IMPROVING INTERPERSONAL SENSITIVITY: AN EVALUATION OF A STEREOTYPE ACCURACY TRAINING PROGRAM

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Morris S. Spier
1969

This is to certify that the

thesis entitled

IMPROVING INTERPERSONAL SENSITIVITY:

AN EVALUATION OF A STEREOTYPE ACCURACY TRAINING PROGRAM

presented by

Morris S. Spier

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in ~~Psychology~~

Date May 29, 19

ABSTRACT

IMPROVING INTERPERSONAL SENSITIVITY:

AN EVALUATION OF A STEREOTYPE ACCURACY TRAINING PROGRAM

by Morris S. Spier

The word "stereotype" has taken on a negative connotation, and has become synonymous with "bias" and "prejudice". But there is ample evidence in the literature that stereotypes do not automatically lead to poor understanding of others. Rather, it is the rigid adherence to an inaccurate stereotype that causes poor understanding. Accurate, flexible stereotypes may actually enhance interpersonal sensitivity. The primary purpose of the present research was to determine whether stereotype accuracy could be improved through participation in a formal training program.

Two experimental training programs used a pretest-training-posttest paradigm to improve stereotype accuracy (SA): the ability to predict group norms. In Study One, a pilot program, training improved predictions about marital stereotypes even though training was provided on executive, psychologist, college student, and professor stereotypes. But the criterion instruments were unwieldy and the single criterion design gave little control over what occurred between pretest and posttest, while a great deal of sample shrinkage took place from the first to the last

SA training session.

Study Two's revised and shortened criterion-training instruments made each training session a complete unit in itself. Significant SA improvement occurred in the second study, but improvement on one stereotype did not assure improvement on others, and vice versa.

Improvement in Study One suggested generality of the underlying principles and training designs. The specificity of SA and its improvement, research design requirements, and ease of administration, however, favor the single session unit design of Study Two. Moreover, combinations of training activities within a program emphasizing both content and process hold the greatest promise for maximization of training impact. Practice and feedback formed a core for the additional discussions, diagnosis, base-rate, noncommitment, pooling, and panel procedures and conditions of the present studies.

An additional finding was that Js who did not make written pretraining predictions improved about as much as Js who had made the written judgments.

Additional data showed that the personality and intellectual correlates were also stereotype specific, and that initially inaccurate Js benefited more from training than initially accurate Js. Both findings had relevance for the selection of trainees.

Finally, moderately difficult criterion items best reflected

training outcome from a statistical point-of-view. From a practical standpoint, however, the differences were not worthwhile. SA measures should continue to reflect judging reality and include the total range of item difficulties. Subjective feelings of improvement and liking for training were unrelated to actual outcome which precludes their use as criteria for training effectiveness.

Implications of the findings for future training and research were discussed.

Approved: _____

Date: _____

IMPROVING INTERPERSONAL SENSITIVITY:

AN EVALUATION OF A STEREOTYPE ACCURACY TRAINING PROGRAM

By

Morris S. Spier

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1969

G61355
4-3-79

To Alice

ACKNOWLEDGMENTS

The completion of a dissertation is an opportunity for reflection on the preceeding years of preparation and on the people that made it possible. The author wishes to express special appreciation to the members of his committee, Drs. T. Allen, J. Wakeley, and C. Williams. Their patience and support were important not only in the completion of this manuscript, but throughout the author's graduate career.

Deepest appreciation is due Dr. H.C. Smith, chairman of the author's committee, who provided guidance, support, and motivation in the present research, and who taught by his personal example what it means to be a professional psychologist.

Special thanks are also extended to Drs. N. Abeles and H. Stern for their considerable contributions.

Finally the author thanks his wife, Alice T. Spier, who seems to instinctively know what it means to be sensitive. She more than earned her Ph.T..

TABLE OF CONTENTS

TABLE OF CONTENTS--(Continued)

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

The primary aim of the present research was the improvement of the ability to understand others through participation in a formal training program. Two groups, each containing nearly 200 Michigan State University students, took part in respective 10 week experimental training programs designed to increase their interpersonal sensitivity (Mietus, 1969; Price, 1969; Smith, 1968a). The present studies were concerned primarily with the training sessions aimed at improving stereotype accuracy: the ability to accurately predict group norms. Since an important practical consideration of the research was the development of general training principles, the focus could as well have been any component.

Chapter I of the following report briefly reviews the recent literature with regard to the definition, assessment, and attempts to improve sensitivity. The data collected in the experimental training programs are reported in Chapters II and III. Chapter IV contains a discussion of the results including their relation to previous research, and Chapter V is a summary of the present research and its findings.

CHAPTER I

HISTORY OF THE PROBLEM

> The majority of men are subjective towards themselves and objective towards all others, terribly objective sometimes -- but the real task is to be objective towards oneself and subjective towards all others.
>
> Soren Kierkegaard
> (The Journals, Oxford University Press, 1847)

Interest in the ability to understand people existed long before experimental attempts to define, measure, or develop it. Indeed, interpersonal sensitivity has long been an elusive goal of the psychologist, manager, educator, and salesman alike. This pervasive interest appears in each man's occupational role: the psychologist wishes to understand his client; the manager, his employee; the educator, his student; the salesman, his customer; and as social beings, all want also to understand their families, friends, and peers.

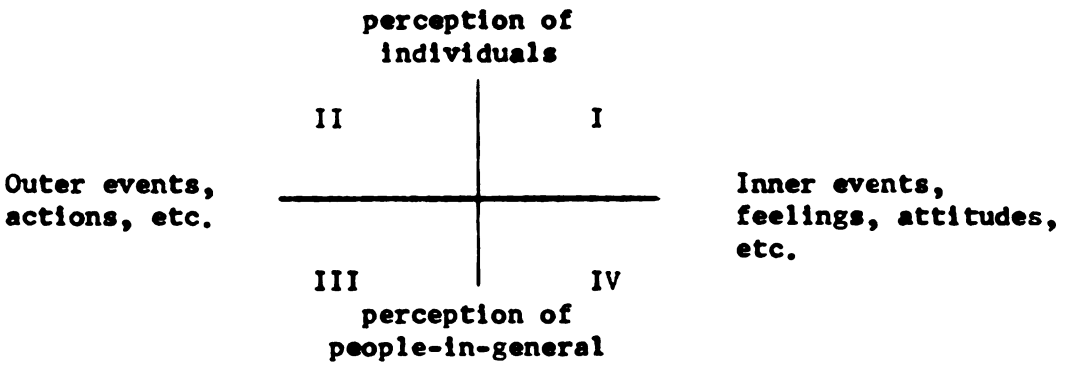
The Nature of the Ability

Everyone wants to be sensitive, but not everyone can convincingly demonstrate his understanding of others, nor even communicate what it means to be sensitive. Smith's definition of sensitivity as "the ability to predict what an individual will feel, say, and do about you, himself, and others" (1966, p. 3),

places the assessment of the ability in a central position. Prediction is the best, if not the only, test of understanding people. To be sensitive is to be able to accurately predict both private data (what a person will feel) and public data (what a person will say and do) about others.

Gough's (1968) grid approach is reproduced in Figure 1 as an aid to conceptualizing the tasks and goals of interpersonal sensitivity. The key question seems to be "What do you want to understand about whom?"

Figure 1: Suggested domain sampling for appraising social acuity. (Reproduced from Gough, 1968, p. 3)



Gough's schema retains an emphasis on predictive accuracy while isolating the "what" and "whom" as the two major dimensions of interpersonal perception. As in Smith's definition, the "what" of understanding is contained on a dimension from inner to outer events (private to public). In the manner of Bronfenbrenner, Harding, and Gallwey (1961) who dichotomized the task of the perceiver into "Interpersonal Sensitivity" and "Sensitivity to the Generalized Other", Gough's "whom" ranges along a dimension

from the understanding of a specific individual to an understanding of people-in-general. While the "good judge of others" should be able tc demonstrate his predictive accuracy in each of the quadrants defined by the interaction of the dimensions, the present investigation was concerned only with predictions in quadrants III and IV: the prediction of behavioral, attitudinal, etc. norms of groups.

The Measurement of the Ability

Concern with "who is sensitive" led to the development of an assessment paradigm with predictive accuracy at its core. Typically, a person (O) fills out a personality, interest, attitude, or other inventory. A judge (J) is introduced to O via filmed or tape interviews, face-to-face contact, personality sketches, case histories and/or other descriptive data. J then completes the personality, interest, or attitude questionnaire as he (J) feels O filled it out. The measure of J's understanding of O then, is the number of items correctly predicted.

In 1955, Cronbach analyzed such accuracy scores and concluded that they represented an oversimplification of what was, in fact, a factorially complex ability. Cronbach pointed out that the judge with the highest score was not necessarily the most understanding of others. Accuracy in judging O might result also from J's response tendencies, his understanding of groups, or even the degree of similarity to O assumed by J. Thus, for example, the J who habitually assumes a great deal of similarity between himself and others would achieve spurious accuracy in cases of a high degree of actual similarity.

Smith (1966, 1968a, 1968b, 1968c) labeled response sets, evaluative accuracy (level); knowledge of groups, stereotype accuracy; and assumed similarity, empathic accuracy. The level component is based on the assumption that our judgments of others invariably contain an evaluative element. Indeed, there is a tendency to implicitly or explicitly assign goodness or badness to our descriptions of others. Campbell (1967), for example, pointed to the evaluative aspects of stereotypes. The English describe themselves as reserved and respecting the privacy of others, while viewing Americans as intrusive, forward, and pushing. Americans, on the other hand, saw the English as snobbish, cold, and unfriendly, and thought themselves friendly, outgoing, and open-hearted.

Our judgments of people are also inevitably influenced by the groups to which they belong. To the extent that knowledge of the groups is incomplete and inaccurate, as it necessarily must be, we will err in our judgments.

A final assumption postulates a strong motivation to determine ways in which people around us are similar or dissimilar to ourselves. Thus, Smith (1966, p. 19) defined empathy as "the tendency of a perceiver to assume that another person's feelings, thoughts, and behavior are similar to his own". In trying to understand others, we usually assume too little or too much similarity between ourselves and them. The goal of sensitivity training, then, should be to introduce accuracy into the evaluative,

stereotyping, and empathic processes.

Training to Improve Sensitivity

Evaluations of training program effectiveness either equate training with amount of clinical experience or psychological course-work of the judge, or administer a pretest and posttest with interpolated training. Evidence for the effectiveness of either type of program is difficult to obtain, and to complicate matters, all too often the actual training procedures remain unspecified (Wakeley, 1961). Some other programs depend on anecdotal evidence and the participants' "liking for training" or confidence in their improved judgment as "proof" of training value.

In 1955, Taft reviewed the previous three decades of research into the effectiveness of training at increasing understanding of others. Most of the studies followed the educational model. Taft concluded, after controlling for similarity of Js and Os in an academic setting, that "physical scientists, and possibly other nonpsychologists, e.g., personnel workers, appear to be more capable of judging others accurately than either psychology students or clinical psychologists.... There is also evidence that suggests that courses in psychology do not improve ability to judge others and there is considerable doubt whether professional psychologists show better ability to judge than do graduate students in psychology" (p. 12).

Eleven years later, in 1966, Smith's review included five studies carried out since Taft's review. The picture remained gloomy: training did not increase sensitivity.

Fancher (1967), as part of a larger study, speculated that those in his sample of student Js who were most similar to clinical psychologists should also make the most accurate predictions. Results: experience (number of psychology courses) and competence (grade in an abnormal psychology course) correlated -.17 and .05, respectively, with accuracy. His findings further suggested that the professional preparation of clinical psychologists makes its greatest impact not on the accuracy of judgments, but on the validity of conceptualizations of others.

Goldberg's (1968) review found little success in training attempts to date. Similarly, Campbell and Dunnette (1968) reviewed the literature covering the effectiveness of T-group experiences. Conclusion: participants emerge with a larger vocabulary of interpersonal terms, but with no demonstrable increase in interpersonal accuracy.

The findings are also generally negative with regard to the usefulness of "confidence in accuracy or improvement" as a criterion of training effectiveness. Oskamp's (1965) judges, for example, estimated the percentage of 25 items they would predict correctly, and then made predictions on the basis of minimal information about an O. While initially only slightly overconfident, Js (including experienced clinical psychologists) became more and more certain of their accuracy as the amount of information about O was increased. Finally, having read all of the data about O, the average J estimated 53% correct. 28% were actually

correct. Oskamp concluded that "the judges' confidence ratings show that they become convinced of their own increasing understanding of the case... While... their certainty about their decisions became entirely out of proportion to the actual correctness of those decisions" (p. 264).

Opsahl (Dunnette, 1968) used a computer technique to control for 3 possible sources of variance in interpersonal prediction: assumed similarity, base rate of item endorsement, and social desirability. Assessment instruments developed by the computer used all combinations of the above variables. In the 5 resulting treatments, target Os appeared in filmed interviews. In a sixth condition, however, Js made predictions based only on stereotypic data (O's age, sex, education, and occupation) with all 3 controls applied. All Js participated in all treatments including a feedback or no feedback condition. For each response, Js indicated their degree of confidence in the correctness of their prediction as either Not Confident, Somewhat Confident, or Very Confident. Among the findings: under the film case conditions with all controls either present or absent, accuracy and confidence were strongly related. When controls were only partially relaxed, Opsahl's Js, like those of Oskamp (1965), were confident of the correctness of more of their answers without actual increases in accuracy. Finally, where Js predicted only on the basis of stereotypic data, confidence was unrelated to actual accuracy.

In a cross-cultural study (Gough, 1968), 231 American Js completed a 64 item questionnaire as they thought male and female Italian Os would answer. The Js viewed films of the Italian Os and completed the criterion instrument a second time. Js reported, both in groups and individually, that they found the task "meaningful", "interesting", and "stimulating". Result: "After viewing the film, and in spite of subjective feelings of greater confidence and insight, the American viewers declined in accuracy" (p. 8). Mean accuracy before viewing the film was 36.87 items correct: after viewing the film, 32.90 items correct, a statistically significant reduction.

Principles of Training

As Gough's experimental design suggests, some researchers have gone beyond the traditional educational training grounds in attempts to develop sensitivity training programs. Many of these attempts have carefully specified the procedures, materials, etc., used in training. Thus, some generally effective training procedures have emerged from the programs, among them: practice at making judgments, feedback about accuracy, discussion, diagnosis, use of an explicit empirical personality theory, an empathic technique, pooling, delay of impression formation, and the use of base-rate data.

Practice and feedback.

Their facilitative effects on learning in general (Cline, 1968; Dunnette, 1968; Dunnette and Hakel, 1968; Locke, 1967;

Smith, 1966) have made practice and feedback (knowledge of results) central components of attempts to improve understanding. Cline's (1968) "feedback in judging accuracy training research" provided experimental evidence for the usefulness of feedback in increasing sensitivity. Eight groups of Js took tests of 3rd person accuracy (i.e., inferred how O's behavior and personality were rated by some significant others; Bronfenbrenner, et al., 1961) after being exposed to one of 8 possible combinations of 3 dichotomous variables: feedback after each response to each item versus no feedback; a brief personality summary versus no summary; filmed interviews versus no film. A ninth group (control) made its predictions only on the basis of O's sex and age. Results: feedback groups made significantly more accurate 3rd person predictions of behavior and personality than the no feedback groups. Furthermore, feedback was the most powerful discriminant on both the behavioral and personality trait predictors.

Goldberg (1965) administered massive doses of practice and feedback to experienced clinicians, psychology graduate students, and naive Js as part of a program to improve diagnostic accuracy. Goldberg established both training and testing sets of tasks from 1530 MMPI profiles with criterion diagnoses of neurosis and psychosis. In 9 weeks of training, Js received intensive practice in predicting neurosis or psychosis on over 4,000 MMPI profiles followed by immediate feedback after each guess. In all, when both training and testing sessions are combined, each J viewed

over 10,000 profiles. Results: expert and middle Js showed little change from their initially similar levels of accuracy, and while naive Js showed substantial improvement, their accuracy was still well below that of the more experienced Js. Practice and feedback, it appears, may be facilitative, but are not in themselves, sufficient for optimal learning.

Participation-discussion.

Jecker, Maccoby, and Breitrose (1965) added a participation-discussion component within the practice and feedback framework to improve teachers' use of non-verbal cues to determine comprehension. Sound film clips collected in classroom settings were edited to show a single student during the instruction of one item and the time allotted to answer the question on that item. Js guessed whether the O answered the question in each clip correctly or incorrectly. Twenty-five such clips comprised a pretest. An experimental group received training: i.e., viewed 15 to 30 clips; identified the non-verbal cues; made judgments; discussed the correct scoring, interpretation, and meaning of the cues; and were given feedback regarding the accuracy of their predictions. A control group viewed and discussed a film on interpersonal communication. Both groups were then posttested. Results: when scores were equated for initial accuracy, experimental Js showed a mean gain in accuracy of 7.2% while the control Js declined .5% in accuracy. The difference was statistically significant.

A programmed learning approach toward increasing understanding (Dailey, 1966a, 1966b) also stressed practice and feedback. Dailey edited published biographies and personally collected case histories to identify key events in a target O's life. A completed programmed case contained 10 to 15 such key episodes arranged chronologically and each paired with two alternative, but incorrect, events. Each episode required J, initially presented only with O's occupation, to predict what would happen in the next segment. Immediate feedback followed the prediction. Thus, a formalized program of practice and feedback could be combined with role-playing exercises, group discussions, etc., to form a complete laboratory (1966a). In one such application, Dailey presented his Js with 12 cases, 2 at a time. Mean predictive accuracy increased 17% from the first 2 cases to the last 2 cases, a statistically significant improvement (1966b).

Kepes' (1965) aim was to improve individual accuracy (Gough's (1968) quadrants I and II) through a program of practice, participation-discussion, and feedback. Js viewed a sound film interview of 3 men and 3 women and completed an inventory based on the film calling for 2nd person predictions in the form of differentiating between the film Os (i.e., inferences about how an O rates himself; Bronfenbrenner, et al., 1961) and recall of specific mannerisms, dress, speech, etc. (observational accuracy) of each O. Since the focus of training was individual accuracy, the 120 item criterion instrument was constructed to be free of

evaluation (level) and stereotype influences. Following pretest administration of the film based inventory, an experimental group of 58 Js went through 8 weekly training sessions. In a typical one hour meeting, Js answered a series of case study items, discussed reasons for responding as they did, received feedback about accuracy, proceeded to the next series of items, and so on, until the case was completed. A control group received no training, but both groups were posttested with the film based inventory. When Js were matched for initial accuracy, there was no significant differences between pretest and posttest for either group. A trend of gains in 2nd person accuracy was found for the experimental group, but no such trend occurred for the control Js. In a second experiment, Js were pretested using only the 3 men interviews on the film based inventory. The training involved discriminating among the 3 women interviewees who were presented in film, on tape, live, and in written cases to 4 groups, respectively. The men film based inventory was readministered after training with the result that training on the women interviews did not generalize to accuracy about men. There was no difference in accuracy among the treatment groups and even the general trends of differential gain noted in the first study failed to replicate.

Diagnosis of error tendencies.

In his first study noted above, Kepes found good gains in observational accuracy among experimental Js. Similar improvement

occurred among the control Js with no training. After pretesting, both groups were given their observational accuracy pretest scores and told that they would be tested again at some later date. While seemingly only a special case of feedback, Kepes felt that such diagnosis of present proficiency motivated both groups to improve (Locke, 1967).

Grossman (1967) made diagnosis of error tendencies a formal part of his training program. Attempts were made with three experimental training groups to improve evaluative accuracy, empathic accuracy, and individual accuracy, respectively. Js were pretested and posttested on each of the components regardless of the focus of training to follow. Grossman's training focussed primarily on individual accuracy and stressed practice, participation-discussion, and feedback. Improvement due to training was statistically significant. Comparable gains in individual accuracy, however, were recorded for members of the experimental groups which received specific training on the other components. Grossman, like Kepes, noted that each J had been apprised of his evaluative and empathic error tendencies and felt that this knowledge acted both as a motivator and a basis for change.

An explicit personality theory.

Grossman (1967) was somewhat less enthusiastic about the effectiveness of an empirical personality trait theory in increasing accuracy. Smith (1966, 1968a) noted that we all use

implicit, untested theories in our attempts to understand others. He proposed (1966) that Js be taught an explicit empirical personality trait theory to provide a framework within which they could change their own theories and thus improve their understanding. Grossman (1967) instructed Js in the nature and intercorrelation of the 5 trait dimensions (Linden, 1965): impulsive-controlled, rational-empirical, introverted-extroverted, cautious-bold, and emotional-calm. Next, he introduced Js to the use of the traits in differentiating between Os and gave opportunity for practice and feedback in making differential predictions. He concluded that the theory contributed little to improvement.

An empathic technique.

We often use ourselves as a basis for prediction about others. Such tendencies may provide valuable input when we are accurate in our assumptions of similarity or dissimilarity. Diagnosis, practice, discussion, and feedback are useful tools also in increasing sensitivity through an empathic technique (Mietus, 1969; Silkiner, 1962; Smith, 1966, 1968a). Such a program should emphasize both the content and process of empathic accuracy.

First Js would be instructed in the nature of empathy and its relation to sensitivity. Second, they would complete, for example, an interest inventory answering for their own likes and dislikes. Then they answer the same inventory as they felt O, a typical group member, would respond. Suppose that a typical O

had actually answered "like" to the first item. Given feedback about Os actual answers, J would find that if he had answered "like" for himself and for O, he would be correctly assuming similarity. If J answered "like" for himself but "dislike" for O, he would be incorrectly assuming dissimilarity. If he answered "dislike" for himself and for O, he would be incorrectly assuming similarity. If he answered "dislike" for himself and "like" for O, he would be correctly assuming dissimilarity. Such an exercise provides both a diagnosis of empathic tendencies and a demonstration of the influence of empathic assumptions on accuracy.

Pooling.

Wakeley (1961) focussed on prediction as a process of information gathering and interpretation culminating in the actual recording of judgments. Experimental training programs dealt with each of these tasks. Two programs aimed at improving J's observation (information gathering) by instructing J to focus on his own reaction to O, or to focus on O when observing. Two other programs (information processing) instructed Js to either concentrate on the uniqueness of each O, or on O's similarity to other people known by J (inferring-pooling). Another program focussed on the use of rating scales, and a final program combined the techniques of the others. Js were pretested and posttested with a test of the ability to differentiate between people regardless of the training Js received. Only the inferring-pooling training, which advocated a kind of stereotyping,

and the program combining all the techniques increased accuracy. A second study replicated the results.

Delay of impression formation.

We almost invariably form instantaneous impressions of others (Smith, 1966; 1968c). Springbett (1958), for example, demonstrated that employment interviewers made their decisions to hire or reject within the first 4 minutes of the interview. We do not seek out a person's specific qualities and then attempt to form a general picture of him, rather, we move from the immediate overall impression to the search for specific traits. Our search, however, is indelibly influenced by our early impression. Accuracy should be improved, therefore, by slowing down the impression formation process. Dunnette and Hakel (1968) were unable to specify a procedure, but felt that sensitivity training should aim to develop a delay of impression formation and instill in J "an awareness of the necessity to gather all information prior to making or 'fixing' early impressions or judgments" (p. 38).

Use of base-rate data.

Opsahl (Dunnette, 1968), it was noted earlier, controlled for response set strategies in his computer generated sensitivity measure by correcting for base-rate of item endorsement. Items were matched according to the relative proportion of people in general who endorse (or reject) a particular item. Such knowledge, however, when accurate, could aid in understanding.

Lorge and Diamond (1954) asked Js to predict whether or not student Os would pass ability test items. After pretesting, Js were classified as poorest, mediocre, or best and "trained" by being told the difficulty level of some specimen items. When Js responded only to the items, the average correlation between predictions and actual outcomes for the best Js was .73, and for the poorest Js, .56. After base-rate information was given, the same Js had mean correlations of .77 and .73. Conclusion: "Apparently the difference between 'poorest', 'mediocre', and 'best' judges is that the 'best' judges have some experiential referent for the percentage of the population that can pass an item. Giving such referents to the 'poorest' and 'mediocre' judges... leads to a significant reorientation of such judgments" (p. 33).

Stelmachers and McHugh (1964) used MMPI norms to develop a scoring key around the base-rate of item endorsement of 3 broad groups: college females, adult females, and adult males. Experienced psychologists, psychiatrists, and psychiatric social workers predicted the responses made by two women and two men to 171 MMPI items. The female Os were a normal college sophomore, and an elderly woman with a long medical and psychiatric history. The male Os were a teenage homosexual and an adult depressive. The authors made similar predictions using only the special base-rate key. Results: the special base-rate keys were more accurate than the expert Js for 3 out of the 4 criterion Os.

Statement of the Problem

The primary purpose of the present research was to evaluate the effectiveness of a pilot and a revised training program designed to improve stereotype accuracy (Gough's (1968) quadrants III and IV).

For decades the social science literature has supported the view that stereotypes are simplistic, virtually impossible to alter, and never accurate (Harding, 1968). Stereotypes have become synonymous with prejudice and insensitivity despite a growing body of empirical evidence which suggests that the converse is true (Cline, 1968; Tagiuri, 1968).

Attempts have been made to increase sensitivity, but attempts to increase stereotype accuracy, i.e., the ability to predict group norms, have been largely ignored. While a number of investigators have developed measures of the ability and begun to explore its nature (Harris, 1962; Johnson, 1963; Shears, 1967; Silkiner, 1962; Zavala, 1960), training activities have attempted to eliminate stereotyping influences on judgments.

Study One, a pilot study, proposed to: (1) form a stereotype accuracy training program based on principles established in the research literature; (2) test the hypotheses and suggest new ones; and (3) generate hypotheses about training effectiveness, means for measuring effectiveness, and the relation of training impact to personality, intellectual, and attitudinal factors. Study Two was a further test of hypotheses derived from the pilot

program as well as an opportunity to apply the findings to the development of a revised program.

The following hypotheses guided Study One.

Hypothesis 1: A training program will be effective in improving stereotype accuracy.

Hypothesis 2: The effectiveness of training varies inversely with initial accuracy.

Hypothesis 3: The effectiveness of training will be reflected more clearly on stereotype accuracy predictor items of moderate difficulty, than on those of either high or low difficulty.

Hypothesis 4: Stereotype accuracy is a situation specific ability.

Hypothesis 5: The effectiveness of training varies with personality, intellectual ability, and attitudes toward training.

Chapter II discusses the methods and results of Study One.

CHAPTER II

STUDY ONE

General Design

Study One used a pretest-posttest design with interpolated training. Subjects were administered a single criterion, received training in making predictions about specific groups or group members, and were posttested with the criterion measure used in the pretesting. Training took place in 4 successive sessions of about 1 hour and 10 minutes each.

Subjects

Judges (Js) in this study were 200 Michigan State University students, representing a wide range of college majors and levels, who were enrolled in a Fall 1967 section of a Psychology of Personality course. While training took place during the twice weekly class meetings and as a part of the regular class-work, attendance was not mandatory.

Measures of Stereotype Accuracy (SA)

The paper-and-pencil measures of stereotype accuracy included tests of accuracy in judging the difference between:

(1) happily married, unhappily married and divorced men and women;

(2) executives and unskilled workers;

(3) a professor and college men in general;

(4) psychologists and men-in-general; and

(5) a particular man (Morgan Johnson) and men-in-general. Copies of the instruments appear in Appendix A. An IBM scoring sheet, provided for use in recording answers, allowed machine scoring of all tests.

1. The criterion instrument used in the present study was based on the research of Johnson and Terman (1935) concerning the relation of marital happiness to similarity (in interests and personalities) of husbands and wives. The development of the criterion measure is of interest in that it illustrates the construction of all of the dependent variables.

Johnson and Terman compared the resemblance of partners in three groups: happily marrieds, unhappily marrieds, and divorced couples. One hundred couples in each category were matched to control for other relevant group memberships (e.g., occupational level, socio-economic status, age, religion, national origin, etc.), responded to lengthy personality and interest inventories, and participated in in-depth clinical interviews. Results: no relationship between marital happiness and husband-wife resemblance, but many of the individual inventory items correlated with marital happiness and thus distinguished between the members of the three groups.

Close (1963) used Johnson and Terman's findings to develop the Marriage Test, the criterion measure of the present study. The instrument measures a J's ability to differentiate between

typical happily married, unhappily married, and divorced people. Part I deals with men (28 items); Part II deals with women (28 items). The two-part construction served to partial out another significant group membership: sex. The instructions read:

> For each of the following statements, mark the one of the three groups you think is best described by it.
>
> The correct answers are based on the answers that the groups made on lengthy questionnaires. For example, 100 happily married women, 100 unhappily married women, and 100 divorced women answered the question: "Do you prefer a play to a dance?". Results:
> 81% of the happily married women answered "yes".
> 58% of the unhappily married women answered "yes".
> 44% of the divorced women answered "yes".
> Therefore, the correct answer to the statement, "Most apt to prefer a play to a dance" is "happily married women."

While these measures have no demonstrated empirical validity, their construction suggests that they do have content validity. The instruments purport to measure stereotype accuracy: the ability to predict group norms. The correct answers throughout, are based on the actual responses of the Os and are not the opinions of the trainer. Thus, a J's stereotype accuracy is the number of correct predictions he makes.

2. The sixty item test of The Interests of Executives vs. Unskilled Workers required the J to differentiate between professional and nonprofessional men in their answers to an interest inventory (Silkiner, 1962). J was informed that:

> A large number of executives and professional men (lawyers, managers, etc.) and a large number of unskilled workers (laborers, porters, etc.) checked whether they liked or disliked each of the interests or activities below. A larger percentage of the unskilled workers like

> half of the interests; a larger percentage of the executive and professional group liked the other half.
>
> Mark "1" if you think more unskilled workers liked a particular interest.
> Mark "2" if you think more professional workers liked it.

3. The Case of the Psychology Professor provided separate exercises for the development of level, stereotype, and empathic accuracy. Js were told the professor's position on five personality trait dimensions and then made predictions about how the professor answered the items on a personality inventory. The Stereotype Training exercise is a two-part test of the ability to predict the ways in which the professor conforms to the stereotype of college men in general (part 1) and the ways in which he differs from the group norm (part 2). There are twenty true-false items on each part of the test. The instructions to part 1 read:

> The statements below were in the Inventory completed by the professor. On these particular statements, as it turned out, he gave the same answers as those given by more than two-thirds of the college men who completed the inventory. The question here, therefore, is not just how well you understand the professor. It is also how well you understand the typical college man.
>
> Circle the answer "true" if you think both the college professor and the typical college man answered it "true".
> Circle the answer "false" if you think both the college professor and the typical college man answered it "false".

The instructions to part 2 read:

> To the following statements, the professor gave a different answer to the one given by two-thirds or more of the college men. The question here, therefore, is, How does the professor differ from the typical college man?
>
> Circle "true" if you think the professor answered "true" but the typical student answered "false".

> Circle "false" if you think the professor answered "false" but the typical student answered "true".

4. The Special Interests of Psychologists (60 items) required J to differentiate between psychologists and men-in-general on the basis of their responses to a series of vocational interest blank items (Silkiner, 1962). The directions for the test read:

> How do the interests of psychologists differ from those of other men? To answer the question, several hundred male psychologists and several thousand other business and professional men checked whether they would "like" each of many different occupations, amusements, activities, and kinds of people.
>
> A higher percentage of the psychologists liked some interests. For example, 41 percent of the psychologists said they would like to be the "author of a novel"; only 32 percent of men in general expressed such a liking. A lower percentage of the psychologists liked some interests. For example, only 29 percent of the psychologists said they would like to be a "sales manager" whereas 37 percent of men in general expressed a liking for this occupation.
>
> Mark for each of the interests below whether you think more or less psychologists liked the interest. Mark "1" if you think a higher percentage of psychologists than men in general liked the interest; mark "2" if you think a lower percentage of psychologists liked the interest.

5. The Case of Morgan Johnson asked J to differentiate between the interests of a college senior majoring in psychology and those of men-in-general. The 64 item test begins with a brief biographical sketch of "Morgan Johnson" (see Appendix A) and contains 2 sets of 32 items each. The first set of items ("Morgan is like the typical man") notes that:

> Morgan filled out the Strong Vocational Interest Blank that requires the respondent to answer "Like," "Indifferent,"

> or "Dislike" to a long list of intersts. The directions of this test ask the respondent to "disregard considerations of salary, social standing, future advancement, etc. ... consider only whether or not you would enjoy the interest regardless of any necessary skills, abilities, or training which you may or may not possess."
>
> In the first group of interests below Morgan gave the <u>same</u> answer as more than half of several thousand representative American men. Mark the answer that you think was given by both Morgan and the typical men. Use "1" for "Like"; "2" for "Indifferent"; and "3" for "Dislike".

In the second series of 32 items J considered the ways in which "Morgan is unlike the typical man."

> In the group of interests below, the answer given by Morgan is the one in small letters (l,i,d). His answers tc these interests, however, were <u>different</u> from those given by the majority of men. From the two possible answers not chosen by Morgan that are indicated by capitals (L,I, or D) mark the one that you think was chosen by the typical man.

<u>Training Procedures</u>

Study One used three basic training procedures:

1. A program of diagnosis of error tendencies, practice in making judgments, feedback regarding accuracy, and discussion;
2. A pooling technique; and
3. An empathic technique.

Each procedure emphasized various aspects of the SA training program.

The first group of procedures (diagnosis, practice, feedback, and discussion) stressed the SA principles.

> -Our impressions of a group and its members are necessarily

based on inaccurate and incomplete information, but stereotypes are not, in themselves, the cause of poor prediction (understanding).

-Rather, it is the rigid adherence to an inaccurate, incomplete stereotype that interferes with prediction.

-An accurate, flexible stereotype can aid in prediction.

The process of diagnosing errors involved the administration of a pretest which required prediction about a specific group. Formally, a 15 item, true-false Test of the American Worker (see Appendix A) demonstrated that an inaccurate stereotype results in poor (inaccurate) prediction. Js predicted workers' responses to statements about their jobs. A low score was indicative of an inaccurate stereotype of workers. On a less formal level too, Js got feedback about the accuracy of their stereotypes as they practiced applying them. They also received knowledge of results and discussed their predictions for a wide range of groups.

Practice in applying a stereotype to a single group member demonstrated the need for maintaining flexibility. A particular O may fit the group norm in some ways, but will almost certainly differ from it in other ways. Hence, the accurate stereotype that is a "best bet" about a person when we first meet him will have to be modified.

Each J recorded his accuracy scores on a Personal Profile Sheet (see Appendix A) which contained norms for the class as a whole. He then compared his accuracy with that of the other members

of the class.

Within the program of diagnosis, practice, feedback, and discussion, the pooling and empathic procedures stressed the process of acquiring information about group norms, and the psychological organization of the information in order to make more accurate judgments.

Wakeley's (1961) pooling method had J pick from among people he already knew well, a representative "pool" of people that most closely matched O. J was instructed to assume that the O judged thought, felt, and behaved as the average O in the matched pool. The following instructions (Smith, 1966, p. 146) were read to the Js.

> In the course of your living you have obtained a great deal of information about many people. The pooling principle simply suggests that you use this information when making inferences about a person with whom you have had little contact. When you are attempting to make inferences about a person whom you do not know well, one of the things which you can do is to form a pool of people whom you do know well who are like the unknown person. You take what you do know about the person, form a pool of people you know well, and then make your predictions or judgments based on the pool. The important things to remember in making these pools are to use people you know well and to use all of the information you have about the person you are trying to judge. You may form a pool that leads to wrong predictions if you use just one piece of information about the person, such as, his skin color, his religious preferences or any other single piece of information. You may also form some pools that lead to wrong predictions if you use people whom you do not know well.

Js listed, at the top of their answer sheets or test booklets, the names of as many individuals that they knew well who matched

the characteristics of the O. Thus, the pooling technique encouraged conscious utilization of information and experience gleaned from past encounters with similar people.

Smith defines "empathy" as "the tendency of a perceiver to assume that another person's feelings, thoughts, and behavior are similar to his own" (1966, p. 19). In using the empathic technique as a part of the SA training program, Js used their own responses to the tests as a base of similarity-dissimilarity to the Os being judged. For example, The Special Interests of Psychologists required J to give his own responses directly on the test booklet. The instructions were:

> Mark "1" if you would say "like" or "yes" to the interest or statement;
> Mark "2" if you would say "dislike" or "no" to the interest or statement.

J completed the test by predicting the interests of the psychologists. The correct answers were read to the class. J compared the number of items he answered correctly with the number of items he answered (or expected to answer) in the same way as psychologists. J was thus encouraged to focus on both the benefit and hindrance to understanding of assuming similarity or dissimilarity between himself and the person being judged.

Contents of the Training Sessions

One or more of the training procedures were combined with each of the SA measures to comprise a single training session. Table 1 summarizes the content of the 4 meetings devoted to SA training within a larger program of training to improve sensitivity

to others. SA training took place in the 6th, 7th, 8th, and 15th sessions of the overall program. The criterion was readministered (posttest) in the 17th meeting.

Table 1. Summary of training activities in Study One.

Session	Stereotype	Pretest	Training Procedure* (1)	(2)	(3)	Posttest
6	Marital (Criterion)	X				
6	American Worker		Diagnosis Only			
7	Morgan Johnson	X	X			
8	Executives-Workers	X	X			
8	Psychologists	X	X	X	X	
15	Professor	X	X	X		
17	Marital (Criterion)					X

*Note: (1) = diagnosis, practice, feedback, discussion; (2) = pooling; (3) = empathic technique.

Training session 1. -- Js were given the criterion instrument (the Marriage Test), the Test of the American Worker, a scoring pencil, and an IBM answer sheet as they entered the classroom. Following a lecture on the nature of stereotypes and SA, and an introduction to the training principles, Js completed the test of marital stereotypes on the separate answer sheet. Test booklets and answer sheets were then collected. The Test of the American Worker was completed directly on the test booklet. Correct answers were given and a distribution of scores was written on the blackboard and discussed by the class as a whole.

Training session 2. -- The Case of Morgan Johnson was handed to each J as he entered the classroom. The principles of training were reiterated and the Js completed the test. The separate answer sheets were collected, but the test booklets were retained. Js completed the 64 item test, 10 items at a time, directly on the test booklet. The correct answers were given immediately after each series of 10 items and responses were discussed.

Training session 3. -- Session 3 focussed on The Interests of Executives vs. Unskilled Workers and The Special Interests of Psychologists. The training materials were distributed as in sessions 1 and 2. Js completed the 60 item Executives vs. Unskilled Workers test and returned the separate scoring sheet. Js then wrote the numbers 1 through 9 on the back of the test booklet and completed the following exercise which was read to them by the trainer.

1. In the following frames one or more words is missing. You will be required to write in the missing word(s) before turning to the next (page, frame) where you find the correct (response).
2. Properly used, this manual will teach you to accurately predict the difference in the interests of executives and unskilled workers. Illustrations or examples will teach you the basic principles involved in the (differences) in interests of the two groups.
3. It will be helpful for you to assume the imaginary role of both the typical executive and the typical unskilled worker. Then, deciding that an executive would be more likely to (prefer) a musical comedy is not at all difficult.
4. As you imagine the typical individual in both groups, note that going to a musical comedy would be more agreeable to (executives) than to (unskilled workers).

5. In the same way, note that unskilled workers would be more likely to (prefer) the sport of hunting.
6. Hunting is a less intellectual and more rugged type of interest than seeing a musical comedy. These differences are typical of the differences between (unskilled workers) and executives.
7. For similar reasons, the usually better educated (executive) would prefer a sport like golf more than would the (unskilled worker).
8. Imagine which would (prefer) to be a corporation lawyer as opposed to being a criminal lawyer. The executive type would probably (prefer) to be a corporation lawyer.
9. The unskilled worker type would prefer the more adventurous job: being a (criminal) lawyer.
10. In your temporary role as an executive you should (prefer) the interest of "scientific research worker". In your other role as an (unskilled worker) you would be (less) likely to prefer such an interest.

(Smith, 1966, p. 148)

The programmed exercise introduced the Js to the technique of training to understand others via a linear program for developing SA. Following the exercise the Js completed the 60 test items, 15 at a time, directly on the test booklet. Feedback and discussion followed each series of 15 items.

Next, the Js were taught, and encouraged to use, the pooling and empathic procedures in judging psychologists' interests. They completed the test, handed in the separate answer sheet, and were given feedback followed by discussion.

Training session 4. -- The Case of the Psychology Professor contains separate exercises for each of the evaluative, stereotype, and empathic components of understanding allowing simultaneous training on all these components. Js were again presented the principles and goals of SA training. They then completed the Stereotype Training exercise and returned the separate answer

sheet before receiving feedback and discussing the case.

Training sessions 1 - 3 were consecutive, but session 4 took place 3 weeks after the third session. The criterion post-test was administered approximately 5 weeks after the first SA training session and one week after the fourth session.

Additional Variables

In addition to the SA scores, data were collected to explore the effects of: (1) J's personality; (2) his intellectual ability; and (3) his attitude towards training, on training outcome.

The personality variables are those measured by the Protebob Personality Inventory developed by Linden (1965). The test contains 200 items, with 40 items measuring each of 5 basic traits: cautious-bold, emotional-calm, introverted-extroverted, impulsive-controlled, rational-empirical. A sixth score, acquiescence, measured the J's tendency to mark "true" to a statement.

The measures of intellectual performance include:

1. Scores on a 56 item, multiple choice Midterm examination. Items 1 - 28 are a measure of the student's knowledge of the text material. Items 29 - 56 measure his comprehension of the lecture (training) concepts and principles.
2. Scores similar to those for the Midterm are also available for the 86 item, multiple choice Final examination. Forty three items tested the text material since the midterm, and forty three items tested knowledge of the lecture (training) material.

3. Grade Point Average: based on the trainee's overall academic record to the beginning of the term of training, i.e., his GPA as of the end of Summer term 1967.

4. The College Qualification Test (Form B, 1956) is a measure of general academic aptitude. Test scores are reported in the form of percentile ranks which specify a student's position on the test relative to all new students in his entering class. The present study used scores for the verbal (V) and information (I) subtests, and the total score (T). "T" included a subtest measuring numerical proficiency. "V" is a measure of (recognition) vocabulary and reflects verbal abilities associated with success in social science, literature, and other similar fields. "I" measures general information in the social and natural science.

An additional, demographic, variable was the sex of each J.

Finally, three variables apply directly or indirectly to an assessment of J's attitudes toward the training experience.

1. Confidence in improvement. When the posttest criterion for the overall training program, of which this study was a part, was administered (Test of the Ability to Understand People, Part II), Js estimated the number of items (out of 60) they expected to answer correctly. They were instructed to evaluate the impact they felt the entire training program had had on their ability to understand others.

2. Liking for training. In the same session in which the Js made their estimates, they were asked to rank, in terms of benefit and interest, their 5 most recent courses. The 5 courses were to be ranked from 1 to 5 (with 1 being highest and 5 lowest).

3. The number of training sessions attended.

The type of sensitivity instruments used in the present study were used in previous research (Kepes, 1965; Grossman, 1967). Control groups that did not receive training did not improve. In as much as this finding was consistent over a number of studies, it seems to be a well established phenomenon that no improvement can be expected without training. Hence, the control groups in the previous literature may serve as the controls for the present research.

Method of Analysis

200 subjects completed the training program. 54.5% of the sample were females; 45.5% males. 100 Js, 50 males and 50 females, were randomly assigned to a validation group (V). Analysis of the V data aimed at generating hypotheses to be tested on the remaining 100 subjects, of whom 39 were males and 61, females. The latter sample provided a cross-validation (C-V).

Since class attendance was not mandatory, not all subjects completed each and every phase of the overall training program. Table 2 shows how many trainees took part in the SA program.

The V group was consistently superior to the C-V group on measures of intellectual ability. The groups did not differ on

measures of attitude toward training. Personality trait patterns are similar for the total group and for females, but C-V males were more extroverted than V group males.

Table 2. The number of subjects who completed stereotype accuracy training.

A. Validation Sample	Total	Males	Females
Session 1	90	48	42
Session 2	89	47	42
Session 3	89	46	43
Session 4	82	44	38
Criterion Posttest	81	42	39
Criterion Pre and Posttest	75	40	35
B. Cross-validation Sample	**Total**	**Males**	**Females**
Session 1	84	34	50
Session 2	84	35	49
Session 3	83	34	49
Session 4	76	32	44
Criterion Posttest	72	31	41
Criterion Pre and Posttest	62	28	34

The basic measure of SA improvement due to training was the difference between pretest and posttest criterion scores. Such scores were available for 75 Js in the V sample and 62 in the C-V group. Tests of differences between means utilized t-tests, and matched t-tests where appropriate.

Reliabilities (internal consistency) were computed for each SA test. Criterion scores were correlated to each other and to improvement, and analyzed according to level of item difficulty.

A secondary analysis concerned itself with the differential

impact of training. These analyses involved a general correlational study using the Missing Data Routine (MDSTAT) on Michigan State University's CDC 3600 computer.

Results

Table 3 presents Kuder-Richardson #20 reliabilities for each of the SA measures. The coefficients are based on the total sample which varied, according to class attendance, for each predictor.

Table 3. Internal consistency of the stereotype accuracy predictors.

Predictor	N	Reliability
Marriage Test (Criterion)	188	.68
The Case of Morgan Johnson	184	.66
The Interests of Executives vs. Unskilled Workers	181	.66
The Special Interests of Psychologists	180	.77
The Case of the Psychology Professor	160	.68

Below are results in light of the respective hypotheses. Wherever possible, the large volume of data has been summarized and the original data source placed in an appendix. A brief discussion section appears after the presentation to summarize the findings.

Hypothesis 1: A training program will be effective in improving stereotype accuracy.

Comparison of pretest and posttest scores for the total V and C-V samples indicates agreement on the effectiveness of training in improving SA. Evidence regarding sex differences in

improvement was less clear. Table 4 summarizes the findings.

Table 4. The results of training.

Subjects	N	Criterion Variance Pre	Post	Criterion Mean Pre	Post	Diff.	t
Validation Sample							
Total	75	35.97	47.86	34.95	37.39	+2.44	2.32*
Males	40	31.12	47.61	34.90	36.33	+1.43	1.02
Females	35	42.59	46.72	35.00	38.60	+3.60	2.26*
Cross-validation Sample							
Total	62	33.91	32.69	35.11	37.66	+2.55	2.48*
Males	28	28.84	27.73	33.89	37.21	+3.32	2.34*
Females	34	36.77	37.42	36.12	38.03	+1.91	1.31

*p < .05

While gain for males in the V group and females in the C-V group did not reach statistical significance, there was a trend shown toward improvement which was consistent for all groups.

Hypothesis 2: The effectiveness of training varies inversely with initial accuracy.

Table 5 shows the relation between initial accuracy and improvement. In all comparisons, pretest scores were negatively

Table 5. The relationship between initial accuracy and amount of improvement.

	Validation Sample Gain	Cross-validation Sample Gain
A. Total Pretest	-.48**	-.54*
B. Men Pretest	-.45**	-.68**
C. Women Pretest	-.52**	-.40*

**p < .01
* p < .05

related to improvement scores supporting the hypothesis. The lower a judge's accuracy score before training, the more improvement he is likely to show, and vice versa. The results are similar for separate male-female analyses.

Hypothesis 3: The effectiveness of training will be reflected more clearly on stereotype accuracy predictor items of moderate difficulty, than on those of either high or low difficulty.

The results of training were analyzed in light of a J's degree of initial accuracy (low, moderate or high) and the level of criterion pretest item difficulty[1] (easy, moderate, or hard).

Js whose criterion pretest raw scores ranged from 0 to 31 (N = 38) were classified as having low pretraining accuracy. Pretest scores of initially moderately accurate subjects ranged from 32 to 37 (N = 58), while pretest scores of 38 or higher (N = 41) placed J in the group considered to have started with high SA. Similarly, indices of item difficulty allowed categorization of criterion items as easy, those missed by 0-29% of the Js (N = 19); moderate, those missed by 30-46% of the Js (N = 19); or hard, those missed by 47+% of the Js (N = 18). Table 6 presents a summary of the analysis. The original data appear in a more complete form in Appendix B (p. 104).

Criterion pretest scores on the easy items declined after training. Scores increased on the moderately difficult items, and on the hardest items as well. The trend was consistent in that all three groups of Js showed greatest improvement on the

[1] The index of difficulty is the percentage of the total group marking a wrong answer or omitting the item.

moderate items and between easy and hard items were statistically significant ($p < .01$), whereas the moderate-hard item difference was not. It would seem that more SA improvement can be expected on difficult items than on easy items. It may be that most "difficult" items did not tap the extremes of difficulty, as well as the easy items tapped the extremes of ease. The items were initially grouped so that item categories contained approximately equal numbers of items. The data lend qualified support for the hypothesis.

Table 6. Summary of the relation of item difficulty to the measurement of improvement due to training.

	Criterion Improvement Scores			
	Item Difficulty			
Initial Accuracy	Easy	Moderate	Hard	Mean Total
Low	.5	3.2	2.4	2.0
Moderate	- .3	1.7	1.2	.9
High	-1.2	0	- .1	- .4
Mean Total	- .3	1.6	1.2	

Note:--These data are mean criterion improvement scores. Tests of the significance of the difference between differences for the Total marginals yielded the following results:

Easy item score vs. Moderate item score: $t = 5.14$, $df = 272$, $p < .01$;
Easy item score vs. Hard item score: $t = 4.55$, $df = 272$, $p < .01$;
Moderate item score vs. Hard item score: $t = 1.00$, $df = 272$.

Hypothesis 4: Stereotype accuracy is a situation specific ability.

If SA is situation specific, accuracy in judging one group should be independent of accuracy in judging another. SA scores for 5 situations were intercorrelated to test the hypothesis. The data appear in Table 7.

Table 7 shows little consistency in the comparison of V and C-V findings. The presence of relationships statistically

different from zero suggested generality, but were not reproduced in cross-validation. The failure to replicate, coupled with the generally small correlation coefficients, offered conflicting evidence with regard to the situational specificity of the ability. Of particular interest was the pattern of relationships observed: statistically significant values occured mainly between scores on instruments requiring predictions about similar groups. Thus, observed generality may be due to the similarity of the tasks involved rather than the generality of the ability. The number of pairs of observations ranged from 67 to 89. Results for men and women were similar to those for the total group.

Table 7. The generality-specificity of stereotype accuracy.

A. Validation Sample - Total					
Stereotype	Marital Pretest	Morgan Johnson	Executive	Psychologist	Professor
Marital Pretest	-	.23*	.21*	.12	.21
Morgan Johnson		-	.13	.31**	.24**
Executive Worker			-	.07	-.05
Psychologist				-	.41**
Professor					-
B. Cross-validation Sample - Total					
Marital Pretest	-	.20	.03	.34**	.10
Morgan Johnson		-	.03	.32**	.18
Executive-Worker			-	.30*	.02
Psychologist				-	.08
Professor					-

**p < .01
* p < .05

Hypothesis 5: The effectiveness of training varies with personality, intellectual ability, and attitudes toward training.

Analyses relating personality trait scores with SA criterion posttest and improvement scores found no relation between any of the 5 personality traits measured and improvement due to training. (See Appendix E , p. 105).

Correlation coefficients were computed between training outcome measures for the total groups and measures of intellectual ability (Appendix b, p.106). Only the midterm examination grade covering the text material was related to criterion posttest scores in both V and C-V samples. Other correlations statistically different from zero (but not replicated) occured in the V sample when comparing criterion posttest scores with GPA and CQT verbal, information, and total test scores. Improvement was statistically related to GPA and CQT information scores without replication. The results with regard to the relation of intelligence and training outcome are inconclusive.

Analyses performed separately on males and females yielded a clearer result. Table 8 summarizes the results of the male analysis. Gain for men was positively related to intellectual ability as measured by GPA: (correlated with both criterion posttest and improvement scores in both V and C-V); CQT-Verbal score (criterion posttest); CQT-Information score (improvement); CQT-Total (criterion posttest). Both measures of training outcome were statistically related to lecture and total final

Table 8. Summary of the relationship of intellectual ability to training outcome for men.

	Post-training Measures	
	Criterion Posttest	Improvement
A. Validation Sample - Males		
Final exam grade: lecture questions	**	**
Final exam grade: total	*	*
Grade Point Average	*	**
CQT-Verbal score	*	
CQT-Information score		**
CQT-Total score	**	**
B. Cross-validation Sample - Males		
Final exam grade: text questions		*
Grade Point Average	*	**
CQT-Verbal score	*	
CQT-Information score	*	*
CQT-Total score	**	

**p < .01
* p < .05

examination scores in the V sample but not in C-V. Inspection of the data for women showed no relationship between training outcome and any of the intelligence measures.

Finally, neither confidence in improvement, liking for training, or number of class sessions attended was related to training outcome. Data for these analyses appear in Appendix B (p. 108).

Summary of the Results of Study One

Study One was largely exploratory, designed to generate hypotheses for further testing, and to serve as the basis for a revised training program. The results of the analyses are summarized below and will be more fully discussed in a later chapter.

Participation in a training program improved SA. There existed, however, an inverse relation between initial accuracy and training outcome: the lower a judge's pretraining accuracy, the more likely he was to benefit from training, and vice versa.

SA predictors had good internal consistency, but moderately difficult and hard criterion items best reflected improvement due to training. Subjects showed either no change or a decline in accuracy on the easy items.

The findings about the generality-specificity of the ability were ambiguous. Overall, however, SA tended to be situation specific: judges who were accurate in their predictions of one group, were not necessarily accurate in predictions of other groups. The intercorrelations observed among the predictor pretest scores may have reflected the similarity of tasks involved (i.e., the similarity of the groups to be judged) rather than any underlying continuum of generality.

SA improvement scores were unrelated to J's personality, sex, or attitude toward training. For men, gain was positively related to some intellectual factors. For women, improvement was unrelated to intellectual ability.

Several limitations of Study One should be noted. First, the frequent failures to reliably replicate results in cross-validation may have been due to wide fluctuations in sample sizes. C-V results were difficult to interpret in light of generally shrinking N's from the V to the C-V analyses. A

second, related problem, concerned the degree of sample shrinkage due to the relative inefficiency (e.g., length, lack of posttest) of the SA measures.

Study Two was an extension of Study One. Training measures and procedures were modified in light of experience with the pilot program. The next chapter describes the methods and results of Study Two.

CHAPTER III

STUDY TWO

Problem

Study Two was an extension of Study One. The pilot research was mainly exploratory with the aim of suggesting revised training procedures and hypotheses. The second study tested the revised hypotheses and applied the findings of Study One to the development of a revised training program. The following hypotheses guided the present research.

Hypothesis 1: Participation in a training program will improve stereotype accuracy. Improvement will be stable over time.

Hypothesis 2: The effectiveness of training varies inversely with initial accuracy.

Hypothesis 3: Stereotype accuracy is situation specific, and improvement is similarly specific.

Hypothesis 4: Specific training with stereotypes of female Os will generalize to similar stereotypes of male Os.

Hypothesis 5: Training will result in greater improvement in accuracy when written judgments are not made before training, than when written judgments are made before training.

Hypothesis 6: The effectiveness of training is unrelated to personality and attitude towards training.

Hypothesis 7: Improvement in stereotype accuracy is positively related to intellectual ability for men, but unrelated for women.

General Design

Study Two used a pretest-posttest design with interpolated training for each stereotype. Subjects were administered a pretest, received training in making predictions about groups or particular group members, and were posttested with the pretest

criterion measure. Each session formed a complete pretest-training-posttest sequence. Training took place in 3 successive sessions of about one hour and ten minutes each.

Subjects

The trainees were 182 MSU students, representing a wide range of college majors and levels, who were enrolled in a Winter 1968 section of a Psychology of Personality course. Training, as in the pilot study, was a part of the course work and took place during the regularly scheduled class meetings. Attendance was not mandatory.

Measures of Stereotype Accuracy

The SA predictors used in the present study were greatly shortened from the original forms used in Study One. The revisions were more in keeping with the relative specificity of the ability and, by making possible a pretest-posttest design for each training session, avoided the wide variation in sample sizes. Item analyses of the Study One measures identified a core of items covering the whole range of item difficulty levels for each instrument. Table 9 presents the original and revised item difficulty data. While items were chosen to represent the continuum from "easy" to "hard", the mean item difficulties (in terms of the average percentage of trainees answering an item correctly) was relatively similar for the original and shortened test forms.

The revised paper-and-pencil tests of SA appear in Appendix

Table 9. Summary of item analyses for original and revised stereotype accuracy predictors (based on the Fall, 1967 data).

Predictor	Mean Percentage Correct			
	Study One	Items	Study Two	Items
The Case of the Psychology Professor (N = 160)	76.9	40	74.4	16
The Special Interests of Psychologists (N = 180)	65.4	60	58.5	16
The Case of Morgan Johnson (N = 184)	57.8	64	53.2	20
Marriage Test (N = 188)	61.6	56	61.2	28
Exercise A. Men	58.6	28	56.5	16
Exercise B. Women	64.4	28	66.0	12

C. An IBM scoring sheet, provided for recording judgments, allowed machine scoring of all instruments.

The following revisions were made.

1. The 2-part Case of the Psychology Professor became two separate exercises in the revised form. The first exercise, The Typical College Man (CM-I), asked J to predict the group norm of typical college men in responding to a personality inventory. Both a pretest and a posttest were administered on 16 items which required J to:

> Mark "1" on the separate answer sheet if you think the typical college man answered "true" to the statement;
> Mark "2" if you think the typical college man answered "false".

16 training items, ordered from the least to the most difficult, were provided for practice and feedback to be interpolated between the pretest and posttest.

2. The second exercise called for the Application of the Stereotype to a Particular Man (CM-II) on the same 16 items

used in CM-I. The same training items are also used between the pretest and posttest as were administered in CM-I. J was informed that most people conform to a stereotype in some respects and differ in other respects. J's task was to consider how a particular professor conformed to and was different from the group norm of college men in general. For each item J was to:

> Mark "1" on the separate answer sheet if you think the professor answered "true" to that particular statement;
> Mark "2" if you think he answered "false".

In this manner, J demonstrated his ability to predict group norms and to apply his ability to predictions about a unique individual.

3. The Typical Psychologist (P-I) exercise was the shortened and revised form of The Special Interests of Psychologists test. The instructions were similar to those for the original test form, except that J took both a pretest and a posttest on 16 items designed to differentiate between psychologists and men-in-general. 16 interpolated training items ordered from least to most difficult gave opportunity for practice.

4. The revised form of the Morgan Johnson test was an Application of the Stereotype to a Particular Psychologist (P-II). Instructions on the revised form were similar to those for the original test. While the P-I measure described above required the trainee to differentiate between psychologists as a group and the group of men-in-general, P-II required J to apply the stereotype to a particular psychologist both as a group member and as a unique

individual. 20 items were provided for use as a pretest and posttest, and 16 additional items ordered by difficulty level, served as training items.

5. The Marital Stereotypes exercise was the revised (28 items) form of the pilot criterion instrument. The instructions for the revised measure were similar to those for the original predictor. The shortened version, however, allowed a pretest and a posttest, in a single session, on 16 items aimed at differentiation between happily married, unhappily married, and divorced men (MM), and 12 items aimed at similarly differentiating between women (MW) in these marital groups. 12 training items were included on the test form to provide practice in, and feedback about, applying the stereotypes to the women marital groups. The training items were arranged so that J proceeded from the easiest to the most difficult items.

Training Procedures

Six training procedures or conditions were used in Study Two:

1. a program involving diagnosis of error tendencies, practice in making judgments, feedback, and discussion;
2. a pooling technique;
3. a panel technique;
4. a trait method;
5. a noncommitment training condition; and
6. a base-rate condition.

It will be useful to view the techniques as stressing allied, but different aspects of training. First, diagnosis, practice, feedback and discussion stressed the basic training principles: Stereotypes are not, in themselves, bad and may actually aid in prediction if accurate and flexible. In the present study, diagnosis was less formalized than in the pilot program. Diagnosis involved demonstration to J, on the training items, that inaccurate stereotypes result in poor prediction of group characteristics, and that inflexible stereotypes result in inaccurate predictions about individual group members. In addition, at the start of each session, Js were instructed to recall their performance on the previous session's exercises as a way of assessing their progress in SA improvement.

Practice, feedback, and discussion were further systematized. Training items were completed in series of, for example, 4 or 5 at-a-time followed by feedback and discussion, completion of the next 4 or 5 items, and so on until all training items were answered.

Second, the methods described below emphasized the process of acquiring information about groups and the psychological organization of the information to make for more accurate judgments.

The pooling technique described in the Training Procedures section of the Study One chapter, stressed the training principle: Use your knowledge of similar people (Smith, 1968a).

The panel method applied the principle: Listen to what others

have to say before making your predictions. A panel of 4 to 6 student volunteers sat in front of the classroom. The panel members completed a series of training items and one member read and explained his predictions. Panel members then discussed their predictions, particularly disagreements. The class members as a whole then made their judgments on the series of training items and discussed the responses. The discussions were led, at all times, by the panel. The role of the trainer was the purely nondirective one of reflecting the proceedings. The correct answers were read and the panel proceeded to the next series of training items.

The trait method encouraged the J to: Shift to an explicit theory of personality. Js were taught and encouraged to use a trait theory involving the dimensions: emotional-calm, introverted-extroverted, impulsive-controlled, rational-empirical, and cautious-bold. As a part of training, each J took a personality test and received a graphic profile of his own trait scores. Further, Js were told the relative position of the group being judged on the continuum of, for example, introversion to extroversion. The J then used the information and the theory to describe the group being judged and to compare the judged group with other groups.

Under the noncommitment condition, Js were instructed to: Delay the formation of impressions. Js whose last names began with the letters "A" through "G" were given the same training materials as the "H" to "Z" Js, but did not answer the pretest

items. "H" to "Z" did answer the items. All Js participated fully, however, in the training. The noncommitment condition was used only with the P-I and P-II stereotypes.

Lastly, the base-rate condition, used only with P-II, told J the percent of men-in-general who expressed a liking for a particular interest vs. the percent of the members of a group (to which the O belongs) who expressed a liking for the particular interest. Again, the Js were instructed in the need for flexibility in applying a stereotype to an individual group member. While psychologists in general say they like a particular interest, any single psychologist might dislike it.

Contents of the Training Sessions

One or more of the training procedures were combined with each of the SA measures to comprise a single training session. Table 10 summarizes the content of the 3 meetings devoted to SA.

Table 10. Summary of training activities in Study Two.

		Training Procedure*							
Session	Stereotype	Pretest	(1)	(2)	(3)	(4)	(5)	(6)	Posttest
5	College Man I	x	x	x	x	x			x
5	College Man II	x	x	x	x	x			x
6	Psychologist I	x	x	x	x	x	x		x
6	Psychologist II	x	x	x	x	x	x	x	x
7	Marital-Men	x							x
7	Marital-Women	x	x	x	x	x			x
20	CM-II,P-I,M-M,M-W	Readministered							

*Note: (1) = diagnosis, practice, feedback, discussion; (2) = pooling; (3) = panel; (4) = trait; (5) = noncommitment condition; (6) = base rate condition.

As in Study One, the revised program was a part of an overall program designed to improve the understanding of others. SA

training occurred in the 5th, 6th, and 7th sessions of the broader program. In the 20th session, a random sample of 68 Js again completed the CM-II, P-I, M-M, and M-W tests.

Training session 1. -- Js took test booklets (CM-I and CM-II) when they entered the room. Following a lecture on the nature of stereotypes and SA and an introduction to the basic training principles and methods, they completed the 16 item CM-I pretest on a separate answer sheet and were introduced to the pooling and trait methods. The norms for college men and women on the personality trait dimensions were graphically depicted. (The norms for the personality inventory appear in Smith, 1968b.) The volunteer panel came to the front of the classroom and the training items were completed in series of 4 items each. Following training, the Js completed the 16 item CM-I posttest on the separate answer sheets. CM-II was now taken up with a procedure similar to that for CM-I.

Training session 2. -- Js took a P-I and P-II test booklet. Students with last names beginning with the letters "A" through "G" sat on the left side of the classroom and did not take the P-I pretest (noncommitment condition). The remaining Js completed the P-I pretest. The class members as a whole were given instructions for pooling and the empiricism of psychologists was emphasized. Martin Luther King and David Hume were contrasted (Smith, 1968b) as representing opposite ends of the rational-empirical continuum. The panel came to the front of the classroom and answered the training items 8 at-a-time.

All Js completed the training items and the P-I posttest. Procedures for P-II were similar to those for P-I.

Training session 3. -- Training materials were distributed as Js entered the classroom. The principles of SA and of training were reviewed. Pretests of Marital Stereotypes of men and women were completed on the separate answer sheets and the panel assembled to begin training only on the items pertaining to stereotypes of women. Js used the pooling technique again. The relative positions of the marital groups on the emotional-calm and introversion-extroversion trait dimensions were graphically presented on the blackboard. The training items for women were completed in 2 series of 6 items each. After training the subjects completed both the MM and MW posttests.

Additional Variables

In addition to the SA measures, data were collected to explore the relationship of: (1) J's personality; (2) his intellectual ability; (3) his attitude towards training; and (4) his sex, age, and scores on a human relations inventory, to training outcome.

The personality variables were those measured by the Protebob Personality Inventory described in Study One.

As in Study One, measures of intellectual performance included scores on multiple choice midterm (56 items) and final (88 items) examinations. Half of the items on each test were devoted to the text material and half to the lectures (training).

A total text and lecture score was available as the sum of the appropriate midterm and final exam scores. A total score comprised the sum of the whole midterm and final exam scores. Similarly, Js' CQT Verbal, Information, and Total scores were recorded, as were cumulative grade point averages (GPA) up to, but not including the Winter 1968 quarter.

Attitudinal factors were measured by: (1) J's ranking of the Psychology 225 course as described in Study One, and (2) a count of the number of SA training sessions attended.

The demographic variables included the Js' sex and age.

Finally, Js filled out The CR Leadership Questionnaire (Dore, 1960). It contains 60, 2-alternative, forced-choice items designed to measure attitudes toward 2 leadership methods: consideration (C) and responsibility (R). The inventory yielded 3 scores: C, R, and a total score. Items 1-30 required a choice between a less and a more considerate statement. Persons choosing a greater number of considerate items tend to think it more important to be employee-oriented than work-oriented; subjects who mark more inconsiderate items, work-oriented rather than employee-oriented. Items 31-60 required a choice between items that are high or low in reponsibility. Persons scoring high on R generally feel that a superior should play a role different from that of his subordinates, spending more time in activities his followers are not able to do: e.g., organizing, planning, providing information, etc. (Smith, 1968b).

Method of Analysis

182 subjects took part in some phase of training. Since attendance was not mandatory, the number of Js varied from instrument to instrument. Table 11 shows the number of Js who attended each training session. In all, 114 Js attended all 3 training sessions. 41 were men; 73, women.

Table 11. The number of subjects who completed the revised stereotype accuracy training program.

		Subjects		
Session	Stereotype	Total	Males	Females
1	College Man I	157	56	101
	College Man II	157	56	101
2	Psychologist I	107	41	66
	noncommitment	51	20	31
	Psychologist II	101	37	64
	noncommitment	55	23	32
3	Marital-Men	138	52	86
	Marital-Women	138	52	86
Retest	CM-II,P-I,M-M,M-W	68	24	44

Kuder-Richardson #20 reliabilities were computed for the shortened criterion instruments.

The basic measures of improvement were differences between pretest and posttest criterion scores available for each stereotype. Tests of differences between means used t-tests and matched t-tests where appropriate.

Correlations were computed on Michigan State University's CDC 3600 computer using a prepared routine (MDSTAT). Criterion scores were intercorrelated and related to the additional variables.

Results

Table 12 presents the reliabilities of the revised SA measures. The revised reliabilities are lower than those for the pilot instruments. The tests used in the second study were greatly shortened from the originals, and contained items chosen to represent the entire available range of content and difficulty and discrimination levels.

Table 12. Reliabilities of the shortened stereotype accuracy predictors.

	N	Reliability KR#20
The Typical College Man (College Man I)	157	.33
The Application of the Stereotype to a Particular Man (College Man II)	157	.14
The Typical Psychologist (Psychologist I)	107	.54
Application of the Stereotype to a Particular Psychologist (Psychologist II)	101	.15
Marital Stereotypes - Men Exercise	138	.44
Marital Stereotypes - Women Exercise	138	.26

As with the pilot program, the data for Study Two are presented in relation to the relevant hypotheses formulated to guide the study. Similarly, wherever possible, the data have been summarized and the original data placed in Appendix D (p. 120ff).

Hypothesis 1: Participation in a training program will improve stereotype accuracy. Improvement will be stable over time.

Table 13 summarizes the results of training for the trainees as a group. Separate analyses of male and female data yielded similar results (see Appendix D, p. 121). Training did not affect scores on the CM-I or CM-II measures. Improvement due to training was statistically significant, however, for the stereotypes of psychologists and of happily married, unhappily married, and divorced

men and women. Additional analyses (Appendix D, p. 122), yielded no reliable differences between men and women in either accuracy before training or the impact of training. Moreover, when Js were retested on 4 stereotypes (CM-II, P-I, M-M, M-W), 7 weeks after training, there was no decline in accuracy (Appendix D, p. 137). Training impact was stable. The results endured over time.

Table 13. The results of the revised training program for the total sample.

Stereotype	N	Pretest Mean	Posttest Mean	Difference	t
College Man I	157	13.69	13.45	- .24	1.09
College Man II	157	10.19	10.23	.04	.20
Psychologist I	107	10.20	11.42	1.22	3.49**
Psychologist II	101	10.88	11.99	1.11	3.70**
Marital-Men	138	9.50	10.84	1.34	5.36**
Marital-Women	138	7.45	9.58	2.13	3.74**

Hypothesis 2: The effectiveness of training varies inversely with initial accuracy.

Table 14 shows the relation between pretraining accuracy and improvement.

Table 14. The relationship between initial accuracy and amount of improvement for each revised criterion.

	Subjects		
Stereotype	Total Gain	Men Gain	Women Gain
College Man I Pretest	-.31**	-.40**	-.27**
College Man II Pretest	-.40**	-.29*	-.45**
Psychologist I Pretest	-.59**	-.60**	-.58**
Psychologist II Pretest	-.53**	-.49**	-.57**
Marital-Men Pretest	-.61**	-.70**	-.56**
Marital-Women Pretest	-.17	-.59**	-.13

**$p < .01$
* $p < .05$

It was found that initial accuracy was negatively related to gains in all comparisons. The results offered support for the hypothesis.

Hypothesis 3: Stereotype accuracy is situation specific, and improvement is similarly specific.

The interrelation of pretraining, posttraining, and improvement scores between all six criteria (Appendix D, p. 123) yielded few statistically reliable relationships. Out of 135 correlation coefficients only 7 were statistically significant at the .05 level. Seven statistically significant relationships out of such a large number could certainly be expected by chance alone.

These data support the hypothesis that SA and SA improvement are specific. The average correlation for 135 comparisons for the total sample was .09. The average intercorrelations were .14 and .10 for the male and female data respectively (Appendix D, p. 123ff).

Hypothesis 4: Specific training with stereotypes of female Os will generalize to similar stereotypes of male Os.

The data in Table 13 (page 59) are relevant to the present hypothesis. Js were pretested on MM and MW stereotypes, trained on MW stereotypes only, and posttested on both MM and MW stereotypes. As shown in Table 13 gains on both MM and MW stereotypes were equally statistically significant. Had there not been generalization, these results could not have occurred. Additional support for the hypothesis was derived from separate

analyses of the male and female data (Appendix D, p. 122) which obtained similar results.

Hypothesis 5: Training will result in greater improvement in accuracy when written judgments are not made before training, than when written judgments are made before training.

For P-I and P-II only, an experimental group of Js did not formally commit themselves to judgments before training, but took a written posttest. Control Js wrote their P-I and P-II predictions on both a pretest and a posttest. Table 13 (page 59) showed that the control group improved significantly on the P-I and P-II measures. DIF accuracy on P-I and P-II was not, however, reliably different from the posttest accuracy scores of the total, male and female control groups. Table 15 summarizes the results of delayed impression formation (DIF). Additional data appear in Appendix D (p. 126).

Table 15. Summary of the results of delayed impression training.

Stereotype	Comparisons of Means	Diff.	df	t
Psychologist I	Total group posttest vs. DIF	.07	156	.17
	Male posttest vs. DIF	-.49	59	-.77
	Female Posttest vs. DIF	.42	95	.78
Psychologist II	Total group posttest vs. DIF	.06	154	.15
	Male posttest vs. DIF	.68	62	1.10
	Female posttest vs. DIF	-.30	94	-.60

**p < .01
* p < .05

The data in Table 15 lend no support for Hypothesis 5.

Additional information about the relationships of DIF accuracy and the additional experimental variables are contained in Appendix D (p. 126f).

Hypothesis 6: The effectiveness of training is unrelated to personality and attitude toward training.

The present analyses separately correlated SA posttest and gain scores with measures of personality, age, and training and leadership attitudes (Appendix D, pp. 131 - 136).

Analysis of the combined male-female data revealed that more emotional Js show greater gains on the P-I test (-.24, $p < .05$), while the calmer Js score higher on the P-II posttest (.21, $p < .05$). The more cautious Js do better on the MM posttest (-.18, $p < .05$), but impulsive Js show more improvement (-.19, $p < .05$) due to training.

Among men, emotionality was related to gain on P-I (-.39, $p < .05$); and caution with improvement on P-II (-.43, $p < .01$). MW posttest scores were higher for controlled and empirical the impulsive (.38, $p < .01$) and rational (.31, $p < .05$).

Among women, introversion was related to posttest scores on CM-II (-.22, $p < .05$), while calmness was related to P-II posttest scores (.33, $p < .01$). Bolder women scored lower on the MM posttest than the more cautious (-.22, $p < .05$). Gain on the MM test was greatest for impulsive (-.27, $p < .05$), cautious (-.35, $p < .01$) women as compared to controlled, bold women.

Furthermore, male and female correlation coefficients differed statistically for the following comparisons:

-P-II posttest vs. cautious-bold scores: for men, -.43; for women, .11. ($Z = 2.59$, $p < .05$);
-MM improvement vs. cautious-bold scores: for men, .20; for women, -.35. ($Z = 3.35$, $p < .01$);

-MW posttest vs. impulsive-controlled scores: for men, .38; for women, -.09. ($Z = 2.76$, $p < .01$).

With regard to the age and attitude variables for the total sample, age was negatively related to gain (-.20, $p < .05$) on the MW test. Liking for training (course rank) was unrelated to training outcome. Attendance at SA sessions was reliably related to P-II improvement (.23, $p < .05$). Finally, Responsibility (R) leadership scores were negatively correlated with gain on CM-I (-.28, $p < .01$) and MM (-.25, $p < .01$) tests.

Both MM posttest (-.34) and gain (-.35) scores correlated negatively ($p < .05$) with the ages of male Js. MW posttest (-.36) and gain (-.36) scores correlated negatively ($p < .05$) with R scores. Attendance was positively related to CM-II improvement (.27, $p < .05$), but inversely related to P-I gain (-.45, $p < .01$).

For women, CM-I posttest (-.25) and gain (-.29) scores were negatively related ($p < .05$) to R scores, and CM-II posttest accuracy correlated positively with the total leadership inventory scores (.29, $p < .05$). P-I improvement was related to R (.28, $p < .05$), and MW improvement (.26, $p < .05$) to attendance.

Personality, age, and attitude, it appears, may be related to gain, but in such a specific manner that the initial hypothesis may not be relevant.

Hypothesis 7: Improvement in stereotype accuracy is positively related to intellectual ability for men, but unrelated for women.

Table 16 summarizes the relation between training outcome and measures of intellectual performance. Total group CM-II

Table 16. Summary of the relation of training outcome and intellectual ability.

	TOTAL				
	College Man II Posttest	Psychologist I Gain	Psychologist II Gain	Marital Women Posttest	Marital Women Gain
Midterm: text	**			*	**
Midterm: lecture	**			*	*
Midterm: total	**			**	**
Final: text				**	
Final: lecture				*	*
Final: total				**	*
Text: total	*			**	*
Lecture: total	*			**	*
Total: total	*	*		**	**
GPA	*	*			
CQT-Verbal	*				
CQT-Information					
CQT-Total	*		*	*	

	MEN				
	College Man II Posttest	Psychologist I Posttest	Psychologist I Gain	Marital Women Posttest	Marital Women Gain
Midterm: text				*	**
Midterm: lecture		*		**	*
Midterm: total				*	**
Final: text				*	*
Final: lecture	*			*	**
Final: total	*			*	**
Text: total				*	**
Lecture: total	*			**	**
Total: total				**	**
GPA			*		
CQT-Verbal				*	
CQT-Information				**	**
CQT-Total				**	**

	WOMEN			
	College Man II Posttest	Psychologist I Gain	Marital Men Gain	Marital Women Posttest
Midterm: text	**			
Midterm: lecture				
Midterm: total	**			
Final: text				**
Final: lecture		*		
Final: total		*		*
Text: total				*
Lecture: total		*		
Total: total				*
GPA			*	
CQT-Verbal	*			

**$p < .01$ *$p < .05$

posttest scores correlated positively with the 3 midterm exam scores, the 3 composite exam measures, GPA, and CQT-V and Total scores. P-I gain was inversely related to Total-total and GPA. P-II gain was inversely related to CQT-Total. MW posttest scores were related to all intellectual measures except GPA and CQT-V and I scores. MW gain was related to all intellectual measures except final text subscores, GPA, and the CQT measures.

For men, CM-II posttrained scores correlated positively with final exam lecture subscore and total grade, and with the composite lecture scores. P-I posttest was related to midterm lecture subscores, while improvement on the test was negatively related to GPA. MW posttest scores were correlated with all measures except GPA; gain scores were related to all intellectual measures except GPA and CQT-V.

For women, CM-II posttest scores correlated positively with midterm text and total grades, and with CQT-V scores. P-I gain was inversely associated with final exam lecture and total scores and with composite lecture subscores. MM gain was similarly negatively related to GPA. MW posttest scores varied positively with final exam text and total grades as well as with the composite measures of text and total exam performance.

Furthermore, tests of the significance of the difference between correlations (male vs. female Js) of MW posttest scores and CQT-V ($Z = 2.12$), CQT-I ($Z = 2.24$), and CQT-T ($Z = 2.53$) scores were statistically significant at the .05 level. Similar tests

compared male and female correlations of MW gain scores with final exam test ($Z = 2.24$), lecture ($Z = 2.24$), and total grades ($Z = 2.06$), and composite text ($Z = 2.24$), lecture ($Z = 2.18$), and total grades ($Z = 2.06$). Each of the comparisons differed at the .05 level. MW-CQT-I ($Z = 2.76$) and MW-CQT-T ($Z = 2.88$) correlations for males and females differed at the .01 level. In all of these analyses, the correlations between intellectual factors and improvement were larger for males than for females.

Summary of the Results of Study Two.

The revised instruments of Study Two were greatly shortened from those in the pilot project. Fewer items allowed a pretest-training-posttest design in each training session. The reliabilities of the new measures declined, however, (except for P-I which increased) even when corrected for length by the Spearman-Brown formula.

Participation in a training program did improve predictive accuracy with regard to the stereotypes of psychologists and married and divorced persons. Training did not affect accuracy in making predictions about college men. Male and female Js did not differ in either pretraining accuracy or in the benefits derived from training. Moreover, improvement due to training was stable.

The lower a J's pretraining accuracy, the greater the benefit from training, and vice versa. This inverse relationship between initial accuracy and gain may account for the failure of training to improve the already high CM accuracy scores.

The overall results point to the relative specificity of the stereotype accuracy ability and its improvement. This finding may make the original hypotheses no longer relevant, or at best, difficult to interpret. A J's accuracy in predicting about one group does not necessarily mean that he will be accurate in his predictions about other groups. Inaccurate prediction about one group, on the other hand, did not preclude accuracy in judging other groups. Similarly, improvement in prediction about one group did not necessarily improve prediction about other groups, nor did failure to improve in accuracy about one group mean that a J will not benefit from training on some other stereotype.

Training on stereotypes of women generalized to stereotypes of men on the Marital Stereotypes exercise.

There was no difference in improvement between judges who make written predictions before training vs. those who delay the formation of impressions. An additional analysis explored the correlates of the delayed impression formation (DIF) accuracy scores.

Turning to the correlates of posttraining accuracy and gain, specificity is once again an important factor. Total group P-II and MM posttest scores were related to calmness and caution, respectively. Gain on P-I and MM were associated with emotionality and impulsiveness, respectively. Controlled, empirical men scored higher on the MW posttest, emotional men

showed greater P-I gain and cautious men improved more on P-II. Introverted women had greater CM-II posttraining accuracy; calm women, on P-II; and bold women on the MM exercise. Impulsive, cautious women showed greater MM improvement. The following sex differences were statistically significant: caution was related to P-II posttest scores for men, but unrelated for women; control was related to men's MW posttest performance, but unrelated for women; and gain in MM accuracy tended to be higher for bold men, but was higher for cautious women.

Age for the total sample was negatively related to MW gain; for male Js, negatively related to MM posttest and gain; and unrelated to women's performance. Liking for training was unrelated to training outcome. Attendance at training sessions was reliably related to P-II gain for the total group; to CM-II and P-II (negatively), for men; and to MW gain for women. Leadership inventory R scores correlated negatively with total group CM-I and MM gain, with male sample MW posttest scores and gain, and with female sample CM-I posttest scores and gain. R was positively correlated with female P-I gain. Total leadership inventory scores were positively related to CM-II posttest scores for women Js. With regard to statistically reliable sex differences, younger men scored higher on post-training measures of MW stereotypes; age was unrelated to women's performance. Attendance was unrelated to P-I gain for females, but negatively related to male performance.

As with other measures, intellectual ability also appeared to be related in a specific manner to SA. Scores of both male and female Js were related to some measures of their intellectual performance. Intelligence was a more salient aspect of male accuracy than for that of female Js. Male Js' overall college academic ability varied positively with MW posttest accuracy, but female CQT scores were unrelated. The difference between male and female correlations was statistically reliable. MW improvement was greatest for men with highest course achievement as measured by three final exam scores and three composite examination scores. The male-female differences between correlations were again statistically significant.

CHAPTER IV

DISCUSSION

Results of the Training Studies

Table 17 summarizes the results of the tests of the hypotheses.

Table 17. Summary of the results of the tests of the hypotheses.

Hypothesis	Outcome	
	Study One	Study Two
Training will improve SA.	++	++
The amount of improvement varies inversely with initial accuracy.	++	++
Moderately difficult criterion items best reflect improvement.	+	NA
SA is a situation specific ability.	+	++
SA improvement is situation specific, but will generalize to opposite sex predictions.	NA	++
The delay of impression formation will enhance improvement.	NA	0
The amount of improvement varies with Js personality and intellectual attributes and with his attitude toward training.	?	?

++ = strongly confirmed; + = confirmed; 0 = rejected;
? = inconclusive; NA = not applicable.

The most significant finding in light of the number of training programs reporting little or no success in improving accuracy, was that both of the present studies improved the SA ability. Explanations for the success of the present program lie in the nature of stereotypes and the training activities, and in the interaction of the two.

Neutrally regarded, stereotypes are concept systems which

organize our experiences much as do other concepts (Vinacke, 1957). But while in the course of normal development we are given continual and disciplined guidance in the application of many concepts, the inevitable categorization of people, which we call stereotyping, is relatively de-emphasized or becomes subtle and more complex. The present program provided some belated concept training.

But why was SA training successful while attempts in other areas of sensitivity have failed? Little success was recorded by the programs that were vague as to both the content and process of understanding: these have mainly been the studies equating training and "education" (Fancher, 1967; Goldberg, 1968; Smith, 1966; Taft, 1955). Similarly, little success was realized when the process of understanding was emphasized to the exclusion of content (e.g., Goldberg, 1965). T-group training has consistently stressed process with little demonstrable improvement in interpersonal sensitivity (Campbell and Dunnette, 1968). Programs reporting some success, however, stressed both the content and process of understanding and have combined training procedures and conditions (Dailey, 1966a; Grossman, 1967; Jecker, et al., 1965; Wakeley, 1961).

The present studies, too, combined specific, relevant training procedures and conditions, and to a greater extent than in previous programs, emphasized both the content (the nature of stereotypes, SA, and understanding) and process (practice

and feedback, etc.) of sensitivity. All of this was done, also, with the most effective criterion and training instruments devised to date.

Thus, practice and feedback in the present study provided the chance to test the appropriateness of present concepts and progress towards improvement, while diagnosis gave information and motivation. Wakeley's (1961) pooling technique encouraged the development of stereotypes and introduced Js to the usefulness of base-rates. The base-rate condition was formalized in Study Two to emphasize the need to move from the instance to the class and vice versa. Modification of the pilot predictors and training prompted the dropping of the empathic technique in Study Two with no apparent loss in the power of the program. The explicit, empirical personality trait theory criticized by both Grossman (1967) and Smith (1968a) was appropriate and useful in SA training since personality and intellectual trait names form the basis of stereotypes. The discussion technique of "Why did you answer like you did?" proved unwieldy with the larger group in Study One and gave way to the panel method in Study Two which was a good compromise and at the same time provided information regarding the social reality. The success of the present program lies, then, in an interaction of the nature of stereotypes and the training procedures.

The data supported previous findings of an inverse relationship between pretraining accuracy and improvement (Goldberg, 1965;

Kepes, 1965; Lorge and Diamond, 1954). The phenomenon, possibly regression to the mean or a ceiling effect, occurred on the College Man (Study Two) tests: the already high pretest scores virtually precluded significant gain.

The generality-specificity controversy in psychology extends also to the area of sensitivity. Empirical evidence has been presented by both the generalists (Cline and Richards, 1960; 1961) and the specificists (Crow and Hammond, 1957) to support their respective positions. The results of the present research support the latter group. The Study One data were cautiously interpreted as supportive of the specificity of SA (there were no consistent findings from the V to the C-V groups). But the power of the test was suspect owing to the degree of overlap of the stereotypes tested. Consider, for example, the tasks involved in each of the pilot training instruments:

Morgan Johnson = differentiate between a psychology student and men in general;
Executive Test = differentiate between professional and unskilled men;
Psychologist Test = differentiate between psychologists and other professional men;
Professor Test = differentiate between a psychology professor and college men in general.

Thus, the evidence for generality may simply reflect similarity of the tasks. The only relatively "pure" stereotype was that of psychologists and this predictor also had the greatest internal consistency. The stereotype for Morgan Johnson and the professor were obscured by other group memberships. It is not clear, for example, whether Js' predictions about Morgan were based on a

"student" or a "psychologist" stereotype. It is not surprising to find statistically reliable but irreproducible relations between a task requiring predictions about "average men" and "unskilled workers", or "professionals", "executives", "psychologists", and "college men". The revised SA measures (Study Two) produced less stereotype overlap (the Marital Stereotype test, for example, further separated out males and females) and provided strong support for the specificity hypothesis. SA improvement was also found to be specific.

The hypothesis that delay of impression formation will enhance improvement (Dunnette and Hakel, 1968; Smith, 1966; 1968a) was not supported, but with some doubt about the conclusions to be drawn. The nature of the experimental condition (no written judgments before training) may not have slowed down the judgment process, but may simply have hindered the extent to which a prediction was "fixed". While such findings are certainly relevant, it is more appropriate to limit generalizations to the noncommitment condition itself. In this present study, noncommitment neither enhanced nor detracted from improvement: posttraining accuracy reached the same levels for both experimental and control Js.

Gough (1968), Opsahl (Dunnette, 1968), and Oskamp (1965) found no relation between Js' confidence in gain and actual training outcome. The data from the present studies are consistent with their findings.

Moreover, the students in the present studies consistently ranked the subjective value of the Psychology of Personality course about 3 on a scale from 1 to 5: that is, the average student felt that the course was about as valuable to him as the average course he was taking. In another program, adult employment agency administrators unanimously rated SA training as valuable, relevant, job related, and so on (Spier, in preparation). Results: no relation of training outcome to liking for training.

Implications for Training to Improve Sensitivity

The results of the present studies had implications for 3 broad aspects of sensitivity training: the improvement of training design, selection of trainees, and evaluation of training outcome.

How can training be improved?

While the optimal weighting of each remains to be determined, future sensitivity training programs should include both lectures and participative exerercises. It is not enough to confront an individual with his shortcomings as a judge of people (process). Some cognitive input is required (content) to enable J to assess and integrate the relevancy of his training experience and to generalize from the training setting to other arenas of interpersonal interaction. Similarly, the design of Study Two, in which each training session was a complete unit in itself proved valuable in terms of research methodology and helped to maintain

Js' interest by providing immediate feedback and closure. Such a future design is also relevant in light of the relative specificity of the ability and its improvement.

In Study One, the SA principles were sufficiently general so that marital stereotypes became more accurate even though Js were trained on other stereotypes. Study Two demonstrated that SA principles and procedures may be applied to a wide variety of stereotypes, but that not only the SA ability, but SA improvement, was specific: Js who improved in accuracy on one stereotype, did not necessarily gain in accuracy on other stereotypes. Thus, future programs should apply the general principles to the development of understanding of specific groups. The most relevant question for future programs to ask is "Who do we want to know what about whom?" (Smith, 1968a). The answer may well be that the psychologist wishes to understand his client; the manager, his employee; the college administrator, his restless students; the salesman, his customer; and so on.

The empathic technique proved of little value in SA training and may be dropped in future SA programs as a separate condition. But the trait method should be included even though it has not proven successful in other areas of sensitivity training. The panel method, while seemingly a good compromise for discussion considering the size of the training groups, should not supplant the need for smaller training groups. Most importantly, combinations of training procedures were most effective. Additional

techniques may prove useful. Goldberg (1968), for example, likened the acquisition of understanding to the research process: we should constantly be forming our impressions of others as hypotheses to be tested and revised and rejected. A trial training session using an hypothesis approach showed promise. Thus, SA was improved in the present studies, and the task of the future may well be the search for new techniques and technique combinations that will further maximize outcome.

Finally, except for the diagnosis, little attention was given the trainees' motivation to improve. Some findings suggested that this might be a serious omission. While a ceiling effect might explain the failure of moderately and highly accurate Js to gain on initially easy items, low accuracy Js showed no improvement on the low difficulty items even though they showed the greatest overall improvement as a result of training. Consider, however, that each J enters training with some "picture" of himself as a judge of people, and that Js have varying degrees of confidence in their numerous predictions. Thus, judgments perceived as easiest (i.e., those having the highest subjective probabilities of correctness) should be the least amenable to change. What, then, was the effect of diagnostic feedback which informed J of his actual prowess?

The Marital Stereotypes test was revised into 4 series (matched for item difficulty) of 14 items each. As a part of SA training, adult employment agency administrators estimated

the number of items they expected to answer correctly before training and received feedback about actual performance after each series. Js unanimously overestimated their actual performance for the first 14 items, and as a group, drastically reduced their estimates for the second series of items. Thereafter, two trends emerged: some Js consistently set higher goals for themselves than their just previous performance; the remaining Js consistently set their goals lower than their just previous performance. The former were superior in SA throughout, and showed signficantly greater gain in accuracy. A second study replicated the results (Spier, in preparation). Failure to improve, then, may result from Js' focus on predicting their own ability when the task at hand is actually the predictions about others. Conclusion: sensitivity training programs should attempt to measure (perhaps through level of aspiration), interpret, and deal openly with Js' feelings about their ability as judges of others. Some perspective needs to be supplied to motivate J to predict for O and to prevent a possible defensive resistence to training.

Who should be trained?

The findings that SA and its improvement were specific suggested that virtually anyone could benefit from training on relevant stereotypes. Similarly, both men and women showed comparable accuracy gains. But additional data suggested greater complexity. A study of the correlates of SA improvement showed them to be similarly stereotype specific. That is,

personality and intellectual factors were not related to SA improvement in a general way. Rather, J's specific personality and intellectual traits were related to improvement on specific stereotypes as the programs were presently conducted. The overall implications were not entirely clear, but further research is certainly necessary. Interesting implications exist should personality and intellectual factors turn out (on cross-validation) to be stereotype specific. Perhaps, to obtain optimal results from an SA training effort, tailor-made training designs may be indicated for groups that differ in personality and/or intellectual make-up.

The specificity issue was also relevant to the finding that training had its greatest impact on initially inaccurate Js. Since SA and gain were specific, diagnosis of pretraining accuracy may serve as a screening device in future programs. Motivation may be a greater problem, for example, among initially high scoring Js. Also, the costs of interfering with an already effective judging system may outweigh the slight benefits that accrued to the accurate Js.

How should training outcome be evaluated?

The criterion problem is an issue in evaluating the effectiveness of sensitivity training as it is in all areas of psychological research. But, perhaps more than in most areas, there has been a tendency to accept anecdotal evidence and subjective feelings about sensitivity as indications of training success. Based

on the present findings, Js' self reported confidence in improvement and liking for training should not be used as criteria of training effectiveness. Clearly, the problem has been greatest for the T-group training efforts, but the search for adequate criterion and training measures is no less acute in studies similar to the present one.

An exploration was made of the relation of training impact and criterion item difficulty in Study One. The data supported the hypothesis that moderately difficult items best reflected training outcomes. But items at each level of difficulty measured a unique aspect of the total training impact. Moreover, there was no difference between outcomes on the moderately difficult and hard items. Only the easy items showed no change. Predictions about others, however, range from the easy to the difficult in reality. Conclusion: future sensitivity predictors and training instruments should contain items at all levels of difficulty.

Implications for Future Research

Emphasis on both the content and process of understanding used combinations of training techniques and conditions and improved SA. Future studies need to explore the relative contribution made to the total outcome by each of the procedures. Training has improved SA, but future programs should seek to maximize the impact through the design of new programs and evaluation and redesign of existing paradigms (e.g., additional concern for J's motivation to improve); development and use of

new training procedures (e.g., the "hypothesis technique"); and continued evaluation and revision of criterion and training measures. An additional question of interest concerns the extent to which training results endure over time.

The most serious shortcoming of the present research was the lack of adequate control groups. The absence of control groups limited the generality of the present results despite the fact that previous studies using similar techniques have shown no improvement in control groups. Future studies should include adequate control groups.

An exploration of the correlates of improvement showed the stereotype specific nature of the personality and intellectual factors measured. The importance of the specificity issue suggested a need to cross-validate the findings. "Purer" (i.e., less stereotype overlap) SA measures may help resolve the question of whether the observed specific relationships are spurious. Additionally, an exploration of the nature, correlates, and measurement of the SA ability itself has relevance to the issue at hand and to the question of "Who do we want to know about whom?". While the present studies used a shot-gun approach, the proposed research should provide the rifle needed to aim programs more precisely and derive maximum outcome.

CHAPTER V

SUMMARY AND CONCLUSIONS

Two experimental training programs used a pretest-training-posttest paradigm to improve stereotype accuracy (SA): the ability to predict group norms. In Study One, a pilot program, training improved predictions about marital stereotypes even though training was provided on executive, psychologist, college student, and professor stereotypes. But the criterion instruments were unwieldy and the single criterion design gave little control over what occurred between pretest and posttest, while a great deal of sample shrinkage took place from the first to the last SA training session.

Study Two's revised and shortened criterion-training instruments made each training session a complete unit in itself. Significant SA improvement occurred in the second study, but improvement on one stereotype did not assure improvement on others, and vice versa. Improvement showed no erosion in Js retested 7 weeks after training.

Improvement in Study One suggested generality of the underlying principles and training designs. The specificity of SA and its improvement, research design requirements, and ease of administration, however, favor the single session unit design of Study Two. Moreover, combinations of training activities within a program emphasizing both content and process hold

the greatest promise for maximization of training impact. Practice and feedback formed a core for the additional discussions, diagnosis, base-rate, noncommitment, pooling, and panel procedures and conditions of the present studies.

An additional finding was that Js who did not make written pretraining predictions improved about as much as Js who had made the written judgment.

Additional data showed that the personality and intellectual correlates were also stereotype specific, and that initially inaccurate Js benefited more from training than initially accurate Js. Both findings had relevance for the selection of trainees.

Finally, moderately difficult criterion items best reflected training outcome from a statistical point-of-view. From a practical standpoint, however, the differences were not worthwhile. SA measures should continue to reflect judging reality and include the total range of item difficulties. Subjective feelings of improvement and liking for training were unrelated to actual outcome which precludes their use as criteria for training effectiveness.

Implications of the findings for future training and research were discussed.

BIBLIOGRAPHY

Bronfenbrenner, U., Harding, J., and Gallwey, Mary The measurement of skill in social perception. In D.C. McClelland (Ed.), Talent and Society. New York: Van Nostrand, 1958, 29 - 108.

Bruni, E. A film test of accuracy in observing people and its correlates. Unpublished master's thesis, Michigan State University, 1963.

Campbell, D.T. Stereotypes and the perception of group differences. American Psychologist, 1967, 22, 817 - 829.

Campbell, J.P., and Dunnette, M.D. Effectiveness of T-group experiences in managerial training and development. Psychological Bulletin, 1968, 70, 73 - 102.

Cline, V.B. Person perception from the perspective of an empiricist. Paper delivered at the 76th annual meeting of the American Psychological Association, San Francisco, California, 1968.

Cline, V.B. and Richards, J.M., Jr. The generality of accuracy of interpersonal perception. Journal of Abnormal and Social Psychology, 1961, 62, 446 - 449.

Cline, V.B. and Richards, J.M., Jr. Accuracy of interpersonal perception: a general trait? Journal of Abnormal and Social Psychology, 1960, 60, 1 - 7.

Close, Patricia A test of stereotype accuracy in judging differences in marital couples. Unpublished manuscript, 1963.

Cronbach, L.J. Processes affecting scores on "understanding of others" and "assumed similarity". Psychological Bulletin, 1955, 52, 177 - 193.

Crow, W.J. and Hammond, K.R. The generality of accuracy and response sets in interpersonal perception. Journal of Abnormal and Social Psychology, 1957, 54, 384 - 390.

Dailey, C.A. The experimental study of clinical guessing. Journal of Individual Psychology, May, 1966, 22, 65 - 79.

Dailey, C.A. Prejudice and decision making. Personnel Administration. September - October, 1966, 6 - 13.

Dore, R. The development and validation of forced choice scales measuring attitudes toward leadership methods. Unpublished master's thesis, Michian State University, 1960.

Dunnette, M.D. Forms of interpersonal accommodation: processes, problems, and research avenues. Paper delivered at the 76th annual meeting of the American Psychological Association, San Francisco, California, 1968.

Dunnette, M.D. and Hakel, M.D. Interpersonal perception and behavior prediction. A proposal for research submitted to the National Science Foundation under NSF Grant GS-1081, 1968.

Fancher, R.E., Jr. Accuracy vs. validity in person perception. Journal of Consulting Psychology, 1967, 31, 264 - 269.

Goldberg, L.R. Learning clinical inference: the results of intensive training on clinicians ability to diagnose psychosis vs. neurosis from the MMPI. Paper delivered at the annual meeting of the Western Psychological Association, Honolulu, Hawaii, June, 1965.

Goldberg, L.R. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483 - 496.

Gough, H.G. An exploratory cross-cultural study of interpersonal perception. Paper delivered at the 76th annual meeting of the American Psychological Association, San Francisco, California, 1968.

Grossman, B.S. Evaluation of a training program to improve the ability to differentiate between people. Unpublished doctoral dissertation, Michigan State University, 1967.

Harding, J. Stereotypes. In David L. Sills (Ed.) International Encyclopedia of the Social Sciences. New York: Macmillan and Free Press, 1968, 15, 259 - 262.

Harris, W. The relation of observational to inferential accuracy in judging people. Unpublished master's thesis, Michigan State University, 1962.

Jecker, J.D., Maccoby, N., and Breitrose, H.S. Improving accuracy in interpreting non-verbal cues of comprehension, Psychology in the Schools, July, 1965, 2(3), 239 - 244.

Johnson, R.L. Correlates of a test of group sensitivity. Unpublished master's thesis, Michigan State University, 1963.

Johnson, Winifred B., and Terman, L.M. Personality characteristics of happily married, unhappily married, and divorced persons. Character and Personality, 1935, 3, 290 - 311.

Kepes, S.Y. Experimental evaluations of sensitivity training. Unpublished doctoral dissertation. Michigan State University, 1965.

Linden, J. The self-centered orientation in interpersonal relationships. Unpublished master's thesis. Michigan State University, 1965.

Locke, E.A. Motivational effects of knowledge of results: knowledge or goal setting? Journal of Applied Psychology, 1967, 51, 324 - 329.

Lorge, I., and Diamond, Lorraine K. The value of information to good and poor judges of item difficulty. Educational and Psychological Measurement, 1954, 14, 29 - 33.

Mietus, J. Development of a multiple component criterion for the evaluation of sensitivity training. Unfinished master's thesis, Michigan State University, 1969.

Oskamp, S.W. Overconfidence in case-study judgments. Journal of Consulting Psychology, 1965, 29, 261 - 265.

Price, K. Is the favorableness of interpersonal evaluations a general trait? Unfinished master's thesis, Michigan State University, 1969.

Shears, A.D. The development and analysis of a criterion of the ability to understand people. Unpublished master's thesis, Michigan State University, 1967.

Silkiner, D.S. A cross-cultural study of the measurement, determinants, and effects of stereotype accuracy. Unpublished master's thesis, Michigan State University, 1962.

Smith, H.C. Bill Wilkins as a model for sensitivity training. Paper delivered at the 76th annual meeting of the American Psychological Association, San Francisco, California, 1968a.

Smith, H.C. Personality Development. New York: McGraw-Hill, 1968b.

Smith, H.C. Sensitivity to People. In H. Toch and H.C. Smith (Eds.) Social Perception: The Development of Interpersonal Impressions. Princeton, N.J.: D. Van Nostrand, 1968c, 10 - 19.

Smith, H.C. Sensitivity to People. New York: McGraw-Hill, 1966.

Springbett, B.M. Factors affecting the final decision in the employment interview. Canadian Journal of Psychology, 1958, 12, 13 - 22.

Stelmachers, Z.T. and McHugh, R.B. Contribution of stereotyped and individualized information to predictive accuracy. Journal of Consulting Psychology, 1964, 28, 234 - 242.

Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1 - 23.

Tagiuri, R. Person perception. In David L. Sills (Ed.) International Encyclopedia of the Social Sciences. New York: Macmillan and Free Press, 1968, 11, 560 - 567.

Vinacke, W.E. Stereotypes as social concepts. Journal of Social Psychology, 1957, 46, 229 - 243.

Wakeley, J.H. The effects of specific training on accuracy in judging others. Unpublished doctoral dissertation, Michigan State University, 1961.

Zavala, A. A test of stereotype accuracy. Unpublished master's thesis, Michigan State University, 1960.

APPENDIX A

STUDY ONE STEREOTYPE ACCURACY MEASURES

Attitudes of American Workers

Several thousand workers in 150 companies in an anonymous questionnaire were asked to indicate whether they thought the statements below were "true" or "false". How well can you predict what their answers were? If you think a majority agreed with the statement, answer "true". If you think a majority disagreed with the statement answer "false". Remember to answer them not as you would but as the typical worker would.

T F 1. The average American worker today really enjoys his work.

T F 2. Most of the time the average worker is worried about being laid off.

T F 3. The average worker takes pride in what he does on the job.

T F 4. He thinks that wages are increased primarily because employees are able to demand an increase.

T F 5. He does not feel that his Company's investment in new labor-saving equipment makes it possible for him to earn more money.

T F 6. He thinks he should be advanced by seniority rather than by ability.

T F 7. He should turn out as much work as he can.

T F 8. He feels that his work is important to the Company's customers.

T F 9. He feels that his management thinks his work is important to the Company.

T F 10. He feels that the Company will prosper whether he does his work well or not.

T F 11. He believes that everyone on the job would benefit if each worker did the best he could.

T F 12. He feels he receives fair treatment from his supervisor.

T F 13. He is more concerned about his job security than about high pay.

T F 14. He is more interested in getting advancement than in having an easy job.

T F 15. He feels that he gets impartial treatment from the management of his Company.

Marriage Test

Directions: The following test measures your ability to identify differences between the typical happily married, unhappily married, and divorced person. Part I deals with men; Part II, with women.

For each of the following statements, mark the one of the three groups you think is best described by it.

The correct answers are based on the answers that the groups made on lengthy questionnaires. For example, 100 happily married women, 100 unhappily married women, and 100 divorced women answered the question: "Do you prefer a play to a dance?". Results:

81% of the happily married women answered "yes".
58% of the unhappily married women answered "yes".
44% of the divorced women answered "yes".

Therefore, the correct answer to the statement "Most apt to prefer a play to a dance" is "happily married women".

PART I. MEN

Mark: (1) if you think the correct answer is "happily married men"; (2) "unhappily married men"; or (3) "divorced men".

Correct		Difficulty Level
1	1. Most tolerant of sick people.	4
1	2. Most prefers to spend a night at home.	6
1	3. Most apt to like religious people.	22
2	4. Slowest in making decisions.	37
3	5. Most likely to enjoy taking risks.	23
1	6. Most prefers to make plans with others.	54
3	7. Most apt to make bets.	48
1	8. Most likely to organize a club or team.	46
2	9. Most dislikes modern languages.	44
2	10. Most apt to be critical of others.	29
2	11. Most touchy on the most subjects.	26
3	12. Most prefers fashionably dressed people.	19
3	13. Most often experiences feelings of loneliness.	50
2	14. Most dislikes foreigners.	40
3	15. Most apt to like the occupation of novelist.	46
2	16. Least often takes the lead to enliven a dull party.	49

Correct		Difficulty Level
2	17. Least interested in artistic activities.	51
3	18. Most likely to like the occupation of stock broker.	60
2	19. Least likes symphony concerts.	43
3	20. Most likes the occupation of criminal lawyer.	56
1	21. Most likely to stress quality in his work.	23
1	22. Most apt to like talkative people.	53
2	23. Least often organizes teams or clubs.	58
2	24. Most likely to think of himself as a radical while actually being conservative.	63
1	25. Most likes cautious people.	46
2	26. Least interest in the occupation of teaching.	57
1	27. Most meticulous and methodological in work.	45
3	28. Most likely to enjoy competition.	60

PART II. WOMEN

Mark: (1) if you think the correct answer is "happily married women"; (2) "unhappily married women"; or (3) "divorced women".

Correct		Difficulty Level
1	29. Most apt to like old people.	18
2	30. Most often has spells of dizziness.	6
1	31. Most apt to like religious people.	26
1	32. Most apt to like music.	28
3	33. Most prefers working on commission to definite salary.	31
3	34. Most willing to be unconventional.	26
3	35. Most prefers taking chances to playing safe.	30
1	36. Most tolerant of minority groups.	31
2	37. Most apt to arrive late for work.	27
2	38. Most often troubled by feelings of inferiority.	31
2	39. Most likely to consider themselves as nervous.	13
1	40. Most likes picnics and excursions.	11
3	41. Most interested in change and travel.	26
3	42. Most prefers work that makes heavy demands.	43
2	43. Least effective in emergencies.	31
1	44. Most conservative in social and political opinions.	34
3	45. Least likely to blush.	41
2	46. Least methodological.	49
3	47. Most self-assertive and self-reliant.	33
2	48. Least willing to work things out for themselves.	29

Correct		Difficulty Level
2	49. Most interested in avoiding technical responsibilities.	51
3	50. Most interested in being an inventor.	41
1	51. Most dislikes quick-tempered people.	55
3	52. Most ambitious	45
1	53. Most apt to dislike working in isolation.	50
3	54. Most apt to like playing chess.	67
1	55. Most apt to like people who never drink.	35
2	56. Most apt to like psychology.	84

The Interests of Executives vs. Unskilled Workers

A large number of executives and professional men (lawyers, managers, etc.) and a large number of unskilled workers (laborers, porters, etc.) checked whether they liked or disliked each of the interests or activities below. A larger percentage of the unskilled workers liked half of the interests; a larger percentage of the executive and professional group liked the other half.

Mark "1" if you think more unskilled workers liked a particular interest. Mark "2" if you think more professional workers and executives liked it.

	Difficulty Level
1 1.	Typist 40
1 2.	Bank Teller 41
1 3.	Civil Service Employee 39
1 4.	Interior Decorator 81
2 5.	Sales Manager 32
1 6.	Secret Service Man 43
1 7.	Office Clerk 19
2 8.	Scientific Research Worker 7
1 9.	Draftsman 43
1 10.	Electrical Engineer 81
2 11.	Editor 15
1 12.	Marine Engineer 69
2 13.	Magazine Writer 29
2 14.	Advertiser 20
2 15.	Lawyer, Corporation 6
1 16.	Lawyer, Criminal 72
2 17.	Manufacturer 46
2 18.	Physics 8
1 19.	Physical Training 13
2 20.	Mathematics 14
2 21.	History 18
1 22.	Agriculture 12
2 23.	Golf 5
1 24.	Hunting 21
1 25.	Boxing 5
2 26.	Musical Comedy 23
1 27.	Pet Monkeys 29
1 28.	Detective Stories 15
2 29.	"New Republic" 13
2 30.	Conservative People 32
2 31.	People Who Chew Gum 97
1 32.	Socialists 38
2 33.	Energetic People 23
2 34.	Optimists 38
2 35.	People Who Assume Leadership 29
1 36.	People Who Talk Very Slowly 15
1 37.	People With Gold Teeth 20
1 38.	Nervous People 40
1 39.	People Who Always Agree With You 25
2 40.	Thrifty People 50
2 41.	Religious People 74
2 42.	Prepare the Advertising for a New Machine 23
1 43.	Courteous Treatment from Superiors 23
2 44.	Opportunity to make use of all one's knowledge and experience 12
1 45.	J. J. Pershing, Soldier 18
2 46.	William H. Taft, Jurist 18
2 47.	John Wanamaker, Merchant 53
2 48.	President of a Club 15
2 49.	Do a job yourself 67
1 50.	Definite Salary 34
2 51.	Work for Yourself 33
2 52.	Great Variety of Work 31
2 53.	Emphasis on Quality of Work 15
1 54.	Opportunity to understand just how one's superior expects work to be done 34
2 55.	Freedom in working out one's own methods of doing the work 27
1 56.	Repairing a Clock 16
1 57.	Repairing Electrical Wiring 21
1 58.	Giving "first aid" assistance 36
2 59.	Adjusting difficulties of others 17
1 60.	Climbing along edge of precipice 38

The Case of the Psychology Professor
(Stereotype Training)

The statements below were in the Inventory completed by the professor. On these particular statements, as it turned out, he gave the same answers as those given by more than two-thirds of the college men who completed the inventory. The question here, therefore, is not only just how well you understand the professor. It is also how well you understand the typical college man.

Circle the answer "true" if you think both the college professor and the typical college man answered it "true". Circle the answer "false" if you think both the college professor and the typical college man answered it "false".

			Difficulty Level
T	F	1. I prefer quiet games to extremely active ones.	44
T	F	2. I am considered rather emotional by my friends.	19
T	F	3. I have occasional difficulty getting the temperature of my bath the way I like it.	39
T	F	4. Divine inspiration is an infallible source of truth.	11
T	F	5. I sometimes think more about my ideas than about the routine demands of daily life.	15
T	F	6. I often think for a long time about an idea that has occurred to me.	13
T	F	7. I am generally active in my everyday life.	11
T	F	8. I am practically always tolerant even in dealing with people that I don't like.	18
T	F	9. I always keep control of myself in an emergency situation.	31
T	F	10. I occasionally neglect serious things in order to have a good time.	26
T	F	11. I always finish one task before taking on others.	32
T	F	12. I have occasionally doubted the reality of God.	6
T	F	13. I like to discuss abstract questions with my friends.	14
T	F	14. Artistic experiences are of great importance in my life	31
T	F	15. Quite a few things make me emotional.	24
T	F	16. The thought of God gives me a complete sense of security.	10

Difficulty Level

T F 17. Radical agitators should be allowed to make public speeches 23

T F 18. I believe that what a person does about a thing is more important than what he feels about it. 29

T F 19. I become emotional fairly easily. 17

T F 20. I would rather be a salesman than a scientific research worker. 15

To the following statements, the professor gave a different answer to the one given by two-thirds or more of the college men. The question here, therefore, is: How does the professor differ from the typical college man?

Circle "true" if you think the professor answered "true" but the typical student answered "false".

Circle "false" if you think the professor answered "false" but the typical student answered "true".

Difficulty Level

T F 21. I believe that competitiveness is a necessary and desirable part of our economic life. 38

T F 22. I am greatly influenced in minor decisions by how I happen to feel at the moment. 19

T F 23. I enjoy work more than play. 17

T F 24. I like ballet performances. 12

T F 25. I think cremation is the best method of burial. 27

T F 26. Compared to your own self-respect, the respect of others means little. 49

T F 27. I think I would like to decorate a room with flowers. 27

T F 28. I have never tried to collect pictures of paintings I have liked. 15

T F 29. I am cautious about undertaking anything which may lead to humiliating experiences. 28

T F 30. I am a very adventurous person. 30

T F 31. I am always taking on added social responsibilities. 54

T F 32. I'm occasionally disorganized if I am called on suddenly to make a few remarks. 25

Difficulty Level

T F 33. If I had the ability, I would enjoy teaching poetry at a University. 20

T F 34. I prefer friends who have well developed artistic tastes. 19

T F 35. I find it rather hard to keep to a rigid routine. 19

T F 36. I believe in getting as much fun as I can out of life. 10

T F 37. I would rather see a movie than read a book. 5

T F 38. I never complain about my sufferings and hardships. 29

T F 39. Sports generally interest me somewhat more than very intellectual affairs. 7

T F 40. I have never been seasick, plane sick, or car sick. 47

The Special Interests of Psychologists

DIRECTIONS: How do the interests of psychologists differ from those of other men? To answer the question, several hundred male psychologists and several thousand other business and professional men checked whether they would "like" each of many different occupations, amusements, activities, and kinds of people.

A higher percentage of the psychologists liked some interests. For example, 41 percent of the psychologists said they would like to be the "author of a novel"; only 32 percent of men in general expressed such a liking. A lower percentage of the psychologists liked some interests. For example, only 29 percent of the psychologists said they would like to be a "sales manager" whereas 37 percent of men in general expressed a liking for this occupation.

Mark for each of the interests below whether you think more or less psychologists liked the interest. Mark "1" if you think a higher percentage of psychologists than men in general liked the interest; mark "2" if you think a lower percentage of psychologists liked the interest.

Correct		Difficulty Level
1	1. Actor	33
1	2. Artist	18
1	3. Astronomer	54
2	4. Corporation lawyer	47
2	5. Manufacturer	14
2	6. Athletic director	27
1	7. Chemist	52
2	8. Cashier in bank	16
1	9. Editor	28
1	10. Foreign correspondent	21
1	11. Inventor	24
1	12. Magazine writer	28
2	13. Office manager	21
1	14. Orchestra conductor	54
1	15. Physician	23
1	16. Poet	27
2	17. Rancher	17
1	18. Sculptor	37
1	19. Statistician	40
1	20. Surgeon	36
2	21. Wholesaler	8
1	22. Geometry	48
1	23. Algebra	46
2	24. Physical training	32
1	25. Physiology	19
1	26. Literature	7

Correct		Difficulty Level
2	27. Hunting	24
1	28. Symphony concerts	22
2	29. Sporting pages	33
2	30. Golf	68
1	31. Chess	13
1	32. Solving mechanical puzzles	37
2	33. Travel movies	53
2	34. Fishing	39
1	35. Making a speech	22
1	36. Teaching adults	12
2	37. Taking responsibility	88
1	38. Doing research work	5
1	39. Writing reports	25
2	40. Regular hours of work	31
2	41. Developing business systems	31
2	42. Saving money	42
2	43. Conservative people	32
2	44. Energetic people	83
2	45. People who are natural leaders	76
2	46. People who make fortunes in business	38
2	47. Thrifty people	48
2	48. Religious people	42
1	49. Socialists	44
1	50. Independents in politics	13
1	51. People who talk about themselves	28
1	52. Carelessly dressed people	56
1	53. Absent-minded people	57
2	54. Outside work	59
2	55. Physical activity	42
1	56. Usually drive myself steadily	31
1	57. Have more than my share of novel ideas	17
2	58. My feelings are easily hurt	19
1	59. My advice is sought by many	9
2	60. Put drive into the organization	68

The Case of Morgan Johnson

Morgan is a twenty-two-year-old unmarried college senior who is planing to study psychology in graduate school. His parents died when he was four, and he and his younger brother were raised by permissive grandparents in Brooklyn. Of his childhood, Morgan said: "As I grew up, I always had the feeling that I was inferior to everybody else because I had no parents. In grade school, I was very loud and boisterous and made persistent attempts to dominate my peers and to excel in everything I did." Today he places emphasis on being a "well-rounded scholar." About his values, he now says:

> "I do not believe there are any determining forces in the universe that make us what we are; everybody rules his own destiny. I can think of nothing more important than being a good friend or having good friends, but I don't think it is possible to have more than a few really close ones. I place little value on material things: cars, clothes, etc."

Morgan filled out the Strong Vocational Interest Blank that requires the respondent to answer "Like," "Indifferent," or "Dislike" to a long list of interests. The directions of this test ask the respondent to "disregard considerations of salary, social standing, future advancement, etc... consider only whether or not you would enjoy the interest regardless of any necessary skills, abilities, or training which you may or may not possess."

DIRECTIONS: In the first group of interests below Morgan gave the <u>same</u> answer as more than half of several thousand representative American men. Mark the answer that you think was given by both Morgan and the typical men. Use "1" for "Like"; "2" for "Indifferent"; and "3" for "Dislike".

Morgan is Like the Typical Man

					Difficulty Level
L	I	<u>D</u>	1.	Auctioneer	64
L	I	<u>D</u>	2.	Auto salesman	75
L	I	<u>D</u>	3.	Auto repairman	65
L	I	<u>D</u>	4.	Typist	27
L	<u>I</u>	D	5.	"Atlantic Monthly"	55
<u>L</u>	I	D	6.	Literature	34
<u>L</u>	I	D	7.	Philosophy	39
<u>L</u>	I	D	8.	Driving an automobile	24
<u>L</u>	I	D	9.	Sporting pages	13
<u>L</u>	I	D	10.	"Time"	19
<u>L</u>	I	D	11.	Meeting and directing people	22
<u>L</u>	I	D	12.	Meeting new situations	27
<u>L</u>	I	D	13.	Adjusting difficulties of others	57
<u>L</u>	I	D	14.	Contributing to charities	75
<u>L</u>	I	D	15.	Progressive people	26
L	<u>I</u>	D	16.	Foreigners	30

Difficulty Level

L I D 17. Sick People 50
L I D 18. Cripples 40
L I D 19. People with protruding jaws 33
L I D 20. People with hooked noses 38
L I D 21. Blind people 41
L I D 22. Deaf Mutes 38
L I D 23. Men who chew tobacco 60
L I D 24. People who chew gum 28
L I D 25. Civil Service Employee 86
L I D 26. Clergyman 74
L I D 27. Printer 76
L I D 28. Drilling in a company 46
L I D 29. Pet monkeys 61
L I D 30. Snakes 44
L I D 31. Acting as yell-leader 47
L I D 32. Quick-tempered people 13

Morgan is Unlike the Typical Man

DIRECTIONS: In the group of interests below, the answer given by Morgan is the one in small letters (l,i,d). His answers to these interests, however, were different from those given by the majority of men. From the two possible answers not chosen by Morgan that are indicated by capitals (L, I, or D) mark the one that you think was chosen by the typical man.

Difficulty Level

1 I D 33. Auto Racer 60
1 I D 34. Consul 40
L i D 35. Factory worker 21
L i D 36. Floorwalker 21
1 I D 37. Poet 52
1 I D 38. Politician 59
L i D 39. Real estate salesman 77
L i D 40. Specialty salesman 65
L I d 41. Algebra 60
L I d 42. Arithmetic 45
L I d 43. Economics 48
L I d 44. Physics 68
L I d 45. Physiology 67
L I d 46. Taking long walks 53
L I d 47. Performing sleight of hand tricks 56
1 I D 48. Fortune tellers 37
L i D 49. Detective stories 7
L I d 50. Educational movies 66
L I d 51. Travel movies 39
L I d 52. Operating machinery 42
1 I D 53. Pursuing bandits in sheriff's posse 57
L I d 54. Regular hours of work 14
L I d 55. Saving money 17
L I d 56. Conservative people 50
L i D 57. Energetic people 9

				Difficulty Level
L	i	D	58. Optimists	10
L	i	D	59. People who are natural leaders	11
L	I	d	60. Thrifty people	50
L	i	D	61. Spendthrifts	15
1	I	D	62. People who always agree with you	37
L	i	D	63. Bolshevists	9
L	i	D	64. Fashionably dressed people	15

PERSONAL PROFILE SHEET

	Percentiles										
	0	10	20	30	40	50	60	70	80	90	100
Session 6. Marriage Stereotypes	16	27	30	31	32	34	36	38	40	43	48
Session 7. Male Stereotypes (Morgan Johnson)	23	30	32	34	36	37	39	40	42	43	49
Session 8. Executives vs. Unskilled	8	35	38	40	41	42	43	45	46	48	50
Session 8. Psychologist Stereotypes	18	31	33	36	37	39	41	43	46	48	53

APPENDIX B

RESULTS OF ANALYSES FOR STUDY ONE: ADDITIONAL DATA

The Interaction of Stereotype Training Improvement, Initial Accuracy of Judges, and Difficulty of Predictions.

(Fall, 1967 Training Data)

Criterion Pretest Accuracy*		Criterion Pretest Item Difficulty** EASY (19 items)	MODERATE (19 items)	HARD (18 items)	TOTAL (56 items)
LOW: (N = 38)	Post	13.8	12.0	8.4	11.4
	Pre	13.3	8.8	6.0	9.4
	Gain	+ .5	+ 3.2	+2.4	+ 2.0
MODERATE: (N = 58)	Post	15.0	13.5	8.7	12.6
	Pre	15.3	11.8	7.5	11.5
	Gain	- .3	+ 1.7	+1.2	+ 1.1
HIGH (N = 41)	Post	15.9	14.5	10.4	13.6
	Pre	17.1	14.5	10.5	14.0
	Gain	- 1.2	0	- .1	- .4
TOTAL: (N = 137)	Post	44.7	40.0	27.5	37.4
	Pre	45.7	35.1	24.0	34.9
	Gain	- 1.0	+ 4.9	+3.5	+ 2.5

* Low criterion pretest accuracy: raw score range = 0 - 31; Moderate raw score range = 32 - 37; High raw score range = 38+

**Easy Item Difficulty range = 0 - 29; Moderate range = 30 - 46; Hard range = 47+

Relation of personality to training outcome. Total samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Impulsive-Controlled	.11	.16	-.14	-.05
Rational-Empirical	-.04	-.08	-.10	-.05
Introverted-Extroverted	-.10	-.15	-.19	-.09
Cautious-Bold	-.06	.03	-.11	-.07
Emotional-Calm	.03	.15	.15	.10
Acquiescence	-.04	-.02	-.13	-.20

Relation of personality to training outcome. Male samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Impulsive-Controlled	.23	.17	-.18	-.06
Rational Empirical	-.03	-.02	.27	.14
Introverted-Extroverted	.04	-.25	-.25	-.22
Cautious-Bold	.02	.04	-.22	-.04
Emotional-Calm	.27	.26	.29	.13
Acquiescence	.03	-.05	-.52**	-.41*

Relation of personality to training outcome. Female samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Impulsive-Controlled	.05	.18	-.05	-.09
Rational-Empirical	.03	-.09	-.37*	-.33
Introverted-Extroverted	-.17	-.01	-.02	-.07
Cautious-Bold	-.19	-.02	.02	-.16
Emotional-Calm	-.09	.12	.10	.05
Acquiescence	-.16	-.02	.19	.06

* $p < .05$
** $p < .01$

Relation of intellectual factors to training outcome. Total samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Midterm-Text	.23*	.21	.24*	.21
Midterm-Class	.04	.01	.20	.08
Midterm-Total Grade	.19	.16	.26*	.24
Final-Text	.19	.13	.11	.19
Final-Class	.20	.20	.05	.08
Final-Total Grade	.21	.18	.10	.10
GPA	.32**	.32*	.17	.12
CQT-Verbal	.28*	.16	.07	.13
CQT-Information	.23*	.25*	.15	.16
CQT-Total	.27*	.21	.15	.13

* $p < .05$
** $p < .01$

Relation of intellectual factors to training outcome. Male samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Midterm-Text	.26	.31	.26	.13
Midterm-Class	-.08	-.12	.16	.24
Midterm-Total Grade	.15	.16	.26	.30
Final-Text	.19	.18	.23	.47*
Final-Class	.41**	.41**	-.02	.29
Final-Total Grade	.33*	.33*	.15	.32
GPA	.34*	.46**	.40*	.50**
CQT-Verbal	.35*	.29	.38*	.28
CQT-Information	.30	.44**	.42*	.42*
CQT-Total	.42**	.44**	.52**	.32

* $p < .05$
** $p < .01$

Relation of intellectual factors to training outcome. Female samples

	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Midterm-Text	.23	.14	.26	.28
Midterm-Class	.18	.17	.26	-.13
Midterm-Total Grade	.25	.18	.31*	.13
Final-Text	.22	.10	.05	-.13
Final-Class	.01	-.02	.10	-.14
Final-Total Grade	.13	.05	.08	-.15
GPA	.31	.17	.06	-.19
CQT-Verbal	.21	.03	-.16	-.01
CQT-Information	.22	.07	.12	-.23
CQT-Total	.21	.05	-.03	-.15

* $p < .05$
** $p < .01$

The relation of attitude toward training and training outcome.

	A. Total samples			
	A. Validation Sample		B. Cross-Validation Sample	
	Criterion Posttest	Gain	Criterion Posttest	Gain
Confidence	-.06	-.19	.02	-.11
Liking for training	.18	-.05	.10	.09
Number of sessions attended	.03	.01	-.08	-.12
	B. Male samples			
Confidence	.05	-.04	-.00	-.27
Liking for training	.11	-.08	.32	.28
Number of sessions attended	.23	-.01	-.15	-.24
	C. Female samples			
Confidence	-.16	-.34	.18	.09
Liking for training	.23	-.06	-.17	-.31
Number of sessions attended	-.17	.03	-.01	-.02

APPENDIX C

STUDY TWO STEREOTYPE ACCURACY MEASURES

EXERCISE 1: The Typical College Man

Part 1: This half of the exercise is designed to improve your understanding about what the typical college man thinks and says about himself as well as to measure your improvement.

1. Pretest. The replies of several hundred men who completed the Protebob Personality inventory were analysed. At least two thirds of the men answered the particular statements below in the same way. To about half of these statements the answer was "true"; the other half, "false".

Mark "1" on the separate answer sheet if you think the typical college man answered "true" to the statement;
Mark "2" if you think the typical college man answered "false".

Correct				Difficulty Level
2	1.	(17)	I think I would like to decorate a room with flowers	27
2	2.	(18)	I prefer friends who have well developed artistic tastes.	19
2	3.	(19)	Quite a few things make me emotional.	24
1	4.	(20)	I believe that competitiveness is a necessary and desirable part of our economic life.	
2	5.	(21)	If I had the ability, I would enjoy teaching poetry at a University.	
2	6.	(22)	I have occasional difficulty getting the temperature of my bath the way I like it.	38
2	7.	(23)	I become emotional fairly easily.	17
1	8.	(24)	I occasionally neglect serious things in order to have a good time.	26
1	9.	(25)	I have never been seasick, plane sick, or car sick.	47
2	10.	(26)	Compared to your own self-respect, the respect of others means little.	49
1	11.	(27)	I'm occasionally disorganized if I am called on suddenly to make a few remarks.	25
1	12.	(28)	Radical agitators should be allowed to make public speeches.	23
2	13.	(29)	The thought of God gives me a complete sense of security.	10
1	14.	(30)	I am cautious about undertaking anything which may lead to humiliating experiences.	28
1	15.	(31)	I have occasionally doubted the reality of God.	6
1	16.	(32)	I am generally active in my everyday life.	11

2. Training. When you have completed the pretest, the instructor will provide you with information about the typical student and an opportunity to practice on the following statements. Do not continue with these materials until instructed to do so.

Coll. I		Coll. II			Difficulty Level
T	F	T	F	1.	Sports generally interest me somewhat more than very intellectual affairs. 7
T	F	T	F	2.	I like ballet performances. 12
T	F	T	F	3.	I often think for a long time about an idea that has occurred to me. 13
T	F	T	F	4.	Artistic experiences are of great importance in my life. 31
T	F	T	F	5.	Divine inspiration is an infallible source of truth. 11
T	F	T	F	6.	I like to discuss abstract questions with my friends. 14
T	F	T	F	7.	I sometimes think more about my ideas than about the routine demands of daily life. 15
T	F	T	F	8.	I have never tried to collect pictures of paintings I have liked. 15
T	F	T	F	9.	I enjoy work more than play. 17
T	F	T	F	10.	I am considered rather emotional by my friends. 19
T	F	T	F	11.	I find it rather hard to keep to a rigid routine. 19
T	F	T	F	12.	I think cremation is the best method of burial. 27
T	F	T	F	13.	I am a very adventurous person. 30
T	F	T	F	14.	I always finish one task before taking on others. 32
T	F	T	F	15.	I never complain about my sufferings and hardships. 29
T	F	T	F	16.	I prefer quiet games to extremely active ones. 44

3. Post-test. When the training is complete, answer the pretest statements again. This time, however, use the numbers in parentheses, "1" is (17); "2" is (18), etc.

Part 2: Application of the Stereotype to a Particular Man

Some men fit the stereotype of the typical man exactly, some do not fit it at all, but most fit it in some respects and differ from it in others. For example, a middle-aged social science professor completed the Protebob Inventory. These are his percentile scores compared with the norms for college men on the five basic traits that the inventory measures: Empiricism - 100 percentile; Introversion - 95 percentile; Calmness - 90 percentile; Controlled - 80 percentile; and Boldness - 75 percentile.

His score on the Empiricism Scale that measures readiness for change, interest in science, nonconformity, and independence was higher than that of any college man who took the test. Also, only one out of twenty college men had a higher introversion score. The Introversion scale measures the degree of interest in the arts and reflective thinking and the lack of interest in economic matters and material things. His scores on the other traits indicate that he sees himself as less emotional, more ambitious and orderly, and more dominating, self-confident, and energetic than the typical college man.

1. The Pretest. The answers of the professor to half of the statements below were the same as those made by the typical student; the answers to the other half were different from those of the typical student.

Mark "1" on the separate answer sheet if you think the professor answered "true" to that particular statement.
Mark "2" if you think he answered "false".

Correct		
1	33. (49)	I think I would like to decorate a room with flowers.
1	34. (50)	I prefer friends who have well developed artistic tastes.
2	35. (51)	Quite a few things make me emotional.
2	36. (52)	I believe that competitiveness is a necessary and desirable part of our economic life.
1	37. (53)	If I had the ability, I would enjoy teaching poetry at a University.
2	38. (54)	I have occasional difficulty getting the temperature of my bath the way I like it.
1	39. (55)	I become emotional fairly easily.
1	40. (56)	I occasionally neglect serious things in order to have a good time.
2	41. (57)	I have never been seasick, plane sick, or car sick.
1	42. (58)	Compared to your own self-respect, the respect of others means little.
2	43. (59)	I'm occasionally disorganized if I am called on suddenly to make a few remarks.
1	44. (60)	Radical agitators should be allowed to make public speeches.
2	45. (61)	The thought of God gives me a complete sense of security.
2	46. (62)	I am cautious about undertaking anything which may lead to humiliating experiences.
1	47. (63)	I have occasionally doubted the reality of God.
1	48. (64)	I am generally active in my everyday life.

2. Training: When you have completed the pretest, the instructor will provide you with information and the opportunity to practice on the training materials on the reverse side.

3. Post-test: When the training is completed, answer the statements above again, using the numbers in parentheses to record your answers, i.e. (49), (50), etc.

EXERCISE 2A: The Typical Psychologist

This exercise is designed to improve your understanding of the typical psychologist and to give you a measure of your improvement.

1. Pretest. The replies of several hundred male psychologists and several thousand business and professional men to the Strong Vocational Interest Blank were analyzed. The results showed that the typical psychologist liked some of the occupations, amusements, activities, and kinds of people listed in the test more than men in general. For example, while 41% of psychologists said they would like to be the "author of a novel" only 32% of men in general said they would. The results also showed that the typical psychologist liked some of the items less. For example, only 29% of psychologists said they would like to be a "sales manager" while 37% of men in general said they would.

Mark "1" on the separate answer sheet for each of the interests in the left column if you think more psychologists than men in general said they would like the interest.
Mark "2" if you think fewer psychologists liked the interest.

2. Training. STOP when you have finished the pretest. The instructor will now provide information, giving you a more accurate understanding of the typical psychologist. When instructed to do so, answer the items in the right column, circling "P" for psychologist and "M" for men in general.

3. Post. After you have heard the information, answered and corrected the right hand items, answer the items in the left-hand column again. Use the numbers in the parentheses, i.e., "1" now becomes (17), "2" becomes (18), etc.

Pretest and Post-Test Correct				Training Materials	
			Difficulty Level		Difficulty Level
2	1.	(17)	Physical Activity 42	P M A.	Literature 7
1	2.	(18)	Carelessly dressed people 56	P M B.	Teaching adults 12
				P M C.	Cashier in bank 16
1	3.	(19)	Absent-minded people 57	P M D.	Rancher 17
2	4.	(20)	Golf 68	P M E.	Chess 13
2	5.	(21)	People who make fortunes in business 38	P M F.	Physiology 19
				P M G.	Office Manager 21
2	6.	(22)	Saving Money 42	P M H.	Poet 27
2	7.	(23)	People who are natural leaders 76	P M I.	Athletic director 27
				P M J.	Developing business systems 31
2	8.	(24)	Regular hours of work 31		
1	9.	(25)	Orchestra conductor 54	P M K.	Socialists 44
2	10.	(26)	Conservative people 32	P M L.	Algebra 46
1	11.	(27)	People who talk about themselves 28	P M M.	Travel movies 53
				P M N.	Chemist 52

Pretest and Post-test Correct		Difficulty Level
1	12. (28) Writing reports	25
1	13. (29) Making a speech	22
2	14. (30) Religious people	42
1	15. (31) Physician	23
1	16. (32) Editor	28

Training Materials		Difficulty Level
P M O.	Outside work	59
P M P.	Energetic people	83

EXERCISE 2B: Application of the Stereotype to a Particular Psychologist.

1. Pretest. Morgan Johnson is a 23 year old graduate in psychology. Some of his answers to the Strong Vocational Interest Blank were typical of psychologists and some were not. After reading the following sketch of Morgan, mark "1" for the items in the left-hand column if you think Morgan said he liked the interest; "2" if you think he said he disliked it.

Morgan's parents died when he was four, and he and his younger brother were raised by permissive grandparents in Brooklyn. Of his childhood, Morgan said: "As I grew up, I always had the feeling that I was inferior to everybody else because I had no parents. In grade school, I was very loud and boisterous and made persistent attempts to dominate my peers and to excel in everything I did." Today he places emphasis on being a "well-rounded scholar." About his values, he now says:

> "I do not believe there are any determining forces in the universe that make us what we are; everybody rules his own destiny. I can think of nothing more important than being a good friend or having good friends, but I don't think it is possible to have more than a few really close ones. I place little value on material things: cars, clothes, etc."

The percent of men-in-general who expressed a liking for each interest is given after the item. "Psy" after the interest indicates that more psychologists expressed a liking for it than men-in-general. "Non" indicates that more non-psychologists than psychologists liked the interest.

2. Training. STOP when you have finished the pretest. The instructor will now provide information to increase your understanding of Morgan. When instructed to do so, answer the items in the right hand column as you think Morgan answered them by circling "L" for like or "D" for dislike.

3. Post-test. After the training, answer the items in the pretest again using the numbers in parentheses.

Pretest and Post-test Correct	Difficulty Level	Training Materials	Difficulty Level
1	33.(53)People who always agree with you (5% Non) 37	L D A.	Spendthrifts (5% Non) 15
1	34.(54)Pursuing bandits in sheriff's posse (15% Non) 57	L D B.	Sporting pages(50% Non)13
2	35.(55)Drilling in a company (18% Non) 46	L D C.	Regular hours of work (58% Non) 14
1	36.(56)Auto racer (12% Non) 60	L D D.	Driving an automobile (77% Non) 24
2	37.(57)Auctioneer (8% Non) 64	L D E.	Progressive people (85% Psy) 26
1	38.(58)Meeting new situations (82% Psy) 27	L D F.	Fortune tellers (5% Non) 37
2	39.(59)Thrifty people (74% Non) 50	L D G.	Consul (34% Psy) 40
2	40.(60)Physics (58% Psy) 68	L D H.	Operating machinery (54% Non) 42
2	41.(61)Acting as yell leader (5% Non) 47	L D I.	Snakes (3% Psy) 44
2	42.(62)Economics (61% Psy) 48	L D J.	Arithmetic (74% Psy) 45
2	43.(63)Pet monkeys (8% Psy)61	L D K.	Sick people (20% Psy) 50
2	44.(64)Algebra (57% Psy) 60	L D L.	Adjusting difficulties of others (58% Psy) 57
2	45.(65)Printer (7% Non) 76	L D M.	Auto repairman (19% Non) 65
2	46.(66)Saving money (51% Non) 17	L D N.	Contributing to charities (52% Non) 75
1	47.(67)Politician (18% Psy) 59	L D O.	Clergyman (14% Non) 74
2	48.(68)Quick-tempered people (7% Psy) 13	L D P.	Auto Salesman (13% Non) 75
1	49.(69)Poet (16% Psy) 52		
1	50.(70)Literature (57% Psy) 34		
1	51.(71)Factory worker (6% Non) 21		
1	52.(72)Philosophy (57% Psy) 39		

EXERCISE 3: Marital Stereotypes

This exercise is designed to improve your understanding of the typical happy, unhappy, and divorced man and woman. The correct answers throughout are based on an analysis of the replies of members of each of these groups to lengthy and confidential questionnaires. For example, 100 happily married, 100 unhappily married, and 100 divorced women answered the question: "Do you prefer a play to a dance?" Results:

81% of the happily married women answered "yes"
58% of the unhappily married women answered "yes"
44% of the divorced women answered "yes".

Therefore, the correct answer to the statement "Most apt to prefer a play to a dance," is "happily married women."

The exercise follows this sequence:

A. Men pretest. Answer statements 1 through 16 for men.
B. Woman Pretest. Answer the statements 33 through 44 for women.
C. Training Period. STOP when you have completed the pretests. The instructor will now provide information giving you a more accurate understanding of the typical happily married, unhappily married, and divorced woman. The statements in the training materials are indicated on the other side by "A", "B", etc.
D. Woman Post-Test. After the training, answer the statements in the woman test again, following the numbers in parentheses, i.e., "33" is (45), "34" is (46), etc.
E. Man Post-Test. Can you apply what you have learned about women to men? To find out, answer the statements in the men test again, this time following the numbers in parentheses, i.e., "1" is (17), "2" is (18), etc.

A. MEN PRETEST AND POST-TEST

Mark: "1" if you think the correct answer is "happily married men"
"2" if you think the correct answer is "unhappily married men"
"3" if you think the correct answer is "divorced men".

Correct		
2	1. (17)	Least interested in artistic activities.
2	2. (18)	Most dislikes foreigners.
3	3. (19)	Most apt to like the occupation of novelist.
1	4. (20)	Most apt to like religious people.
2	5. (21)	Slowest in making decisions.
2	6. (22)	Most dislikes modern languages.
3	7. (23)	Most often experiences feelings of loneliness.
2	8. (24)	Least often takes the lead to enliven a dull party.

Correct

2 9. (25) Least interested in the occupation of teacher.
1 10. (26) Most likely to organize a club or team.
3 11. (27) Most likely to enjoy competition.
1 12. (28) Most meticulous and methodical in work.

1 13. (29) Most likely to stress quality in his work.
3 14. (30) Most likely to enjoy taking risks.
3 15. (31) Most prefers fashionably dressed people.
1 16. (32) Most prefers to make plans with others.

B. WOMEN PRETEST AND POST-TEST:

Mark: "1" if you think the correct answer is "happily married women"
"2" if you think the correct answer is "unhappily married women"
"3" if you think the correct answer is "divorced women"

Correct

3 33. (45) Most willing to be unconventional.
3 34. (46) Most interested in being an inventor.
2 35. (47) Most often troubled by feelings of inferiority.
2 36. (48) Most apt to arrive late for work.

3 37. (49) Most prefers work that makes heavy demands.
2 38. (50) Most interested in avoiding technical responsibilities.
1 39. (51) Most apt to like religious people.
2 40. (52) Least effective in emergencies.

3 41. (53) Most ambitious.
1 42. (54) Most apt to like old people.
1 43. (55) Most conservative in social and political opinions.
1 44. (56) Most apt to like people who never drink.

C. TRAINING MATERIALS. The materials are designed to give you practice and feedback in applying the principles outlined in the training discussion. Do not answer any of these statements until told to do so. "H" stands for happy, "U" for unhappy, "D" for divorced women.

		Pretest Difficulty
H **U** D	A. Most often has spells of dizziness.	6
H U D	B. Most likes picnics and excursions.	11
H **U** D	C. Most likely to consider themselves as nervous.	13
H U **D**	D. Most interested in change and travel.	26
H U D	E. Most apt to like music.	28
H **U** D	F. Least willing to work things out for themselves.	29
H U **D**	G. Most prefers taking chances to playing it safe.	30
H U **D**	H. Most self-assertive and self reliant.	33

Pretest Difficulty

H U D I. Least methodical. 49
H U D J. Most apt to dislike working in isolation. 50
H U D K. Most dislikes quick-tempered people. 55
H U D L. Most apt to like playing chess. 67

D. WOMAN POST-TEST. After you have answered the training statements and corrected them, answer the women test again to determine whether your stereotype accuracy has improved. Use the numbers in parentheses, i.e., (45) instead of "33", etc.

E MAN POST-TEST. Try, finally, to apply what you have learned about women to men. Answer again the statements in the Man Test. This time, however, use the numbers of the statements in parentheses, i.e., (16), (17), etc., in recording your answers.

F FINAL FEEDBACK. The correct answers for both the men and women tests will be read to you at the end of the exercise.

Student · _____
Initials _____ M____F__

Stereotype Accuracy
Personal Profile Sheet

		PERCENTILE								
		0	10	20	30	50	70	80	90	100
EXERCISE 1.										
COLLEGE MAN I	PRE	8	12	12	13	14	15	15	16	16
(GENERAL)	POST	4	10	12	13	14	14	15	16	16
COLLEGE MAN II	PRE	6	8	9	9	10	11	12	13	15
(INDIVIDUAL)	POST	6	8	9	9	10	11	12	13	14
EXERCISE 2.										
PSYCHOLOGIST I	PRE	4	7	8	9	11	12	13	14	15
(GENERAL)	POST	6	8	9	10	12	13	14	15	16
PSYCHOLOGIST II	PRE	7	8	9	10	11	12	13	14	15
(INDIVIDUAL)	POST	4	9	10	11	12	13	14	15	17
EXERCISE 3.										
MARRIED MEN	PRE	3	6	7	8	10	11	12	13	14
	POST	5	8	9	10	11	12	13	14	15
MARRIED WOMEN	PRE	3	5	6	7	8	8	9	10	12
	POST	4	7	8	8	9	10	11	12	12

How accurate are your stereotypes?
The circled numbers are your scores. To compare your accuracy with that of other members of the class, note the percentile at the top under which your score falls. EXAMPLE: If your score on the first horizontal line was "14", you were between the 50th and 60th percentile, i.e., your accuracy score was slightly above the average of the class.

How much do you benefit from the training program?
To answer this question fill in the following table:

	Pretest	Post-test	Gain
College Man I	_____	_____	____
" " II	_____	_____	____
Psychologist I	_____	_____	____
" II	_____	_____	____
Married Men	_____	_____	____
Married Women	_____	_____	____

The more of the tests on which you gained, and the larger the gain, the more you learned as a result of the training sessions.

APPENDIX D

RESULTS OF ANALYSES FOR STUDY TWO: ADDITIONAL DATA

Results of the revised training program for the male sample.

Stereotype	N	Mean Pretest	Mean Posttest	Diff.	t
College Man I	56	13.59	13.27	- .32	- .89
College Man II	56	10.56	10.64	.08	.30
Psychologist I	41	10.07	11.54	1.47	2.53*
Psychologist II	37	10.86	11.49	1.43	2.80**
Marital-Men	52	9.81	11.14	1.33	3.33**
Marital-Women	52	7.02	9.29	2.27	6.88**

**p < .01
* p < .05

Results of the revised training program for the female sample.

Stereotype	N	Mean Pretest	Mean Posttest	Diff.	t
College Man I	101	13.72	13.56	- .16	- .02
College Man II	101	10.06	10.05	- .01	.00
Psychologist I	66	10.27	11.35	1.08	2.57*
Psychologist II	64	10.89	12.27	1.38	4.18**
Marital-Men	86	9.31	10.66	1.35	4.09**
Marital-Women	86	7.80	9.76	1.96	2.21*

**p < .01
* p < .05

Results of the revised training program. Additional total sample data.

Stereotype	N	Pretest Variance	Posttest Variance	Pretest S.D.	Posttest S.D.
College Man I	157	2.40	5.06	1.55	2.25
College Man II	157	2.92	2.99	1.71	1.73
Psychologist I	107	7.13	5.57	2.67	2.36
Psychologist II	101	3.92	4.62	1.98	2.15
Marital - Men	138	5.43	4.08	2.33	2.02
Marital - Women	138	42.38	2.99	6.51	1.73

Results of the revised training program. Additional male sample data.

Stereotype	N	Pretest Variance	Posttest Variance	Pretest S.D.	Posttest S.D.
College Man I	56	1.99	4.88	1.41	2.21
College Man II	56	2.07	2.50	1.44	1.58
Psychologist I	41	8.01	5.62	2.83	2.37
Psychologist II	37	5.24	5.76	2.29	2.40
Marital - Men	52	5.29	3.31	2.30	1.82
Marital - Women	52	2.82	2.72	1.68	1.65

Results of the revised training program. Additional female sample data.

Stereotype	N	Pretest Variance	Posttest Variance	Pretest S.D.	Posttest S.D.
College Man I	101	2.62	5.15	1.62	2.27
College Man II	101	3.31	3.17	1.82	1.78
Psychologist I	66	6.81	5.57	2.61	2.36
Psychologist II	64	3.24	3.80	1.80	1.95
Marital - Men	86	5.48	4.54	2.34	2.13
Marital - Women	86	65.13	3.17	8.07	1.78

The interrelation of the revised accuracy measure scores for the total group.

	College Man I			College Man II			Psychologist I			Psychologist II			Marital Men			Marital Women		
	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
College Man I																		
Pre	-	39**	-31**	-00	-06	-05	00	04	03	16	22*	07	05	-08	-12	-02	02	-08
Post		-	75**	00	08	10	16	16	03	08	26*	17	-03	04	06	05	17	07
Gain			-	02	12	12	17	14	-05	-05	10	13	-07	09	15	06	16	13
College Man II																		
Pre				-	64**	-40**	29**	20	-15	15	-04	-16	-00	-04	-03	02	10	17
Post					-	44**	29**	16	-18	18	-00	-15	-05	03	07	11	06	20*
Gain						-	-00	-05	-05	03	03	00	-06	07	12	11	-04	04
Psychologist I																		
Pre							-	50**	-59**	16	15	-00	-05	07	10	06	02	08
Post								-	38**	18	17	01	-01	03	03	-01	17	11
Gain									-	02	-03	-04	07	01	-06	-08	13	01
Psychologist II																		
Pre										-	34**	-53**	23*	07	-17	11	-06	09
Post											-	62**	26*	11	-17	-03	18	04
Gain												-	05	04	-02	-11	20	-04
Marital-Men																		
Pre													-	47**	-61**	12	-01	-08
Post														-	41**	19	-08	-05
Gain															-	05	-06	03
Marital-Women																		
Pre																-	-36**	-17
Post																	-	41**
Gain																		-

**p < .01 *p < .05 Note: All numbers are correlation coefficients to two decimal places.

The interrelation of the revised accuracy measure scores for the male sample.

	College Man I			College Man II			Psychologist I			Psychologist II			Marital Men			Marital Women		
	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
College Man I																		
Pre	-	22	-40**	05	-05	-12	12	31	25	21	23	04	-03	-35*	-23	12	02	-09
Post		-	81**	03	24	29*	17	25	08	34	17	-11	-06	-22	-11	00	19	16
Gain			-	-00	26	34*	01	-09	-20	15	-06	-18	-04	02	05	-08	18	21
College Man II																		
Pre				-	70**	-29*	-01	00	02	07	-01	-07	-16	-30*	-07	-24	17	33
Post					-	49**	17	02	-23	-12	-19	-08	-27	-18	12	-28	11	33
Gain						-	28	03	-38*	-30	-28	-02	-16	13	25	-09	06	02
Psychologist I																		
Pre							-	61**	-60**	18	34*	17	-06	-06	01	-07	19	22
Post								-	26	19	34*	16	-02	-09	-05	05	31	22
Gain									-	-01	-09	-07	04	-01	-04	15	09	-05
Psychologist II																		
Pre										-	45**	-49**	21	-12	-26	-27	-02	33
Post											-	56**	37	08	-27	08	08	09
Gain												-	20	18	-07	20	00	-18
Marital-Men																		
Pre													-	35*	-70**	13	11	-02
Post														-	42**	08	14	05
Gain															-	-07	-00	06
Marital-Women																		
Pre																-	31*	-59**
Post																	-	58**
Gain																		-

**p < .01 *p < .05 Note: All numbers are correlation coefficients to two decimal places.

The interrelation of the revised accuracy scores for the female group.

	College Man I			College Man II			Psychologist I			Psychologist II			Marital Men			Marital Women		
	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
College Man I																		
Pre	-	48**	-27*	-01	-06	-02	-05	-11	-09	14	20	05	09	04	-05	-04	01	-07
Post		-	71**	01	03	03	17	13	-05	-02	31*	29*	-01	15	16	06	16	01
Gain			-	04	06	02	22	23	02	-12	18	26*	-09	13	21	09	15	08
College Man II																		
Pre				-	61**	-45**	39**	29*	-15	18	-03	-18	05	05	-01	08	10	07
Post					-	43**	31*	23	-11	31	11	-15	04	09	05	20	08	11
Gain						-	-09	-08	03	13	15	02	-01	05	06	14	-03	06
Psychologist I																		
Pre							-	43**	-58**	15	-02	-13	-04	13	16	09	-05	-01
Post								-	44**	17	06	-08	-01	09	08	00	13	02
Gain									-	05	03	-01	08	02	-07	-11	16	04
Psychologist II																		
Pre										-	24	-57**	24	16	-10	17	-13	-04
Post											-	65**	20	14	-08	-08	20	03
Gain												-	-02	-01	01	-18	25	05
Marital-Men																		
Pre													-	52**	-56**	15	-07	-14
Post														-	41**	24*	-17	-14
Gain															-	07	-09	02
Marital-Women																		
Pre																-	-50**	-13
Post																	-	33**
Gain																		-

**p < .01 *p < .05 Note: All numbers are correlation coefficients to two decimal places.

Results of delayed impression training for all judges.

Stereotype/Subjects	Delayed Mean	N	Pretest Mean	N	Posttest Mean	N
Psychologist I						
Total	11.49	51	10.20	107	11.42	107
Males	11.05	20	10.07	41	11.54	41
Females	11.77	31	10.27	66	11.35	66
Psychologist II						
Total	12.05	55	10.88	101	11.99	101
Males	12.17	23	10.86	41	11.49	41
Females	11.97	32	10.89	64	12.27	64

Results of delayed impression training. Additional data.

		S.D.	N	Variance
Psychologist I	**Total**	2.44	51	5.95
	Men	2.31	20	5.34
	Women	2.53	31	6.40
Psychologist II	**Total**	2.36	55	5.57
	Men	2.23	23	4.97
	Women	2.48	32	6.15

Correlates of delayed impression scores.

	Psychologist I			Psychologist II		
	Total	Men	Women	Total	Men	Women
Impulsive-Controlled	.05	-.28	.21	-.14	-.05	-.20
Rational-Empirical	-.00	.05	-.03	.20	.26	.15
Introverted-Extroverted	.03	.18	.01	-.12	-.11	-.18
Cautious-Bold	-.05	-.02	-.15	.18	.24	.18
Emotional-Calm	.16	.18	.29	.10	-.04	.20
Acquiescence	-.18	-.21	-.17	-.34*	-.43*	-.27
Midterm Text	.10	-.05	.24	.44**	.55**	.36*
Midterm Class	.35*	.09	.50**	.32*	.47*	.22
Midterm Total	.25	.02	.42*	.41**	.54**	.32
Final Text	.14	.02	.25	.30*	.53*	.12
Final Class	.11	-.16	.29	.36**	.46*	.29
Final Total	.13	.07	.31	.36**	.53*	.23
Text Total	.17	-.03	.43*	.35*	.54*	.17
Class Total	.25	-.09	.48**	.36**	.49*	.26
Total Total	.23	-.06	.51**	.38**	.54*	.27
GPA	.18	.13	.26	.38**	.38	.38*
CQT-Verbal	.18	.27	.12	.27	.35	.21
CQT-Information	.23	.16	.32	.17	.25	.10
CQT-Total	.17	.24	.16	.24	.28	.20
Age	.16	.14	.32	.06	.00	.07
Attendance	-.03	.08	-.18	.29*	.30	.35*
Course Rank	-.12	-.50	.07	.19	.29	.13
Consideration	.28	.27	.20	.11	.18	.06
Responsibility	.07	.13	.03	-.04	-.05	-.03
Leadership Total	.27	.36	.17	.06	.12	.02
College Man I Pre	.12	-.23	.24	-.02	.02	-.02
Post	.31*	.26	.33	.04	.04	.06
Gain	.22	.34	.12	.06	.03	.11
College Man II Pre	.21	.30	.17	-.12	.26	-.12
Post	.30*	.29	.33	-.05	.00	-.09
Gain	.12	.08	.17	-.03	-.25	.04
Marital Men Pre	-.17	-.01	-.24	.06	.05	.07
Post	.00	.06	.02	-.08	.19	-.29
Gain	.19	.08	.27	-.15	.14	-.33
Marital Women Pre	.45**	-.02	.67**	-.01	.23	-.14
Post	.15	-.22	.43*	.09	.40	-.20
Gain	-.33*	-.20	-.46*	.10	.20	-.01
Psychologist I Delay	-	-	-	-.16	.10	-.28

**p < .01
* p < .05

Relation of training outcome to intellectual ability for the total sample.

	College Man I		College Man II		Psychologist I		Psychologist II		Marital Men		Marital Women	
	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain
Midterm-text	.02	.03	.29**	-.01	.08	-.12	.06	-.14	-.06	-.07	.19*	.29**
Midterm-class	.07	.01	.21**	-.03	.12	-.09	.14	.03	-.07	-.09	.21*	.18*
Midterm-total	.05	.02	.29**	-.02	.12	-.12	.11	-.07	-.07	-.09	.23**	.27**
Final-text	-.02	-.01	.11	-.15	.03	-.14	.19	.11	-.11	.00	.28**	.16
Final-class	.06	.05	.15	-.06	.04	-.23	.13	-.12	-.13	-.11	.20*	.20*
Final-total	.02	.02	.16	-.12	.01	-.22	.15	-.02	-.08	-.06	.26**	.22*
Text-total	.01	.02	.19*	-.09	.03	-.14	.15	-.00	-.10	-.03	.25**	.22*
Class-total	.09	.06	.18*	-.03	.06	-.19	.14	-.10	-.13	-.11	.23**	.21*
Total-total	.05	.04	.22*	-.08	.03	-.19*	.15	-.05	-.08	-.07	.26**	.26**
GPA	.11	.04	.16*	-.05	.05	-.20*	-.08	-.07	-.03	-.14	.02	.14
CQT-Verbal	.13	.10	.17*	-.05	.15	.04	-.07	-.16	-.05	.03	.12	.10
CQT-Information	.04	-.00	.14	-.06	.08	-.04	-.00	-.16	.11	.09	.01	.16
CQT-Total	.11	.11	.17*	-.05	.12	.01	-.05	-.20*	-.00	.04	.19*	.16

*p < .05
**p < .01

Relation of training outcome to intellectual ability for the male judges.

	Stereotype											
	College Man I		College Man II		Psychologist I		Psychologist II		Marital Men		Marital Women	
	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain
Midterm-text	.03	.11	13	-.17	.11	-.17	.11	-.03	-.05	-.04	.28*	.43**
Midterm-class	.05	.05	26	-.03	.34*	-.04	.12	.06	-.21	-.20	.36**	.29*
Midterm-total	.04	.09	.21	-.11	.24	-.12	.13	.01	-.14	-.14	.35*	.40**
Final-text	-.01	.02	.21	-.08	.15	-.16	.22	.14	-.05	.07	.28*	.36*
Final-class	.12	.11	.32*	.03	.06	-.21	.20	-.04	-.16	-.09	.35*	.41**
Final-total	.06	.07	.27*	-.03	.11	-.20	.21	.05	-.11	-.01	.32*	.40**
Text-total	.02	.09	.16	-.12	.12	-.17	.19	.05	-.05	.03	.31*	.41**
Class-total	.12	.13	.30*	.01	.16	-.15	.19	-.04	-.21	-.14	.41**	.41**
Total-total	.07	.12	.24	-.06	.15	-.16	.20	.01	-.13	-.05	.37**	.43**
GPA	.12	.11	.22	.08	.12	-.35*	.09	.02	-.15	.03	.15	.25
CQT-Verbal	.01	.05	.05	-.16	.18	.08	-.07	-.20	-.24	-.02	.34*	.27
CQT-Information	.06	.01	.02	-.10	.27	.12	.18	-.07	-.13	.01	.43**	.38**
CQT-Total	.12	.15	.10	-.09	.28	.12	.02	-.24	-.20	.05	.48**	.42**

* $p < .05$
** $p < .01$

Relation of training outcome to intellectual ability for the female judges.

	Stereotype											
	College Man I		College Man II		Psychologist I		Psychologist II		Marital Men		Marital Women	
	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain	Post	Gain
Midterm-text	.04	.01	.34**	.06	.06	-.11	.06	-.18	-.10	-.08	.17	.16
Midterm-class	.10	-.00	.15	-.03	-.06	-.16	.23	.06	-.04	-.01	.17	.08
Midterm-total	.08	.00	.29**	.02	.01	-.16	.16	-.09	-.08	-.06	.20	.15
Final-text	-.01	-.02	.03	-.19	-.06	-.14	.21	.11	-.18	-.05	.32**	-.02
Final-class	.04	.02	.03	-.10	.04	-.26*	.11	-.16	-.14	-.13	.15	.03
Final-total	.02	-.00	.07	-.16	-.07	-.25*	.15	-.05	-.10	-.09	.26*	.05
Text-total	.02	-.01	.18	-.07	-.05	-.14	.15	-.01	-.18	-.08	.27*	.03
Class-total	.09	.03	.08	-.05	-.02	-.25*	.15	-.11	-.13	-.09	.17	.04
Total-total	.06	.00	.17	-.08	-.07	-.24	.16	-.07	-.10	-.09	.25*	.08
GPA	.13	.02	.08	-.12	.00	-.16	.17	-.05	.00	-.26*	-.01	.03
CQT-Verbal	.19	.13	.24*	.01	.11	-.01	-.03	-.11	.04	.07	-.01	-.04
CQT-Information	.06	.00	.13	-.05	-.08	-.21	-.01	-.14	.21	.14	.05	-.09
CQT-Total	.12	.09	.16	-.04	-.01	-.11	-.02	-.12	.08	.03	.05	-.07

* $p < .05$
** $p < .01$

Relation of training outcome to personality for the total sample.

Stereotype		Personality Trait Dimension					
		Impulsive-Controlled	Rational-Empirical	Introversion-Extroversion	Cautious-Bold	Emotional-Calm	Acquiescence
College Man I	Post	.10	-.07	.05	-.03	.05	-.01
	Gain	.01	-.12	-.02	-.07	-.04	-.04
College Man II	Post	.09	.13	-.05	-.00	-.01	-.14
	Gain	.05	-.04	.04	.05	.03	-.03
Psychologist I	Post	-.02	.12	.03	-.12	-.02	-.05
	Gain	-.10	.04	-.10	-.12	-.24*	.03
Psychologist II	Post	.07	.04	-.07	-.07	.21*	-.05
	Gain	.04	-.02	-.03	-.05	.12	-.00
Marital-Men	Post	-.06	-.03	.07	-.18*	.00	.01
	Gain	-.19*	.05	.08	-.16	-.16	-.06
Marital-Women	Post	.07	.09	-.11	-.04	.07	-.10
	Gain	.08	.05	-.10	.07	-.04	-.10

*$p < .05$

Relation of training outcome to personality of the male judges.

Stereotype		Personality Trait Dimension					
		Impulsive-Controlled	Rational-Empirical	Introversion-Extroversion	Cautious-Bold	Emotional-Calm	Acquiescence
College Man I	Post	.22	-.02	.20	-.12	.00	-.00
	Gain	.11	-.07	-.01	-.00	-.20	.01
College Man II	Post	.20	.22	.09	.20	-.11	-.15
	Gain	.22	-.08	.06	.16	.03	.13
Psychologist I	Post	-.04	.12	-.02	-.16	-.13	-.04
	Gain	-.17	.16	-.19	.02	-.39*	-.04
Psychologist II	Post	-.05	-.08	.11	-.43**	.08	.11
	Gain	-.02	-.01	-.03	-.32	.23	-.06
Marital-Men	Post	-.04	-.02	.05	-.06	-.10	.17
	Gain	-.08	.04	.18	.20	.16	.04
Marital-Women	Post	.38**	.31*	-.21	.14	.09	-.21
	Gain	.20	.19	-.11	.10	-.03	-.21

* $p < .05$
** $p < .01$

Relation of training outcome to personality of the female judges.

Stereotype		Personality Trait Dimension					
		Impulsive-Controlled	Rational-Empirical	Introversion-Extroversion	Cautious-Bold	Emotional-Calm	Acquiescence
College Man I	Post	.03	-.10	.00	-.01	.12	-.02
	Gain	-.06	-.14	-.01	-.11	.09	-.07
College Man II	Post	.02	.08	-.22*	-.06	-.03	-.14
	Gain	-.04	-.03	.02	.02	.03	-.09
Psychologist I	Post	.01	.12	.05	-.10	.03	-.06
	Gain	-.06	-.05	-.07	-.18	-.16	.07
Psychologist II	Post	.18	.15	-.11	.11	.33**	-.14
	Gain	.10	.01	.06	.09	.09	.03
Marital-Men	Post	-.09	-.06	.03	-.22*	.04	-.06
	Gain	-.27*	.06	.02	-.35**	-.17	-.13
Marital-Women	Post	-.09	-.02	-.01	-.13	.09	-.04
	Gain	-.04	-.06	-.18	.07	-.07	-.02

* $p < .05$
** $p < .01$

Relation of training outcome to demographic, attitude, and leadership variables for all judges.

		Additional Measures					
Stereotype		Age	Attendance	Course Rank	Consideration	Responsibility	Leadership Total
College Man I-	Post	.10	.01	-.00	.09	-.17	-.06
	Gain	-.04	.00	-.00	.15	-.28**	-.09
College Man II-	Post	.13	.06	-.04	.07	.14	.18
	Gain	.03	.05	.06	-.05	.13	.07
Psychologist I-	Post	.13	.06	-.04	.09	.15	.20
	Gain	-.03	-.13	.09	-.08	.09	.01
Psychologist II-	Post	.18	.16	.03	-.04	-.06	-.08
	Gain	.05	.23*	-.01	.11	-.08	.02
Marital-Men-	Post	-.06	.02	-.12	-.04	-.07	-.10
	Gain	-.11	.05	-.06	.05	-.25**	-.16
Marital-Women-	Post	-.09	.05	-.03	.05	-.03	.02
	Gain	-.20*	.15	-.05	.04	-.05	-.01

* $p < .05$
** $p < .01$

Relation of training outcome to demographic, attitude, and leadership variables for male judges.

Stereotype		Age	Attendance	Course Rank	Consideration	Responsibility	Leadership Total
				Additional Measures			
College Man I-	Post	.04	.13	.13	-.09	-.03	-.11
	Gain	-.15	.12	.08	.09	-.23	-.13
College Man II-	Post	.07	.22	-.27	.03	-.00	.03
	Gain	.11	.27*	-.04	.01	.02	.03
Psychologist I-	Post	.11	.06	-.16	.00	.11	.09
	Gain	-.16	-.45**	.26	-.21	-.01	-.17
Psychologist II-	Post	.12	.20	-.05	-.15	.08	-.05
	Gain	.31	.23	-.16	.08	.05	.11
Marital-Men-	Post	-.17	-.17	-.06	.06	-.36*	-.26
	Gain	-.17	.17	.01	.15	-.36*	-.19
Marital-Women-	Post	-.34*	-.14	.00	.13	-.17	-.04
	Gain	-.35*	-.03	-.03	.23	-.28	-.04

* $p < .05$
** $p < .01$

Relation of training outcome to demographic, attitude, and leadership variables for female judges.

		Additional Measures					
Stereotype		Age	Attendance	Course Rank	Consideration	Responsibility	Leadership Total
College Man I-	Post	.16	-.05	-.14	.17	-.25*	-.06
	Gain	.06	-.07	-.15	.16	-.29*	-.10
College Man II-	Post	.11	-.02	.14	.17	.20	.29*
	Gain	-.01	-.03	.10	-.11	.21	.07
Psychologist I-	Post	.14	.09	.05	.19	.17	.15
	Gain	.02	.17	.06	.03	.28*	.14
Psychologist II-	Post	.25	.06	-.01	-.03	-.09	-.10
	Gain	-.04	.19	-.00	.07	-.11	-.03
Marital-Men-	Post	-.04	.11	-.12	-.04	.06	.02
	Gain	-.05	-.02	-.12	-.00	-.17	-.13
Marital-Women-	Post	.10	.14	-.14	-.06	.10	.03
	Gain	-.12	.26*	-.03	-.07	.14	.05

* $p < .05$

Stability of SA training outcome.

Stereotype		Mean Accuracy Total	Males	Females
College Man II	Posttest	10.26	10.64	10.05
College Man II	Retest	10.66	10.92	10.52
Psychologist I	Posttest	11.42	11.54	11.35
Psychologist I	Retest	11.37	12.13	10.96
Marital-Men	Posttest	10.84	11.14	10.66
Marital-Men	Retest	10.71	11.00	10.55
Marital-Women	Posttest	9.51	9.29	9.64
Marital-Women	Retest	9.17	9.25	9.13