccurate___.

3. Most of the time, I had the feeling that I was: (1) overestimating___; (2) underestimating___, (3) neither consistently over- or under-estimating___.

4. During the task, I was: (1) highly motivated___, (2) somewhat motivated___, (3) very little motivated___, (4) not motivated at all___.

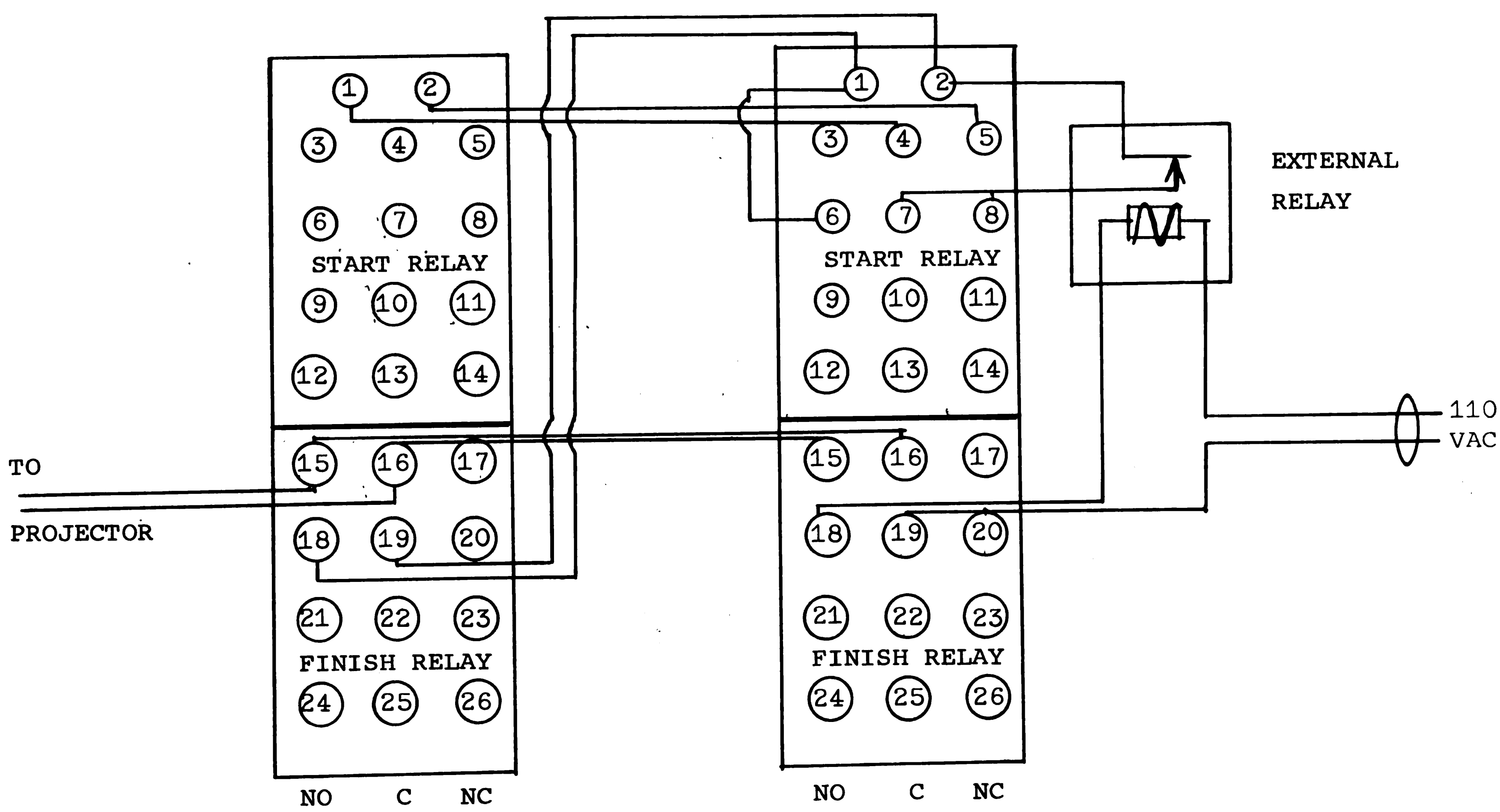
5. Who, do you think do better (on the average) on this task? (1) men___, (2) women___, (3) both the same___.

6. Briefly describe your general mood when you were taking the test.

7. If you always, or almost always, used whole numbers to make your size estimates, why did you do so?

APPENDIX D

Schematic Wiring Diagram (Hunter Timers and Projector)



APPENDIX E

Distributions of LSE Values

	LSE (inches)								
Condition	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	21
<u>Males</u>									
1-7 Regular	0	1	24						
1-7 Fractional	0	2	23						
1-12 Regular	1	4	5	6	8	1			
1-12 Fractional	0	6	6	6	3	4			
1-16 Regular	0	5	4	7	4	3	1	1	
1-16 Fractional	1	4	7	5	4	0	3	1	
<u>Females</u>									
1-7 Regular	0	1	24						
1-7 Fractional	0	2	23						
1-12 Regular	2	7	3	3	3	7			
1-12 Fractional	1	6	4	5	6	3			
1-16 Regular	1	3	4	3	10	3	0	1	
1-16 Fractional	0	3	7	8	3	2	0	2	
No Information									
Males	8	11	3	0	1	1	0	0	1
Females	7	13	1	3	0	0	0	1	

APPENDIX F

Distributions of PRA Scores

Condition	PRA Scores 50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-
Males										
1-7 Regular					9	4	6	3	3	
1-7 Fractional					2	5	6	10	2	
1-12 Regular	1	0	2	2	7	5	4	2	2	
1-12 Fractional			1	0	4	4	8	4	3	1
1-16 Regular					9	6	3	4	3	
1-16 Fractional					6	9	3	6	1	
Females										
1-7 Regular		1	0	2	9	5	5	2	0	1
1-7 Fractional					3	3	10	5	4	
1-12 Regular		2	1	5	4	5	5	2	1	
1-12 Fractional				1	1	5	5	10	2	1
1-16 Regular				5	4	6	6	4		
1-16 Fractional			1	1	2	5	5	7	4	
No Information										
Males			3	4	3	5	6	1	2	1
Females			3	5	4	6	4	2	1	