

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A026 269

THE PAPER-PEOPLE ANALOG IN INVESTIGATIONS
OF INTERVIEWERS' DECISION PROCESSES

AIR FORCE INSTITUTE OF TECHNOLOGY

MARCH 1976

189066

ADA 026269

THE PAPER-PEOPLE ANALOG IN INVESTIGATIONS
OF INTERVIEWERS' DECISION PROCESSES

Charles Douglas Gorman

A Dissertation

Submitted to the Graduate School of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

March 1976

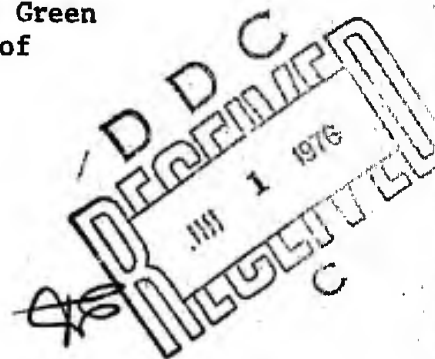
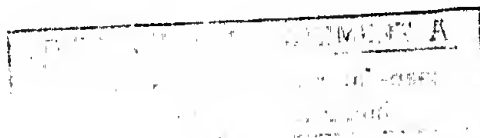
REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

Approved by Doctoral Committee

Michael E. DeBorty, Advisor
Department of Psychology

R. K. Tucker
Graduate School Representative

Kwesi H. Kipper
William O. Smith
Ruth E. Smith



Handwritten signature or initials in the top right corner.

Handwritten number '76-35'.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CI-76-35	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Paper-People Analog in Investigations of Interviewers' Decision Processes, Mar '76		5. TYPE OF REPORT & PERIOD COVERED PhD Dissertation
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Charles Douglas Gorman Captain, USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT student at Bowling Green State University Bowling Green, Ohio		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/CI Wright-Patterson AFB OH 45433		12. REPORT DATE March 1976
		13. NUMBER OF PAGES 116 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES JERRY C. HILL, Captain, USAF Director of Information, AFIT APPROVED FOR PUBLIC RELEASE AFR 190-17.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Attached		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

A common paradigm in psychology involves presenting subjects with sets of multi-cue profiles, and requiring responses based on the information in each profile. One characteristic of this paradigm is its analog nature. Subjects deal with sheets of paper supposedly representing real people; but the people represented are often completely unknown to the subjects. This paper-people method is frequently used to investigate the employment interview. But if subjects respond differently to paper people than to real people, the method may be inappropriate for studying interviewer behavior. The primary purpose of this study was to examine the responses of subjects to both real and paper people. A secondary purpose concerned investigation of the incremental validity of the interview.

Eight graduate student judges each interviewed 55 or more introductory psychology students. Based on both objective and interview information, judges predicted students' scores on a psychology examination and an attitude scale. Later, judges predicted the same students' scores on the measures, based only on objective information.

Results indicated differential responses of judges to real versus paper people. Evidence for this conclusion included higher means and variances of judgments, lower multiple correlations between cues and judgments, and increased confidence in the accuracy of judgments in the presence of real people. Judges also made differential use of the objective cues in response to real versus paper people. Correlations between judgments of real people and judgments of paper people were generally low.

The incremental validity of the interview was not impressive. Alternative procedures (such as unit weighting, linear regression, or subjective weighting) provided, on the average, more accurate predictions of subjects' examination performance than did interview judgments. For the attitude criterion, validities were slightly better for interview judgments, but were nevertheless quite low.

It was concluded that the paper-people analog suffers serious deficiencies in terms of representativeness. Differential behavior on the part of judges in response to real versus paper people makes the use of the paper-people analog inappropriate in the study of questions involving personal interaction between the judge and the judged.

Acknowledgements

I would like to take this opportunity to express my appreciation to several people who have influenced me in the pursuit of this degree. Among those I wish to thank are my parents, for raising me with an appreciation for the value of a good education; the United States Air Force, for giving me the opportunity to pursue this degree; and Drs. Guion, Schipper, Smith, and Tucker, members of my doctoral committee, for their guidance in the accomplishment of my research activities.

Dr. Michael Doherty, my doctoral committee chairman, deserves special recognition. His dedication to the professions of teaching and research is truly inspiring, and without his technical guidance, counseling, and support I would have been hard-pressed to succeed. I offer him sincere thanks and warmest personal regards.

Finally, to my wife Frannie and my son Scott, who provided the love, understanding, and encouragement (not to mention occasional typing and other logistical support) necessary to sustain me through the best of times and the worst of times, thanks seem somehow inadequate. I dedicate this paper to them with the sincere hope that by so doing they will realize how much they mean to me.

Table of Contents

	<u>Page</u>
Introduction.	1
The Paper-People Method.	2
A pilot study	4
Paper people in interview research.	5
The primary purpose of this dissertation.	5
Incremental Validity of the Employment Interview	6
Method.	8
Overview	8
Development of Criterion Measures.	8
The examination	9
The attitude scale.	10
Development of Objective Cue Profiles.	15
Selection of Judges.	15
Interview Treatment.	16
Developing the interview format	16
Preparing the judges.	16
Soliciting the interviewees	17
Collecting the data	18
Paper-People Treatment	19
Preparing materials	19
Collecting the data	20
Collection of Criterion Data	21
Results and Discussion.	21
Results Pertaining to the Judgment Phase	22

Discussion of Judgment Results	32
Higher means and variances in the interview treatment	32
Lower degrees of cognitive control in the interview treatment.	35
Heavy reliance on subjective cues	38
Increased confidence in the interview treatment	39
Low correlations between the judgments in the two treatments	40
Differential use of objective cues in the two experimental treatments.	41
Summary	42
Results Pertaining to the Prediction Phase	42
Background and rationale for the analyses	42
Results	49
Discussion of Prediction Results	55
Incremental validity of the interview	55
Comparison of judgments and judgment policies with other prediction procedures.	58
Relationship of these findings to previous research	59
Conclusion.	61
References.	65
Appendix.	68

List of Tables

Table

1	Number of subjects available for each judge at various stages of the dissertation.	23
2	Comparison of judges' behaviors in the interview and paper-people treatments, examination judgments.	24
3	Comparison of judges' behaviors in the interview and paper-people treatments, attitude scale judgments	25
4	Normalized usefulness indices for 12 cues in the interview treatment, examination judgments.	28
5	Normalized usefulness indices for 12 cues in the interview treatment, attitude scale judgments	29
6	Normalized usefulness indices for 6 cues in the paper-people treatment, examination judgments	30
7	Normalized usefulness indices for 6 cues in the paper-people treatment, attitude scale judgments.	31
8	Normalized group usefulness indices for objective and interview cues for each judge, examination and attitude scale judgments.	33
9	Correlations of usefulness indices for the six objective cues in the interview treatment with the usefulness indices for the six objective cues in the paper-people treatment, examination and attitude scale judgments.	34
10	Comparison of lens model indices for the interview and paper-people treatments, examination criterion.	50
11	Comparison of lens model indices for the interview and paper-people treatments, attitude scale criterion	51
12	Comparison of predictive validities of various prediction procedures, examination criterion.	53
13	Comparison of predictive validities of various prediction procedures, attitude scale criterion	54

List of Figures

Figure

1	Coefficient alpha as a function of number of items for the examination criterion measure	12
2	Coefficient alpha as a function of number of items for the attitude scale criterion measure.	14
3	The lens model, as applied in the analysis of the paper-people treatment.	45
4	The modified lens model, as applied in the analysis of the interview treatment	48

The employment interview enjoys widespread acceptance as a selection device in business and industry today. Ulrich and Trumbo (1965) reported survey results which indicated that as many as 99 per cent of the companies contacted made use of the interview prior to hiring. Drake, Kaplan and Stone (1973) surveyed 195 organizations, and found that 91 per cent of the companies required not one, but two interviews of a job applicant. Wright (1969) reported a survey which indicated that three-fourths of the companies that conducted on-campus interviews of prospective employees from the college ranks later required a second interview.

Not only is the interview widely used in business and industry, but the information obtained in the interview is regarded as important by employers. For example, Drake et al. (1973) found that 62 per cent of the respondents in their study felt the interviewer's impression of the applicant was the most important factor in hiring when compared with objective information like courses studied or participation in extra-curricular activities. Seventy-three per cent indicated the interviewer's opinion had more impact on the final hiring decision than the applicant's resume. Asher (1970) cited a study by Munday in which over 90 per cent of the personnel people surveyed said that of all possible sources of information, they had the most confidence in the interview.

With so many interview-based decisions affecting so many people, it is disturbing to find that there is little empirical basis for the confidence displayed in the above findings. Several reviewers (e.g., Mayfield, 1964; Ulrich and Trumbo, 1965) concluded that of all selection devices, the interview is probably the most unreliable source of information about prospective employees. In fact, some investigators

recommended that practitioners cease using the selection interview until better reliabilities and validities have been shown (Asher, 1970).

This dissertation deals with two issues. The essential focus is on research that has been done on the selection interview which has made use of a method which relies on hypothetical applicants, or simulated interviews. A secondary concern is with the incremental validity of the interview as a prediction device. Each of these issues will be briefly reviewed.

The Paper-People Method

A research paradigm common to several areas within psychology involves presenting the subject with sets of multi-cue profiles, each profile representing a person, and requiring the subject to make some response based on the information in each profile. For example, the subject may be asked to decide whether or not a company should hire the person represented by the paper profile, or to make a medical diagnosis based on the set of cues presented, or to decide whether or not he would like any of the paper people as his friends.

Various experimental treatments can be evaluated through the application of regression analysis or other statistical techniques to responses based on evaluations of paper people. For example, one might be interested in determining if the order in which favorable or unfavorable information is presented affects a subject's decision to hire a given job applicant (see, for example, Blakeney and MacNaughton, 1971), or if the relationship between a judge's perceived importance of the various cues agrees with their importance as determined mathematically (see, for example, Slovic, 1969).

One interesting characteristic of research employing the paradigm just described is its analog nature. Subjects are asked to deal with sheets of paper which are supposed to represent real people; but frequently, the people represented are completely unknown to the subject. Such a use of paper people does have its advantages. For example, if a researcher is interested in capturing the judgment policy of a judge, he does not have to find a sufficiently large number of subjects for the judge to evaluate. Rather, he can simply generate realistic cue profiles for an appropriate number of hypothetical ratees, and present these to the judge. Such a generation of hypothetical ratees also allows the researcher to manipulate such dimensions as cue values, cue variances, and relationships between cues. Further, in areas such as medical or psychiatric diagnosis, generation of hypothetical patients permits investigation of a judge's behavior in situations which might otherwise occur too rarely for reliable analysis.

While the use of paper people has its advantages, there are potential weaknesses, especially under circumstances where a judge normally has personal contact with the ratees being evaluated. In such a case, the extent to which the use of paper people could be considered, in Brunswikian terminology, a representative research design (Brunswik, 1952) is unknown. There are many unanswered questions with respect to the differences in the behavior of a judge in response to real versus paper people. For example, what impact does the presence of a real person have on a judge's use of already existing paper cues? Will the presence of the real person cause the judge merely to add new information to the paper cues, or will he alter his perception and use of those cues in light of other cues now

available as a result of personal interaction with the ratee? Will the judge behave with the same degree of consistency in the presence of real people as he does with paper people, or will he be influenced differentially from one real person to another by factors such as attractiveness or assertiveness of the ratee? Will the presence of a real person "soften" the judge, so that his ratings become more lenient than in the relatively impersonal case of evaluating paper people? Questions like these have provided one major motivation for this dissertation.

A pilot study. Only one investigation of judges' behavior in the presence of a real person on the one hand, and a paper person on the other has been made (Gorman, 1975). The author was interested in determining whether or not judges' personal knowledge of the ratee served as an additional set of cues which, when present, caused their collective rating policy to differ from the policy employed without such personal knowledge. Undergraduate psychology students rated their introductory psychology professor and an unknown, hypothetical psychology professor on their overall effectiveness, based on five-cue profiles for each. Data were analyzed within a lens model framework, and consistency indices were found to be lower under conditions where the subjects had personal knowledge of the ratee than where such personal knowledge was not present. The lower consistency indices for the personal knowledge treatments were interpreted as resulting from subjects' using cues other than those included in the experimental design (i.e., personal knowledge) in the assessment of their professor's overall effectiveness. Other results, including usefulness indices for the various cues which differed as a function of the various treatments, and values for the lens model statistic C tended to support

such an interpretation. The author tentatively concluded that personal knowledge affects rating policy.

Paper people in interview research. One primary user of paper-people methods is the group of researchers investigating various aspects of the employment interview. While a comprehensive review of interview studies using paper people would be too cumbersome here, one study will be described as a reference experiment. Many others are reviewed elsewhere (Mayfield, 1964; Ulrich and Trumbo, 1965; Wright, 1969; Clover, 1975).

Carlson (1971) pointed out that many impression formation studies used college sophomores and attempted to remedy that shortcoming by using a simulated employment interview setting. Descriptive personal history statements were combined into a nine-cue profile for each hypothetical job applicant, and life insurance agency managers were asked to rate each applicant on a seven-point scale with respect to his potential for success as a life insurance sales representative. Order of presentation of favorable and unfavorable items was varied from profile to profile, as was the amount of favorable versus unfavorable information. Findings indicated that "the effect of subjective interview information in altering valid impressions is dependent on the amount, kind, and order in which it is presented" (p. 70). Carlson goes on to make the point that "the laboratory setting or simulation experiment may appear to be very real or natural in its conception, but it does not eliminate the need for testing these findings and hypotheses in natural settings" (p. 71).

The primary purpose of this dissertation. Carlson's point is well taken, for if the behavior of subjects is dependent upon whether they are

dealing with natural or laboratory settings, with real or simulated environments, or with real or paper people, then results of studies employing only facsimiles of the real world may prove little more than interesting demonstrations from a practical viewpoint. The primary purpose of this dissertation is to examine the behavior of judges in the presence of paper people on the one hand and real people on the other. The paper person will be defined to be just that: a person represented by a set of objective cues presented to a judge on a sheet of paper. The real person will be one who is represented by the same objective cues in paper form plus a brief personal interaction with the judge in a realistic interview setting where subjective information can be obtained.

Incremental Validity of the Employment Interview

The secondary purpose of this dissertation is concerned with investigating the incremental validity of the interview. Much of the research accomplished on the validity of the interview has failed to single out the contribution of the interview to validity coefficients based on some combination of interviewer ratings and objective information such as test scores, biographical data, or the like. However, a few studies have made this relevant distinction.

Ulrich and Trumbo (1965), in their review of interview research since 1949, presented several studies which assessed the incremental validity of the interview. Results were equivocal. For example, they reported that Anderson found interviewer summary ratings to be substantially more valid than non-interview data in predicting criterion ratings, while Tupes found that test scores, as well as interviews, contributed to the validity of interviewer predictions. They also noted that Kelley and Fiske found

addition of a one-hour interview to biographical data alone did not raise the validity of interviewers' predictions. Ulrich and Trumbo conclude:

The evidence . . . supports Wagner's plea for wider use of ancillary data. For the most part, those few studies which isolated the contributions of ancillary and interview data indicated that the majority of valid variance in predictions was contributed by tests and/or credentials, rather than by the face-to-face interview. (p. 113)

A similar opinion was voiced by Mayfield (1964) after his comprehensive review of the interviewing literature:

When an . . . interviewer has tests of proven validity available, his predictions based on the interview and test scores are generally no more (and frequently less) accurate than those based on test scores alone. . . . It would seem safe to assume . . . that where tests have a reasonably high relationship with success, the use of the typical face-to-face interview is of little, if any, value. (p. 251)

A study by Grant and Bray (1969) illustrates one current approach to research on the incremental validity of interview information. After having interview reports from assessment centers coded by independent judges, the data were analyzed to determine if information from the reports contributed anything to assessment center evaluations. The results indicated that the interview's contribution was substantial. Information from the interview was shown to be predictive of the criterion "success in management." Independent interview variables "reflecting career motivation, dependency needs, work motivation, and interpersonal skills were related to individual differences in salary increases" (p. 33).

While some researchers, such as Grant and Bray, say that the interview shows promise as a significant contributor to valid predictions, the

consensus appears to question the interview's practical utility. The research to be presented in this paper is designed to provide information relevant to the question often asked in industrial settings: "Should we conduct employment interviews and use their results as one input into the hiring decision; or should we rely only on test scores and other objective information?" The interview is thus seen as serving the function of providing additional, subjective information over and above the objective information provided by test scores, application blanks and the like. The incremental validity of the interview will therefore be assessed by examining the predictive effectiveness of interviewers as a function of whether their interviewees are real (and hence, evaluated to some degree on a subjective level) or paper (and thus evaluated on a more objective level).

Method

Overview

Each of eight graduate student judges interviewed 55 or more introductory psychology students. Based on objective information and information and impressions obtained from the interview, a judge predicted each student's scores on a psychology examination and an attitude scale. Later, the judges predicted the same students' scores on the same criterion measures, based only on the objective information available on each student. Finally, actual criterion data were collected from each student. The methods used in developing materials and collecting data are presented, roughly chronologically, in the remainder of this section.

Development of Criterion Measures

Two criterion measures were used. One was an examination designed

to measure mastery of areas of introductory psychology common to all instructors' courses. This measure will be referred to simply as the examination. The second was a scale designed to measure the attitude of each student toward psychology and will be referred to as the attitude scale.

The examination. Test item manuals which accompanied several introductory texts not in use at Bowling Green State University were consulted to develop a test-item pool. Twenty items were selected for each of nine major areas within psychology (learning, sensation, perception, personality, psychopathology, personality assessment, methodology, developmental, and social) according to the criterion that no item should be text-specific. This was done to insure some degree of generality of the examination. The 180 items so selected were reviewed by three faculty members and a research associate, all of whom had had extensive experience with the introductory psychology course at Bowling Green. Items deemed inappropriate for any reason by any one reviewer were deleted from the item pool. Also deleted were those items to which the experimenter could not find the answer in the required reading areas of both introductory psychology texts used at Bowling Green. This left 52 items, which were pretested on a final examination given to 53 students completing a course in introductory psychology during the summer quarter, 1975.

Item analysis procedures similar to those suggested by Nunnally (1967) were used to select items for the examination. For each item, a part-whole correlation was computed. The items were then rank-ordered from high to low with respect to the magnitude of their part-whole correlations. Coefficient alpha (Cronbach, 1951) was then computed for

various numbers of items (i.e., the top five items, the top 10 items, etc.). Coefficient alpha (also known as the Kuder-Richardson Formula 20) is a measure of internal consistency which can be conceptualized as "an average of all the split-half coefficients that could have been obtained using all possible ways of dividing the test" (Guion, 1965, p. 44). Coefficient alpha is dependent upon, among other things, the number of items in the examination (see Figure 1). An examination consisting of the 25 items with the highest part-whole correlations was constructed since this number virtually maximized coefficient alpha. A high degree of internal consistency was desired since judges would be asked to predict total scores.

A copy of the examination is in the appendix (pp. 68-70). It had a coefficient alpha of .83, a mean of 61 per cent, and a standard deviation of 20 per cent, based on the pretest sample of 53 subjects.

The attitude scale. Twenty statements developed by Carrier (1966) to measure attitude toward psychology were administered to 76 students just completing summer quarter and summer session introductory psychology courses at Bowling Green. For statements positive toward psychology, responses of "strongly agree" received five points, responses of "moderately agree" received four points, and so on, down to one point for responses of "strongly disagree." For statements negative toward psychology, the scoring scale was reversed. The items were then subjected to an item analysis identical to the one just described.

Figure 2 shows the effect of scale length on the level of coefficient alpha for this set of statements. An attitude scale containing the 15 items with the highest part-whole correlations was constructed with a

Figure 1. Coefficient alpha as a function of number of items for the examination criterion measure.

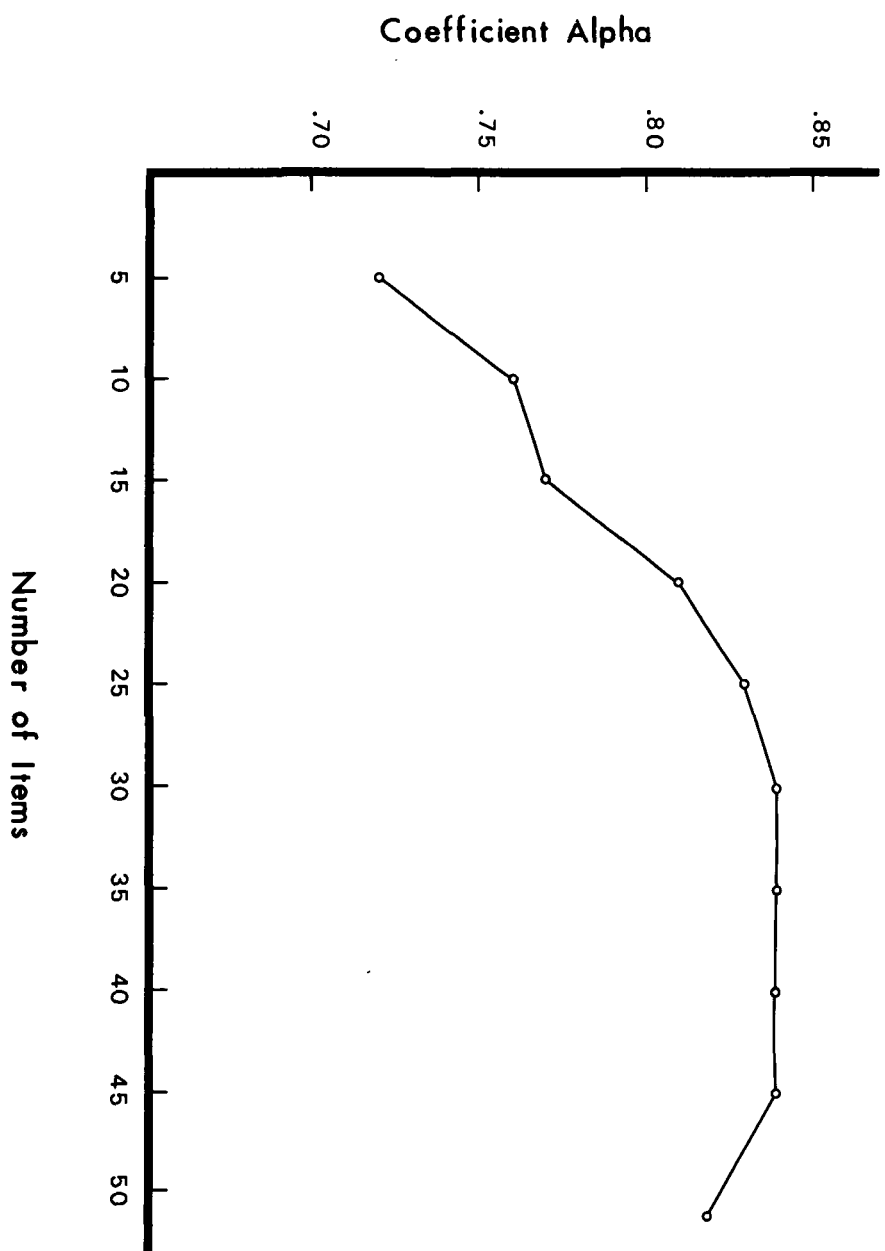
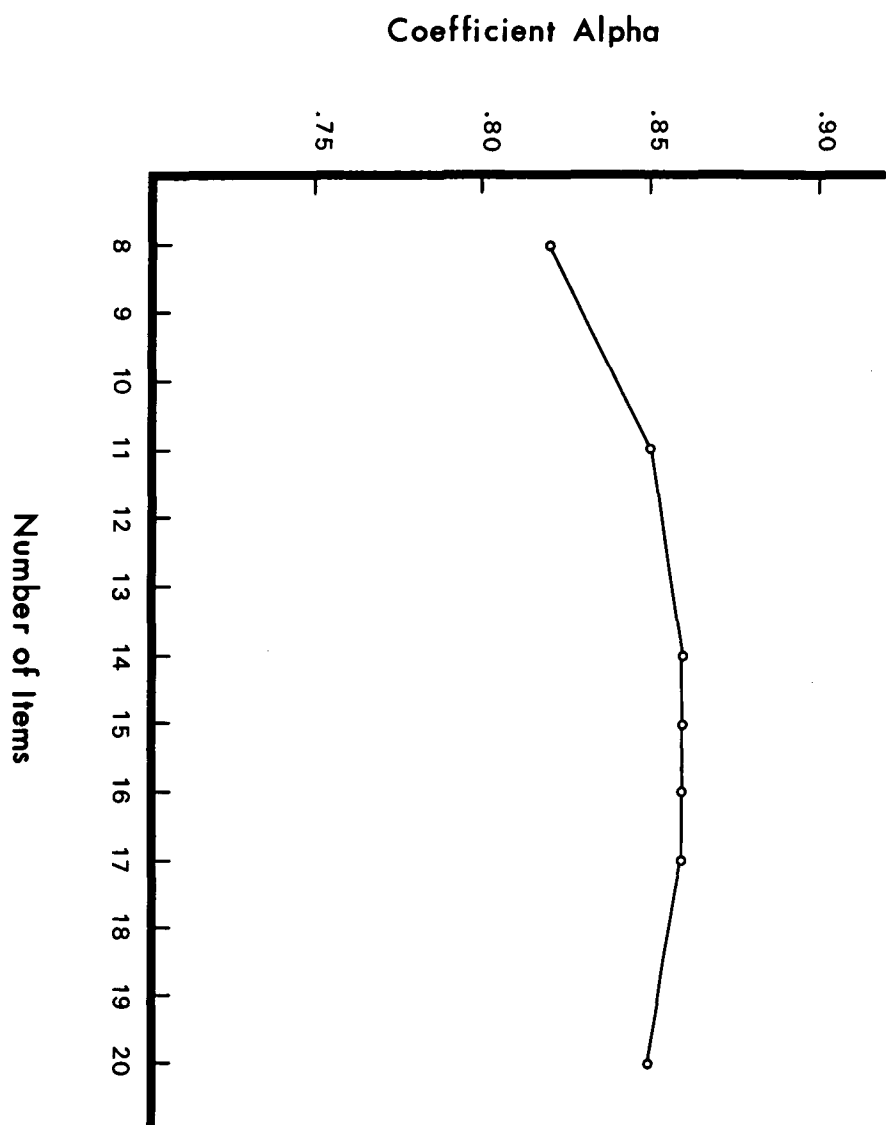


Figure 2. Coefficient alpha as a function of number of items for the attitude scale criterion measure.



coefficient alpha of .86. The scale had a mean of 56.3 and a standard deviation of 8.5, with a possible range of scores from 15 to 75 points. These statistics are based on the sample of 76 pretest subjects. A copy of the attitude scale is in the appendix (p. 71).

Development of Objective Cue Profiles

To provide judges with objective data on each interviewee cues were selected according to three criteria: a) presumptive relevance to the prediction of the criterion measures, b) similarity to the kinds of objective data generally used by industrial interviewers, and c) availability. Since the interviewees were to be sampled from the fall quarter introductory psychology classes at Bowling Green, data in the registrar's records were examined. Six cues were selected: sex, high school decile, and scores on four standard American College Testing (ACT) tests (i.e., English, mathematics, social studies, and natural sciences).

These data were obtained for 1122 students who were pre-registered for introductory psychology in the fall quarter. For 164 of these students, one or more of the cues was missing. These students were deleted from the sample. For each of the remaining 958 students, an objective cue profile like the one shown in the appendix (p. 72) was prepared. These individuals comprised the pool of potential subjects.

Selection of Judges

Eight graduate students in the industrial psychology program at Bowling Green State University agreed to participate in a study which would require them to conduct interviews of from 50 to 75 introductory psychology students. They were told that they would be asked to make predictions concerning each interviewee's future performance in and

attitude toward psychology, and would be required to spend about 12 hours of their time in the completion of their tasks. They were further informed that they would be paid \$25 for their efforts.

The eight judges were not selected in a random fashion. Rather, each had had at least one year of graduate training in industrial psychology, was essentially ignorant of the purpose of the study, and had taken at least one graduate course in interviewing. Six were male, and two female.

Interview Treatment

Developing the interview format. Judges were asked by memo (see the appendix, pp. 73-74) to suggest the kinds of information they felt might be helpful in predicting scores on the criterion measures. The suggestions were organized by the experimenter into four major areas. Interview questions were developed to obtain the various types of information requested by the judges. A tentative, structured interview format was then presented to the judges in a group meeting. Modifications of the interview format during this meeting resulted in a structured interview that required the judge to assess subjectively five characteristics of the interviewee: motivation toward psychology; expectations with respect to course content; expectations with respect to course grade; past experience with psychology; and level of general adjustment and maturity. For each of these major areas, a rating scale was given to judges to record their impressions of the interviewee. A copy of the interview form is provided in the appendix (pp. 75-78).

Preparing the judges. One week before the interviews began, each judge was given a package of materials to familiarize him further with the upcoming tasks. This package contained a cover letter from the

experimenter; general instructions concerning the interview session; a description of each objective cue; copies of the criterion measures, with statistical descriptions of each; and a copy of the structured interview form. Copies of those materials not previously presented can be found in the appendix, beginning on page 79.

Judges' questions were answered at a group meeting with the experimenter two days before the interviews began. Judges then paired off and practiced the interview so they would be familiar with it prior to interviewing actual subjects.

Soliciting the interviewees. Each judge indicated to the experimenter when he would be available to interview by providing him with 12 one-hour blocks between the hours of 9:00 AM and 9:00 PM for the second, third, and fourth days of the fall quarter. The experimenter divided each of these one-hour blocks into six 10-minute interview sessions. A computer card indicating a date and time was printed for each of these sessions. Thus, each judge had 72 potential interview sessions, and the experimenter had one schedule card for each session.

Each introductory psychology instructor was provided with a packet of schedule cards, and one handout for each student in his class. During the first class meeting of the fall quarter, the instructor explained that each student would be required to participate in two hours of research as part of the course. Then the handouts (see the appendix, p. 83) were distributed. The handout invited those students whose social security numbers appeared on an attached list to volunteer for an experiment for which they would receive one-half hour of credit. Their participation, they were told, would require them to take part in "an interview with a

graduate student interviewer." Students whose social security numbers appeared on the attached list were those for whom all objective data were available. If the student desired to participate, he selected a schedule card from those made available to him by his instructor.

Of the 958 eligible students, 581 indicated a desire to participate as interviewees by choosing a schedule card. A few extra cards over and above the originally planned 576 were provided to accomodate additional volunteers. Of these 581 volunteers, 516 actually showed up to take part in the study. When a judge had completed 50 interviews, he was permitted to stop interviewing at his discretion. All judges, however, completed at least five additional interviews, so that 461 interviews were conducted by the eight judges during the three-day period. The remaining 55 participating subjects completed an alternate task extraneous to this dissertation to receive legitimate credit for participating in an experiment.

Collecting the data. For each of the 958 potential interviewees, a manila folder containing the interviewee's objective cue profile and a blank interview form was prepared. Each folder was identified with the interviewee's social security number.

When a student arrived to be interviewed, he was greeted by the experimenter, who took his schedule card and asked for his social security number. The experimenter located the student's manila folder, asked him if he was a freshman or an upperclassman, and marked "F" or "U" on the outside of the folder as appropriate. The experimenter then gave the folder to the judge, who led the interviewee to an interview room. The interview room was approximately 2.5 by three meters, and contained one 75 by 115 centimeter table with two chairs, one on either side.

The judge was instructed to note first whether the interviewee was a freshman or an upperclassman so he could ask the appropriate questions during certain parts of the interview. After introducing himself and studying the interviewee's objective cue profile, the judge proceeded with the interview, marking the rating scales on the interview form as he came to them. On completing the interview, the judge dismissed the interviewee, who returned to the experimenter, picked up his schedule card (now stamped with the experimenter's name to allow the interviewee to obtain credit for research participation), and received a debriefing sheet (see the appendix, p. 84).

After the interviewee left the interview room, the judge considered all available information (objective cues from the profile and impressions from the interview) and predicted the interviewee's percentage score on the examination and raw score on the attitude scale at the bottom of the objective cue profile. He then recorded a previously assigned judge number and returned the entire folder to the experimenter.

Paper-People Treatment

Preparing materials. A notebook of materials for use in the next phase of the study was prepared for each graduate student judge. The first page provided the instructions (see the appendix, p. 85). Next came the same introductory materials as had been provided the judge prior to the interviewing task. That is, a description of the objective cues, a sample profile, statistical descriptions of the criterion measures, and copies of the criterion measures were provided in that order. No sample interview form was included, as it was not relevant to this task.

The objective cue profiles followed the introductory materials. The

first 15 cue profiles represented students who had not been interviewed by any judge. These 15 profiles were common to all eight notebooks. Profiles for those students previously interviewed by the judge whose notebook was being compiled came next, in random order. Next came 10 more profiles for students not interviewed, common to all notebooks. This meant that a judge would see from 80 to 89 profiles, depending on the number of students the judge had interviewed.

Forms for recording subjective importance weights, six for the paper-people treatment and 12 for the interview treatment, followed the cue profiles (see the appendix, pp. 86-87). A series of open-ended questions concerning the judge's subjective impressions of his tasks (see the appendix, pp. 88-90) completed the notebook.

Collecting the data. During the third week of the fall quarter, the experimenter gave each judge his notebook, and instructed him not to thumb through the pages before beginning the task, but rather to work straight through the notebook from front to back. Judges were not required to complete the entire notebook in one sitting, but rather were permitted to work at their own pace. They were asked to discuss the task with no one until all judges had completed their notebooks.

The basic task for each judge was to look at a profile and predict that student's performance on the criterion measures in the spaces provided on the profile. The judge then completed the subjective weighting tasks, answered the open-ended questions at the end of the notebook, and returned the notebook to the experimenter. All judges completed the task by the end of the fourth week of the fall quarter.

Collection of Criterion Data

All introductory psychology students were requested to be present for a final examination during the last week of the fall quarter, or during the final examination period immediately following the tenth week of the fall quarter. Each student present was given the examination and the attitude scale and was instructed to respond to each question or statement on a single accompanying IBM answer sheet. Of the 486 students involved in this study (461 interviewees and 25 students not interviewed), 433 completed the examination and 429 completed the attitude scale.

Results and Discussion

The results will be presented and discussed in two phases. First, those findings pertaining to the primary question of differential behavior on the part of judges in the presence of real and paper people will be covered. This will be referred to as the judgment phase of the analysis. Then, results addressing the question of the incremental validity of the interview will be presented, and their implications discussed. This will be called the prediction phase.

Throughout the remainder of this paper, the terms "cue" and "subjective cues" will be used extensively. In this study, "subjective cue" is used for convenience to represent what actually consists of, in the case of those cues obtained in the interview, subjectively scaled judgments of unspecified clusters of cues. For example, the values for the subjective cue "experience" which were entered into the analyses carried out in this study were actually judges' ratings on the scale labeled "experience" on the interview form. The actual cues, in the traditional sense of the term, which entered into these ratings may have been several in number, and are unknown.

Table 1 should prove helpful to the reader in following this organizational scheme. In it are presented the numbers of cases for each judge on which various analyses were performed. For example, judge 3 initially interviewed 58 subjects. However, he failed to record one or more interview cue values for 8 of those subjects, so only 50 subjects were available for judgment analyses involving the interview cues. Finally, three additional subjects failed to complete one or both criterion measures, reducing to 47 the number of cases available for analyses involving criterion scores.

Results Pertaining to the Judgment Phase

To determine if there were significant differences between judges in their mean judgments of the examination and attitude scale scores, a multivariate analysis of variance with the examination and attitude scale judgments as dependent measures was carried out. The 25 paper-people profiles common to all judges' notebooks served as stimulus materials for obtaining the judgments from the judges, so results of this analysis pertain only to the paper-people treatment. Those results indicated a highly significant multivariate F-ratio ($F = 4.05$; $df = 14, 382$; $p < .0001$), as well as significant F-ratios for both dependent measures. These findings are consistent with those from almost every study reporting individual judge's behavior (see, for example, Slovic and Lichtenstein, 1971, pp. 678-679). That judges did differ from one another confirmed the author's previous bias to carry out analyses on a judge by judge basis rather than to treat judges as a group.

Tables 2 and 3 present comparisons of each judge's behavior in the interview and paper-people treatments for the examination and attitude

TABLE 1
NUMBER OF SUBJECTS AVAILABLE FOR EACH JUDGE AT
VARIOUS STAGES OF THE DISSERTATION

Stage	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8	Total
Number of subjects interviewed and available for judgment analyses where only objective cue values are required	56	55	58	55	62	56	64	55	461
Number of subjects available for judgment analyses where interview cue values are required in addition to objective cue values	56	54	50	55	61	55	61	48	440
Number of subjects available for analyses involving the criterion values	50	49	47	51	58	44	59	42	400

Note: The difference between the first and last row of this table represents subjects initially interviewed who were lost because of missing interview cue values, failure to complete one or both criterion measures, or both of these reasons.

TABLE 2
COMPARISON OF JUDGES' BEHAVIORS IN THE INTERVIEW AND
PAPER-PEOPLE TREATMENTS, EXAMINATION JUDGMENTS

Judge No.	No. Cases	Treatment	Mean	Variance	R ^a	r ^b	r ^c
1	56	Interview	67.3	175.03	.93	.68	.87
		Paper-People	63.4	131.33	.93		
2	55	Interview	62.0	444.37	.91	.58	.71
		Paper-People	60.6	227.71	.93		
3	58	Interview	68.5	263.09	.91	.75	.88
		Paper-People	72.8	173.71	.86		
4	55	Interview	61.5	206.21	.75	.46	.66
		Paper-People	60.0	267.65	.86		
5	62	Interview	63.3	201.64	.94	.85	.93
		Paper-People	67.2	216.97	.96		
6	56	Interview	74.4	157.25	.88	.67	.76
		Paper-People	74.3	129.50	.95		
7	64	Interview	73.4	169.26	.86	.69	.82
		Paper-People	64.1	167.70	.95		
8	55	Interview	61.4	171.61	.79	.57	.77
		Paper-People	57.4	98.01	.93		

- a. Multiple correlation between judgments and 12 cues in the interview treatment, and judgments and six cues in the paper-people treatment. Due to missing data, the number of cases per judge for which this statistic was computed in the interview treatment is reduced to those numbers shown in the middle row of Table 1.
- b. Correlation between a judge's judgments in the interview and paper-people treatments.
- c. Correlation between a judge's predicted judgments in the interview and paper-people treatments, again with reduced numbers of cases (see note above).

TABLE 3

COMPARISON OF JUDGES' BEHAVIORS IN THE INTERVIEW AND
PAPER-PEOPLE TREATMENTS, ATTITUDE SCALE JUDGMENTS

Judge No.	No. Cases	Treatment	Mean	Variance	R ^a	r ^b	r ^c
1	56	Interview	59.3	54.02	.87	.14	.27
		Paper-People	55.4	40.83	.89		
2	55	Interview	55.3	131.79	.84	.39	.49
		Paper-People	53.5	140.90	.91		
3	58	Interview	57.2	53.44	.81	.52	.57
		Paper-People	59.2	50.13	.86		
4	55	Interview	55.6	84.46	.78	.13	.40
		Paper-People	53.5	73.10	.79		
5	62	Interview	58.6	91.01	.74	.31	.35
		Paper-People	57.4	57.15	.95		
6	56	Interview	63.9	92.35	.80	.26	.41
		Paper-People	62.1	66.91	.82		
7	64	Interview	61.6	59.91	.81	.36	.49
		Paper-People	56.8	60.22	.96		
8	55	Interview	54.9	83.17	.71	.20	.38
		Paper-People	58.5	69.39	.94		

- a. Multiple correlation between judgments and 12 cues in the interview treatment and judgments and six cues in the paper-people treatment. Due to missing data, the number of cases per judge for which this statistic was computed in the interview treatment is reduced to those numbers shown in the middle row of Table 1.
- b. Correlations between a judge's judgments in the interview and paper-people treatments.
- c. Correlation between a judge's predicted judgments in the interview and paper-people treatments, again with reduced numbers of cases (see note above).

scale judgments, respectively. Means and variances of the judgments and various correlations are presented.

The multiple correlations between the judgments and the 12 cues in the interview treatment, and the judgments and the six cues in the paper-people treatment, are presented in the columns headed "R". If one conceptualizes the judgments based on the cues as the judgment side of the lens in a traditional application of the lens model, then each of these multiple Rs would be a lens model R_s , traditionally referred to as the consistency index. However, as pointed out by Hammond, Stewart, Brehmer, and Steinman (1974), these multiple Rs, or consistency indices, should not be confused with consistency, in the sense of reliability, which results from similar responses to identical cue profiles on two or more occasions.

Hammond and Summers (1972) attempted to avoid this unfortunate semantic confusion by referring to the consistency index as an indicator of the degree of a judge's "cognitive control." In other words, the correlation between a judge's judgments and his predicted judgments (derived from a linear combination of the cues) represents the degree of control the judge is exercising over the application of his linear policy. A high multiple R in Table 2 or 3 indicates substantial cognitive control on the part of the judge. He is applying his policy in a consistent fashion across a number of different cue profiles. A lower multiple R indicates less consistent application of the policy by the judge.

The first column of bivariate correlations in Tables 2 and 3 represents agreement between a judge's judgments in the two experimental treatments. Low values in this column indicate differential behavior in

the two treatments. The correlations in the second column of bivariate correlations show the degree of agreement between a judge's predicted judgments (based on 12 cues) in the interview treatment, and his predicted judgments (based on six cues) in the paper-people treatment. Low values in this column also indicate differential behavior.

Because of moderate intercorrelations between the objective cues across the total sample (see the appendix, p. 91), usefulness indices (Darlington, 1968) were computed for the cues. According to Darlington, a cue's usefulness index can be interpreted as an indication of its unique contribution to the variance in a judge's judgments. These usefulness indices, presented in normalized form for each judge in each treatment, and for both types of judgments, can be found in Tables 4, 5, 6, and 7.

In a modification of the usefulness index concept, usefulness indices were computed for objective cues and interview cues in the interview treatment, taken as two groups. These group usefulness indices provide an insight into the extent to which judges used subjective interview information when it was available. Group usefulness indices are conceptually identical to the individual cue usefulness index, and are computed in an identical fashion. In the present instance, a multiple R^2 for all 12 cues (six objective cues and six interview cues) was computed. Then, one multiple R^2 for the six objective cues alone, and another multiple R^2 for the six interview cues alone, were computed. The group usefulness index for the six objective cues was the decrement in R^2 when the R^2 for the six objective cues alone was subtracted from the R^2 for the 12 cues taken together. The group usefulness index for the six interview cues was the decrement in R^2 when the R^2 for the six interview

TABLE 4

NORMALIZED USEFULNESS INDICES FOR 12 CUES IN THE
INTERVIEW TREATMENT, EXAMINATION JUDGMENTS

Cue	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8
Sex	.01	.04	.11	.03	.01	.19	.05	.01
Decile	.03	.04	.01	.01	.09	.12	.10	.44
ACT English	.12	.01	.07	.03	.01	.04	.15	.00
ACT Math	.13	.00	.25	.09	.02	.03	.04	.01
ACT Social Studies	.11	.08	.17	.05	.48	.19	.00	.03
ACT Natural Sciences	.07	.01	.18	.01	.01	.02	.12	.10
Class	.05	.01	.08	.01	.02	.01	.02	.08
Expectations	.00	.07	.01	.03	.07	.08	.06	.18
Grade Expected	.01	.08	.01	.08	.01	.01	.00	.01
Motivation	.18	.03	.02	.09	.01	.02	.23	.07
Experience	.03	.07	.07	.07	.03	.03	.06	.05
Adjustment and Maturity	.26	.57	.02	.51	.25	.27	.18	.03

TABLE 5

NORMALIZED USEFULNESS INDICES FOR 12 CUES IN THE
INTERVIEW TREATMENT, ATTITUDE SCALE JUDGMENTS

Cue	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8
Sex	.01	.02	.32	.02	.19	.00	.18	.01
Decile	.07	.00	.04	.00	.12	.12	.16	.01
ACT English	.00	.00	.06	.15	.00	.04	.01	.01
ACT Math	.04	.00	.21	.01	.05	.01	.06	.01
ACT Social Studies	.06	.09	.02	.03	.02	.01	.01	.01
ACT Natural Sciences	.02	.01	.01	.03	.00	.04	.06	.01
Class	.12	.00	.02	.05	.02	.00	.05	.01
Expectations	.00	.02	.12	.03	.21	.09	.02	.32
Grade Expected	.00	.19	.07	.02	.01	.04	.03	.20
Motivation	.42	.00	.02	.11	.22	.03	.26	.22
Experience	.07	.19	.09	.34	.01	.03	.04	.01
Adjustment and Maturity	.20	.45	.02	.23	.15	.58	.13	.21

TABLE 6

NORMALIZED USEFULNESS INDICES FOR 6 CUES IN THE
PAPER-PEOPLE TREATMENT, EXAMINATION JUDGMENTS

Cue	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8
Sex	.02	.05	.02	.02	.01	.00	.00	.00
Decile	.02	.40	.12	.49	.34	.19	.03	.58
ACT English	.05	.01	.17	.03	.00	.40	.80	.19
ACT Math	.32	.00	.46	.05	.01	.19	.08	.23
ACT Social Studies	.36	.53	.09	.40	.64	.15	.08	.00
ACT Natural Sciences	.22	.00	.15	.01	.00	.06	.01	.00

TABLE 7

NORMALIZED USEFULNESS INDICES FOR 6 CUES IN THE PAPER-
PEOPLE TREATMENT, ATTITUDE SCALE JUDGMENTS

Cue	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8
Sex	.00	.02	.01	.18	.00	.01	.01	.23
Decile	.03	.27	.34	.28	.16	.03	.00	.73
ACT English	.03	.02	.03	.02	.01	.05	.05	.00
ACT Math	.05	.01	.38	.02	.00	.45	.03	.03
ACT Social Studies	.82	.65	.23	.41	.82	.46	.90	.00
ACT Natural Sciences	.07	.03	.01	.08	.00	.00	.02	.00

cues was subtracted from the R^2 for the 12 cues taken together. This operation was performed twice for each judge, once for each criterion. The group usefulness indices are presented in Table 8.

To determine if the usefulness indices for the objective cues were affected by introducing a real person into the judgment situation, the usefulness indices for the six objective cues in the paper-people treatment were correlated with the usefulness indices for the same objective cues in the interview treatment. Low or negative correlations are indicative of a large impact of the person on objective cue utilization by the judge. These correlations, presented for each judge, and for each criterion, can be found in Table 9.

Discussion of Judgment Results

There is evidence in the data just presented which supports the contention that judges do respond differently to real people than to paper people regardless of criteria. This evidence is summarized in the following sections.

Higher means and variances in the interview treatment. Three judges (numbers 1, 7, and 8) for the examination and two judges (numbers 1 and 7) for the attitude scale had significantly higher ($p < .01$) means in the interview treatment than in the paper-people treatment. These higher mean judgments in the interview treatment may reflect a response bias: the physical presence of a real person may have resulted in a slightly positive set, an empathy on the part of the judges which was not present when they dealt with paper people. Whatever the reason, differences in mean judgments between treatments were generally not large.

Increased variability of the judges' responses in the interview

TABLE 8

NORMALIZED GROUP USEFULNESS INDICES FOR OBJECTIVE
AND INTERVIEW CUES FOR EACH JUDGE, EXAMINATION
AND ATTITUDE SCALE JUDGMENTS

Judge No.	Obj. Cues Exam	Int. Cues Exam	Obj. Cues Attitude	Int. Cues Attitude
1	.75	.25	.09	.91
2	.27	.73	.09	.91
3	.90	.10	.61	.39
4	.34	.66	.24	.76
5	.70	.30	.12	.88
6	.85	.15	.45	.55
7	.61	.39	.42	.58
8	.58	.42	.02	.98

TABLE 9

CORRELATIONS OF USEFULNESS INDICES FOR THE SIX OBJECTIVE
CUES IN THE INTERVIEW TREATMENT WITH THE USEFULNESS
INDICES FOR THE SIX OBJECTIVE CUES IN THE
PAPER-PEOPLE TREATMENT, EXAMINATION
AND ATTITUDE SCALE JUDGMENTS

Judge Number	Examination Judgments	Attitude Scale Judgments
1	.67	.47
2	.88	.87
3	.61	-.06
4	-.17	-.42
5	.94	-.22
6	-.42	-.44
7	.58	-.50
8	.82	.00 ^a

- a. Correlation is zero due to no variation in the usefulness indices across the six cues in the interview treatment (i.e., all six objective cues had normalized usefulness indices of .01).

treatment is reflected in the three judges (numbers 2, 3, and 8) who had significantly greater ($p < .05$) variance in their exam judgments in the interview treatment than in the paper-people treatment. Such increased variability of responses may have been a function of additional information available to the interviewers in the interview treatment which was not available in the paper-people treatment. Because of the limited amount of information available in the paper-people treatment, judges may have felt that their best estimate of the criterion would be something close to the mean. Therefore, judgments for a given judge based on the relatively limited amount of information provided by the paper-people profiles were bunched more about the mean of that judge's judgments. However, the presence of a real person (and hence, more information) in the interview treatment may have contributed to the judges' willingness to make more extreme judgments. This interpretation gains some support from data to be discussed later, which indicate greater confidence on the part of judges with respect to the accuracy of their judgments in the interview treatment than in the paper-people treatment.

Lower degrees of cognitive control in the interview treatment. Quite revealing with respect to differential behavior on the part of judges as a function of experimental treatment are the systematic trends in the multiple correlations between 12 cues and the judgment in the interview treatment, and six cues and the judgment in the paper-people treatment. Here, it is generally the case that judges behave with more cognitive control in the paper-people treatment than in the interview treatment. This finding is not inconsistent with the results of previous work (Gorman, 1975; Hoffman and Blanchard, 1961; Hayes, 1964; and Einhorn, 1971).

Introduction of a real person into the judgment situation apparently results in a less consistent application of the judges' respective linear policies perhaps because of the increased number of cues available to them.

The differences in the multiple Rs for the two treatments, which are reflected in Tables 2 and 3, are conservative estimates of the actual differences in the amount of cognitive control being exercised by judges in the two treatments. This is true because of the high cue to case ratio in the interview treatment as compared to the paper-people treatment. In the interview treatment, not only were there more cues (12 as compared to six in the paper-people treatment) than in the paper-people treatment, but there were also somewhat fewer cases. The larger cue to case ratio for the interview treatment results in spuriously higher multiple Rs than would otherwise be obtained, and yet, even under these circumstances, only one judge behaved with more control in the interview treatment than in the paper-people treatment.

To investigate the stability of the differences in the multiple Rs for the interview versus paper-people treatments reflected in Tables 2 and 3, each judge's policies for both the examination and attitude scale judgments in both treatments were cross-validated. Twelve cases from each judge's group of interviewees were randomly selected to serve as hold-out samples. Four policy equations were then developed for each judge, based on the remaining cases (43 to 52, depending on the judge), one for each set of criterion judgments in each treatment. Each judge's policies were used to obtain predicted judgments for each criterion in each treatment, based on the cue profiles for the 12 cases in the judge's

hold-out sample. The resulting 12 predicted judgments for a given criterion in a given treatment were correlated with the actual judgments of that criterion in that treatment for the 12 hold-out cases. This procedure resulted in the cross-validated multiple Rs presented in the appendix (pp. 92-93). The relationship between the multiple Rs for a judge in the interview and paper-people treatments is generally the same in the cross-validated case as in the original case presented in Tables 2 and 3.

In multiple correlation it is assumed that errors of measurement are uncorrelated. While this is probably a reasonable assumption over the long run, Guion (1965) indicates that for any given sample there is likely to be some deviation from such a zero correlation. Such chance correlations between errors of measurement influence both the individual predictors' correlations with the criterion and the intercorrelations among the predictors. Since it is these factors which determine the magnitude of the multiple correlation, correlated errors of measurement result in a larger multiple correlation than would otherwise be obtained.

When the regression equation based on a sample containing correlated errors of measurement is applied to a new sample the multiple correlation is usually diminished. This is so because correlations among errors of measurement in the second sample are bound to be different from correlations among errors of measurement in the first sample. This different set of correlations between errors in the second sample results in relationships between individual predictors and the criterion, and in intercorrelations among predictors, which are different from those

relationships and intercorrelations in the first sample. With the problem of sampling error added to this, one finds that cross-validated multiple correlations generally shrink.

Such shrinkage of multiple Rs was generally not the case in the present study. If one remembers that the predictors in this case are cues to judgment rather than measurements, the results presented in the appendix (pp. 92-93) may seem more reasonable. When predictors serve as cues to judgments, they are apparently not subject to the same problems with measurement errors as when they represent measurements in relation to a criterion. A person with a "true score" of 25 on an ACT examination may score 20 at one time, and 28 another time. Thus error of measurement is interjected into the prediction situation. But in the case of a judge presented with a set of cue profiles, a score of 20 on an ACT examination is likely to be interpreted in the same manner time after time, other things being equal. There are virtually no errors of measurement in this case, therefore shrinkage of multiple correlations upon cross-validation should be minimal. This is precisely what occurred in the present study.

Heavy reliance on subjective cues. The reason for less cognitive control in the presence of real people is unknown but it may very well be because more cues are present in the interview treatment and are used by the judges. This hypothesis receives some support upon examination of both individual and group usefulness indices for the judges in the interview treatment (see Tables 4, 5, and 9). Here it can be seen that generally heavy reliance is placed on some of the subjective cues, particularly in making the attitude judgments. It is in the attitude judgments that the greatest differences in cognitive control between the interview and

paper-people treatments, as indicated by the original multiple Rs in Table 3, exist. Further support is gained by looking at the one judge (judge 3) whose control was greater in the interview treatment than in the paper-people treatment. This occurred in his examination judgments. Judge 3's usefulness indices (both individual and group) for the examination criterion indicate very little reliance on the subjective interview cues in the interview treatment.

The data just discussed indicate differential behavior in the two experimental treatments in the sense that different information was important in determining judgments in the two treatments. Although forced to attend only to objective cues in the paper-people treatment, judges seized upon the subjective cues when they were made available, especially in making their attitude scale judgments.

Increased confidence in the interview treatment. The generally heavy reliance on subjective cues is further substantiated upon inspection of judges' responses to the open-ended questions presented them at the end of their notebooks in the paper-people treatment. Seven of the eight judges felt more confident in the accuracy of their predictions when they had benefit of the interview as a supplement to the objective cue profiles. This is perhaps understandable in the case of the attitude scale judgments where the objective cues were not particularly good predictors of the criterion. Here, satisfying the judges' desire for more information relevant to the criterion could be expected to boost their confidence in their judgmental accuracy. But in the case of the examination judgments such increased confidence has a much weaker logical basis. Objective cues predicted the criterion fairly well, as will be discussed

at a later point. It is particularly surprising that judges so familiar with literature on the pitfalls of the interview as a predictor of performance would rely so heavily on the interview and place so much confidence in their judgments based on data from it. Perhaps this is but another manifestation of the overwhelming "need" of the interviewer to "get to know" the interviewee before he is willing to predict performance with much confidence. It is a replication of the finding that confidence in judgments increases as the amount of information on which the judgments are based increases (see, for example, Oskamp, 1965; Fero, 1975).

Low correlations between the judgments in the two treatments. Further evidence that judges behave differently in the presence of real versus paper people can be found by examining the bivariate correlations in Tables 2 and 3. The first column of bivariate correlations in these two tables represents intra-judge agreement in the two experimental treatments. The second column of bivariate correlations represents agreement of a judge's policy in one treatment with his policy in the other treatment. If judges made the same predictions for the same cases in the two experimental treatments, these correlations would all be 1.0. While many of the correlations are substantial, none is perfect, and several are quite low, particularly in the case of attitude scale judgments (Table 3). Median coefficients of alienation for the first column of bivariate correlations in Tables 2 and 3 are .54 for the examination judgments and .92 for the attitude scale judgments. These values are quite large, and indicate substantial differences in the behavior of judges in the two treatments.

Differential use of objective cues in the two experimental treatments.

So far the discussion of differential behavior of judges has focused primarily on the judges' use of the subjective interview cues when they were available in the interview treatment. A related and equally important question concerns the effect of the presence of real people on the judges' use of the objective cues. Was subjective information obtained from the interview merely added to the objective information already available? Or, did introduction of real people cause a reorganization of the judges' use of the objective cues? To investigate this question, usefulness indices for the six objective cues in the interview treatment were correlated with usefulness indices for the same objective cues in the paper-people treatment. Table 9 presents those correlations.

Assuming low intercorrelations between objective and subjective cues, if subjective information was merely being added to the objective information, the correlations in Table 9 would all be 1.0. The extent to which these correlations are not 1.0 may be attributed to substantial intercorrelations between objective and subjective cues in the interview treatment, to some degree of reordering of the importance of the objective cues by the judge in the presence of real people, or to both of these conditions. For the former to occur, an objective cue must be substantially correlated with a subjective cue which is heavily used in the interview treatment. Such a high intercorrelation between the subjective and objective cues might result in much of the common variance being assigned to the subjective cue when in reality it was the objective cue upon which the judge was relying.

Inspection of the tables of usefulness indices in conjunction with

tables of intercorrelations among cues for each judge (see the appendix, pp. 94-101) indicates there are few high intercorrelations between objective cues and heavily used interview cues. If the less than perfect correlations in Table 9 are not due to problems of highly intercorrelated cues, then they are probably a result of a reordering of the importance of the objective cues as a function of experimental treatments. In this regard, some judges (i.e., judges 1 and 2) did not reorder the objective cues to any great extent with respect to either their examination or attitude scale judgments. Others, judges 3, 5, 7, and 8, were consistent in their use of cues from treatment to treatment with respect to the examination judgments but changed order or emphasis from treatment to treatment in making their attitude scale judgments. Finally, judges 4 and 6 were inconsistent in their use of the objective cues from treatment to treatment with respect to both types of judgments.

Summary. Based on the evidence discussed to this point, it seems fair to say that not only do judges make use of interview cues when they are available but those cues also affect the way in which they use the objective cues already available to them. This, coupled with less cognitive control in the interview treatment when compared to the paper-people treatment, leads to the conclusion that judges do evaluate real people differently than paper people.

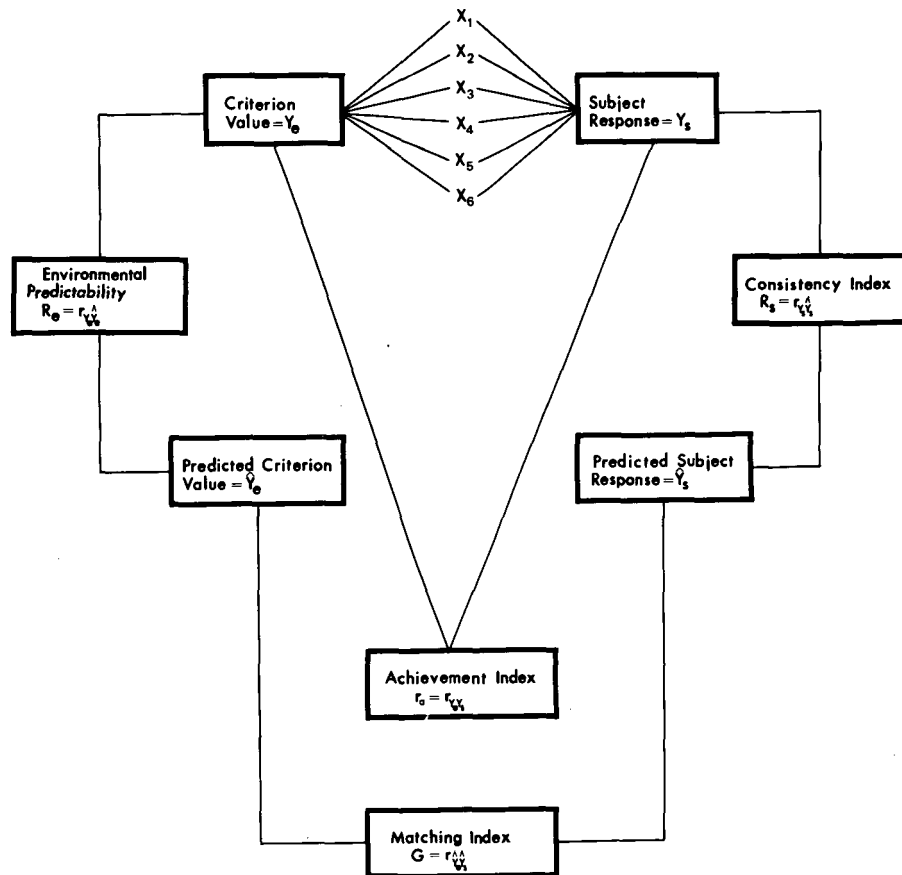
Results Pertaining to the Prediction Phase

Background and rationale for the analyses. Data from the prediction phase of the dissertation were first subjected to lens model analyses for each judge for each of the two criteria in each of the two experimental treatments. That is, a total of 32 lens model analyses were completed,

four for each judge. Figure 3 shows the framework for a lens model analysis. Application of the model uses a single judge's judgments of a relatively large number of cases based upon a set of cue values for each case. Those cue values for each case are represented in the diagram by the cues labeled $X_1, X_2, \dots, X_6$. The judge's judgment for each case is represented on the figure as Y_s . After development of the linear regression equation from the cue values presented to the judge and the judge's responses to them (which represents a model of the judge's "decision making policy"), the judge's predicted response ($\hat{Y}_s$) can be determined for any set of cue values. By following identical procedures using a set of criterion values (Y_e), one can also determine the predicted criterion value ($\hat{Y}_e$). Comparisons between the subject's responses and the criterion values, both actual and predicted, can then be made through examination of the achievement index (r_a), the matching index (G), and the environmental predictability and consistency indices (R_e and R_s , respectively).

For the paper-people treatment the traditional lens model in Figure 3 was used with one important exception. Normally, the regression equation from which the predicted criterion values ($\hat{Y}_e$) are computed is based only on those cases for which a corresponding subject response, and hence predicted subject response, exists. For example, judge 1's policy equation is based on 50 cases (see the bottom row of Table 1). Traditionally, the regression equation for the environmental portion of the lens would be based on those same 50 cases. However, in this study, the predicted criterion values ($\hat{Y}_e$) for judge 1 are the result of applying a regression equation based on all 400 cases for whom all relevant data were available

Figure 3. The lens model, as applied in the analysis of the paper-people treatment.



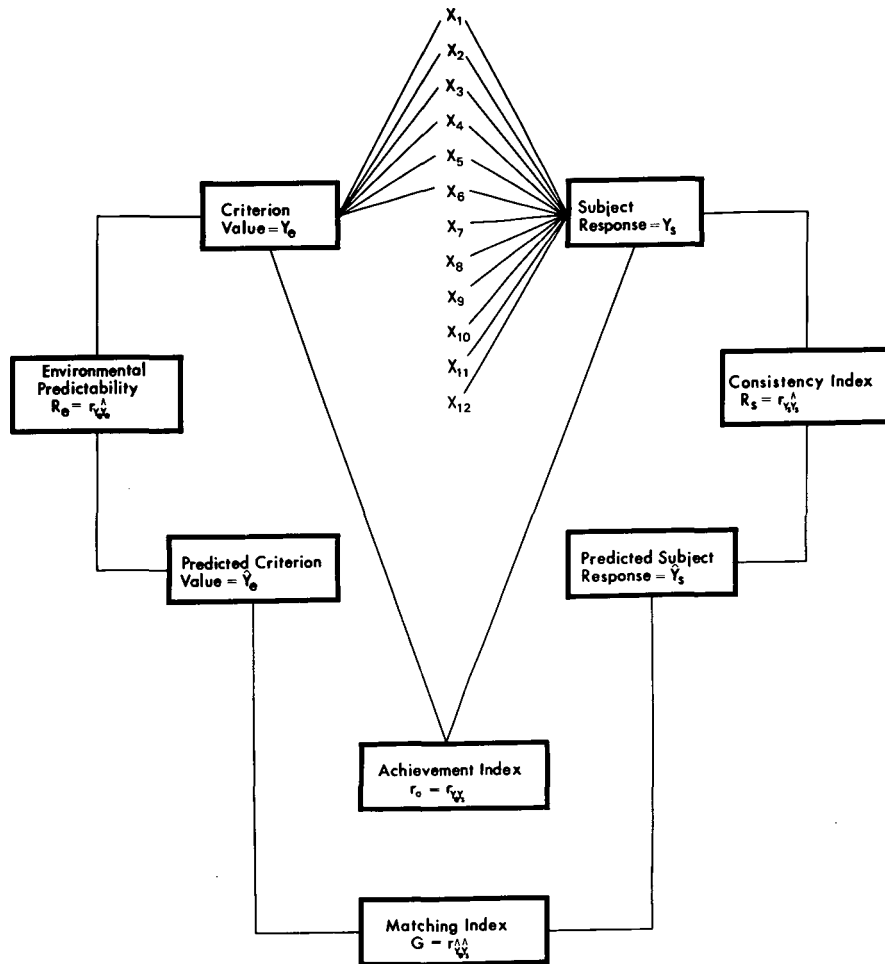
to the 50 cases' cue profiles in question.

This modification was introduced for two reasons. First, in the interest of employing a representative research design, it was felt that the environment which an interviewer is typically attempting to model is one containing a very large number of interviewees, only a small sample of whom he actually judges. In evaluating an interviewer's performance, then, he should be compared with the entire environment he is attempting to match.

In addition to this theoretical reason for basing the environmental regression equation on 400 cases, there is a more practical statistical reason mentioned in the previous discussion section. Because of a relatively large cue to case ratio when only the smaller sample of cases is considered, values of R_e , the environmental predictability index, are spuriously high. Under those circumstances, actuarial prediction of the criterion would appear in a more favorable, and perhaps misleading light due to capitalization on chance factors. This is in fact what occurred in the present study. The value for R_e for the examination criterion based on the 400 cases was .52, while the mean R_e for the eight judges when each judge's R_e was based on only those cases he evaluated was .61 (see page 102 in the appendix). For the attitude scale, the value of R_e for 400 cases was .18, while the mean R_e for the eight judges was .35.

In the analysis of the interview treatment, another modification in addition to the one just discussed was introduced. That additional modification can be seen in Figure 4. Note that there are 12 cues on the subject side of the lens and only six cues on the environmental side of the lens. Cues X_1 through X_6 represent the objective cues as in the

Figure 4. The modified lens model, as applied in the analysis of the interview treatment.



paper-people treatment, while cues X_7 through X_{12} represent the interview cues.

There are two reasons for this additional modification. First, it is the essence of this dissertation that a determination be made concerning the extent to which a judge can predict an environment through the use of objective data only on the one hand, and objective data plus subjective interview data on the other. To compare the two experimental treatments in question requires a common standard, which is a common environment. This design permits investigation of the incremental validity of the interview within the representative environment.

There is also a statistical reason for excluding the interview cues from the environmental side of the lens. Since the regression equation for the environmental side of the lens is based on 400 cases, including all 12 cues in the environmental regression equation would require including cue values for the six interview cues which had been obtained from eight different sources. Evidence already presented in the judgment phase of the dissertation indicates large differences between judges.

Results. Tables 10 and 11 contain lens model indices for each judge under each treatment for the examination and attitude scale criteria, respectively. The only lens model index not previously discussed is "C," which is simply the correlation of residuals from the subject side of the model with residuals from the environment side of the model. A residual is the difference between the predicted value (e.g., $\hat{Y}_e$) and the actual value (e.g., Y_e) for a given case. The potential importance of C will be discussed in detail in the next section.

The accuracy of predictions made by the judges under the two

TABLE 10

COMPARISON OF LENS MODEL INDICES FOR THE INTERVIEW AND
PAPER-PEOPLE TREATMENTS, EXAMINATION CRITERION

Judge No.	No. Cases ^a	Treatment	R _S	R _E	r _a	G	C
1	50	Interview	.93	.52	.50	.88	-.03
		Paper-People	.93	.52	.30	.95	-.09
2	49	Interview	.90	.52	.61	.77	.14
		Paper-People	.93	.52	.67	.93	.07
3	47	Interview	.90	.52	.42	.88	-.05
		Paper-People	.86	.52	.27	.96	-.18
4	51	Interview	.73	.52	.59	.71	.42
		Paper-People	.87	.52	.43	.92	-.01
5	58	Interview	.95	.52	.53	.88	-.07
		Paper-People	.96	.52	.60	.89	.06
6	44	Interview	.89	.52	.17	.81	-.10
		Paper-People	.95	.52	.42	.98	.00
7	59	Interview	.86	.52	.40	.79	-.05
		Paper-People	.95	.52	.41	.93	-.07
8	42	Interview	.82	.52	.44	.79	-.06
		Paper-People	.94	.52	.47	.81	.09

a. Equations from the environmental side of the lens, and any related statistics such as R_E, are based on 400 cases.

TABLE 11

COMPARISON OF LENS MODEL INDICES FOR THE INTERVIEW AND
PAPER-PEOPLE TREATMENTS, ATTITUDE SCALE CRITERION

Judge No.	No. Cases ^a	Treatment	R_s	R_e	r_a	G	C
1	50	Interview	.88	.18	.34	.23	.17
		Paper-People	.89	.18	.02	.65	-.16
2	49	Interview	.85	.18	.11	.50	-.10
		Paper-People	.90	.18	.11	.77	-.16
3	47	Interview	.80	.18	.32	.76	.26
		Paper-People	.86	.18	.04	.86	.07
4	51	Interview	.75	.18	.28	.58	.15
		Paper-People	.79	.18	-.01	.87	-.20
5	58	Interview	.75	.18	.07	.53	.22
		Paper-People	.95	.18	-.04	.71	.03
6	44	Interview	.81	.18	.27	.80	.19
		Paper-People	.81	.18	.09	.85	-.15
7	59	Interview	.81	.18	.08	.72	-.35
		Paper-People	.96	.18	.18	.83	.01
8	42	Interview	.72	.18	.36	.42	.41
		Paper-People	.93	.18	.03	.84	-.08

a. Equations from the environmental side of the lens, and any related statistics such as R_e , are based on 400 cases.

experimental treatments was compared with the accuracy of several other methods of prediction to determine the relative effectiveness of the judges and their policies in predicting the criteria. For the statistical reasons stated previously, analyses involving the interview cues were conducted only on the specific cases with whom the particular judge in question had had contact. The resulting validity coefficients for those various procedures are presented in Tables 12 and 13.

The six-cue unit weighting validity coefficients were obtained by first converting objective cue values for all 400 cases to z-scores. Then, a predicted score was obtained for each case by algebraically summing the cue values (in z-score form) for each case, the signs for the cues in the unit weighting equation being determined by the signs of the bivariate correlations between the cues and the criterion. All cues were positive except decile. The resulting 400 predicted values were correlated with the actual criterion values to obtain the 6-cue unit weighting validity coefficients of .49 for the examination criterion and .17 for the attitude scale criterion. The 12-cue unit weighting coefficients were obtained in a similar fashion, except that cue values were converted to z-scores only within each judge's group of cases. The six objective cues had the same signs, determined in the same manner, as in the six-cue unit weighting procedure. The six subjective cues were given signs as deemed appropriate by the experimenter. All were positive except for class and grade expected.

Subjective weighting validity coefficients were obtained by procedures similar to those just described. The same sets of z-scores were used for evaluating the six-cue and 12-cue subjective weighting policies

TABLE 12

COMPARISON OF PREDICTIVE VALIDITIES OF VARIOUS PREDICTION
PROCEDURES, EXAMINATION CRITERION

	6-Cue Multiple Regression	6-Cue Unit Weighting	12-Cue Unit Weighting	6-Cue Subjective Weighting	12-Cue Subjective Weighting	Paper-People Judgments	Interview Judgments	Paper-People Policy	Interview Policy
No. Cases Per Judge	400	400	Reduced ^a	400	Reduced	Reduced	Reduced	400	Reduced
Judge 1	.52	.49	.53	.50	.52	.30	.50	.50	.55
Judge 2	.52	.49	.59	.51	.62	.67	.61	.47	.63
Judge 3	.52	.49	.42	.50	.42	.27	.42	.49	.48
Judge 4	.52	.49	.53	.51	.51	.43	.59	.48	.47
Judge 5	.52	.49	.59	.49	.57	.60	.53	.47	.58
Judge 6	.52	.49	.37	.50	.35	.42	.17	.50	.23
Judge 7	.52	.49	.42	.48	.39	.41	.40	.48	.49
Judge 8	.52	.49	.48	.50	.41	.47	.44	.43	.57
Mean	.52	.49	.50	.50	.48	.46	.47	.48	.51

a. The reduced number of cases per judge refers to the bottom row of Table 1.

TABLE 13

COMPARISON OF PREDICTIVE VALIDITIES OF VARIOUS PREDICTION
PROCEDURES, ATTITUDE SCALE CRITERION

	6-Cue Multiple Regression	6-Cue Unit Weighting	12-Cue Unit Weighting	6-Cue Subjective Weighting	12-Cue Subjective Weighting	Paper-People Judgments	Interview Judgments	Paper-People Policy	Interview Policy
No. Cases Per Judge	400	400	Reduced ^a	400	Reduced	Reduced	Reduced	400	Reduced
Judge 1	.18	.17	.38	.15	.28	.02	.34	.12	.30
Judge 2	.18	.17	.27	.15	.18	.11	.11	.13	.19
Judge 3	.18	.17	.11	.17	.12	.04	.32	.16	.20
Judge 4	.18	.17	.31	.15	.17	-.01	.28	.16	.25
Judge 5	.18	.17	-.02	.16	-.01	-.04	.07	.14	-.10
Judge 6	.18	.17	.15	.16	.12	.09	.27	.15	.20
Judge 7	.18	.17	.35	.15	.31	.18	.08	.13	.33
Judge 8	.18	.17	.15	.16	.12	.03	.36	.15	.11
Mean	.18	.17	.22	.16	.16	.05	.23	.14	.19

a. The reduced number of cases per judge refers to the bottom row of Table 1.

as were used in the six-cue and 12-cue unit weighting analyses, respectively. For subjective weighting, however, each cue's z-score for a given case was multiplied by the subjective weight given to that cue by the judge in question at the end of his notebook during the paper-people treatment. Signs for the cues were the same as above.

Finally, to obtain validity coefficients for the judges' paper-people and interview policies, regression weights (including signs) for each cue, plus the regression constant, were applied to the raw cue values for each case (400 cases in the paper-people treatment, and a reduced number of cases in the interview treatment) to obtain a set of predicted values. These predicted values were then correlated with the corresponding actual criterion values to obtain the validity coefficients.

Discussion of Prediction Results

Incremental validity of the interview. The question of whether the interview provides any incremental validity over and above validity obtained in a paper-people situation is most directly addressed by examining the values for r_a in Tables 10 and 11. The achievement index in lens model terminology is simply a bivariate correlation between the judge's judgments and the actual criterion value. In other words, the value for the achievement index r_a is the validity of the judge's judgments.

The interesting difference here is in the relative magnitude of the validity coefficients with respect to the examination criterion and the attitude scale criterion. The magnitudes of the validity coefficients for the exam were generally substantial, regardless of which experimental treatment one considers. On the other hand, validity coefficients for the attitude scale were generally quite low in the interview treatment, and

were virtually zero in the paper-people treatment.

Low validity coefficients with respect to the attitude scale criterion in the paper-people treatment were not surprising since the objective cues on which judges had to base their judgments in the paper-people treatment were not good predictors of the attitude scale criterion. What was surprising was that the interview contributed so little incremental validity beyond that provided by judgments based on relatively poor objective predictors, especially since a major reason for including the attitude scale criterion in the study was to provide the interview with a chance to shine. The median coefficient of alienation for the interview treatment r_a values in Table 11 is .92. Such a large value is indicative of poor predictive efficiency in the interview treatment.

Why didn't the interview demonstrate more incremental validity over the validity of judgments based solely on paper people? To discuss this question an understanding of the lens model components which comprise the validity coefficient r_a is important.

Under normal circumstances (that is, when the lens model is applied in its traditional form), the achievement index is related to the other lens model indices by an equation called, appropriately enough, the lens model equation (Tucker, 1964):

$$r_a = GR_s R_e + C \sqrt{(1-R_e^2)(1-R_s^2)}.$$

In the present study, since the environmental side of the lens is based on 400 cases, while the subject side of the lens is based on a much smaller number of cases (from 42 to 59, depending upon the judge in question), the lens model equation does not strictly apply. Nevertheless, it should be

approximately true and still serve as a conceptual aid in understanding the results with respect to r_a .

It can be seen from the lens model equation that the achievement index, or validity of a judge's predictions, is made up of two primary components. One of these (GR_eR_s) is a function of the linear predictability of the environment (R_e), the linear predictability of the judge (R_s), and the degree to which the judge's policy matches the policy of the environment (G). The other component ($C\sqrt{(1-R_e^2)(1-R_s^2)}$) is a function of the coefficients of non-determination for the environment and the judge ($1-R_e^2$ and $1-R_s^2$) and the value of C .

In looking at the linear component of r_a , we see a constant value across treatments for R_e . The environment is equally predictable in the two situations. Consistency indices (R_s) are generally lower in the interview treatment, thus contributing to a lower value of r_a . The value of G is lower in every case for the interview treatment than for the paper-people treatment, thus resulting in lower values of r_a for the interview treatment, other things being equal. Lower values of G in the interview treatment are not unexpected, since it has already been shown that judges were attending to the subjective cues picked up in the interview.

Turning to the second component of r_a , the primary concern is with the value of C . Large positive values of C indicate high correlations between the residuals in the judge's prediction system and the residuals in the environment's prediction system. High positive values for C may be a function of a judge's accurate configural or nonlinear use of cues present in the design; his accurate linear, nonlinear, or configural use

of cues not included in the design; or some combination of these reasons. In the paper-people treatment, since there are no other cues available for the subject to use except those in the design, any high values of C must be attributed to accurate configural or nonlinear use of the cues in the design. In the interview treatment, large positive values of C may be due to any combination of the above causes. In most cases, values of C are higher in the interview treatment than in the paper-people treatment. This is not surprising, since the subjective interview cues present in the interview treatment were not included on the environmental side of the lens. Had they been included there, to the extent they were used in an accurate linear fashion by the judge, the value of G would have increased and the value of C would have decreased. What the above discussion leads to is a conclusion that judges in the interview treatment make use of cues other than those objective cues available in the paper-people treatment in a manner such that their linear consistency indices and matching indices are decreased to a greater extent than they are able to compensate for either through the accurate linear, nonlinear, or configural use of additional cues, or through the accurate use of cues already present in an other than linear fashion.

Comparison of judgments and judgment policies with other prediction procedures. How did the judges' judgments and policies compare in terms of validity with other frequently proposed or employed methods of prediction? Validity coefficients for several such methods are presented in Tables 12 and 13 for the examination and attitude scale criteria, respectively. In looking at these tables, the most striking finding is that there is generally little difference in the mean validity coefficients

for any of these procedures. Apparently, simply unit weighting the six objective cues and algebraically adding them up (in z-score form) provides essentially as valid a prediction of the criteria employed in this study as any more elaborate process.

Introduction of a human judge into the prediction system, whether it was for the purpose of aggregating objective cues (as in the paper-people judgments), making judgments concerning subjective cues and aggregating them with objective cues (as in the interview judgments), or simply determining the relative importance of the various cues (as in the subjective weighting predictions) added little to the validity of predictions obtained by the application of the strictly mathematical procedures of linear regression or unit weighting.

Relationship of these findings to previous research. The interview's failure to contribute much additional validity to judgments in the present study is in general agreement with most of the relevant interviewing literature. Comments by various reviewers cited in the introduction to this paper would have led to an expectation of the results eventually obtained, but much of the previous research employed paper people exclusively. That the present results were obtained in a more representative setting makes them even more compelling.

Some of the other results presented in Tables 12 and 13 have important relationships to previous research. Comparing the validity coefficients of paper-people judgments with the paper-people policies, and interview judgments with the interview policies has been referred to as bootstrapping (Goldberg, 1970; Dawes, 1971). The bootstrapping technique consists of using a judge to obtain a judgment policy, and then replacing

the judge with his policy. The idea is that by replacing a judge, who may be unreliable for any of a number of reasons, with a perfectly reliable policy equation, a researcher may be able to improve the validity of the judge's predictions.

The bootstrapping phenomenon did not occur in every case presented in Tables 12 and 13. Those cases where it did not occur (e.g., judge 4, examination criterion, interview treatment) can be at least partially explained by examining the corresponding value for the lens model index C . In almost every case where the bootstrapping phenomenon did not occur, the value for the corresponding C was positive and substantial. As has been noted earlier, a large positive value of C is indicative of the judge's having accurately used cues not present in the design, or cues present in the design in an other than linear fashion. Since the bootstrapping technique relies, at least in the present instance, on a linear combination of cues included in the design, judges who are accurately using cues in other than linear fashion, or are accurately using cues not in the design, would be expected to do better than their linear policy equation.

The important point to be made is that bootstrapping may be a useful technique, provided that most of the relevant cues are specified in the design, and the linear model is appropriate. However, even then, the appearance of the bootstrapping phenomenon is far from guaranteed. Dawes and Corrigan (1974) presented results which indicated that linear models whose weights had been chosen randomly (except for sign) were, on the average, superior to either judges' judgments or their optimally weighted models with respect to their accuracy of prediction. They indicate that

this is so because "linear models are robust . . . over deviations from optimal weighting"(p. 103).

Finally in examining the subjective cue weighting validity coefficients for the corresponding judge policies, it can be seen that often a judge's subjective policy provides predictions that are nearly as valid as those provided by his mathematically derived policy. While this finding is somewhat at odds with much of the published literature on the subject (see Slovic and Lichtenstein, 1971, for a brief review), the discrepancy may be a function of the approach taken to compare subjective and mathematical policies. Many of the studies reported by Slovic and Lichtenstein made direct comparisons of subjective and regression weights, and found little relationship between them. Cook and Stewart (1975) indicate that low relationships between subjective and mathematical weights need not necessarily mean poor insight into his policy on the part of a judge. They suggest what amounts to comparing the validity coefficient for a judge's subjective policy with the validity coefficient for his mathematical policy to get at the amount of insight a judge has with respect to his policy. This procedure was the one employed in this study, and the results here tend to square with those of Cook and Stewart; that is, judges seem to have reasonable insight into what they are actually doing (as described mathematically) during the judgment task.

Conclusion

For the last 50 years, much of psychology has been heavily influenced by the behavioristic philosophy so strongly espoused by Watson. Breger (1969) cites as the whole impetus for this approach to psychology the desire on the part of many in the field to make psychology a science by

making psychology look scientific. One of the easiest ways to create this scientific appearance was by making the methods employed in psychological research more stringent. In making psychology appear scientific, the subject matter of psychology was thus limited to that which could be studied through the application of rigorous methods. An initial concern with concepts such as operational definitions and laboratory control evolved to the present state of what may be referred to as a veritable plethora of analog research in psychology.

Analog research allows the psychologist to take a complex, naturally occurring situation and, through the application of various analogies, to bring that situation into the laboratory, where rigorous "scientific" methods of investigation can be employed. The problem that all too frequently occurs, and which is so deftly pointed out by Breger, is that in the process of redefining the problem in terms of an analog, the phenomenon of interest is destroyed by being reduced to triviality. By the time an analog of the problem is refined to the point that the research is methodologically beyond reproach, the analog is no longer a valid representation of the problem. This is precisely the concern that Brunswik (1952) demonstrated nearly 25 years ago.

Paradoxically, the very researchers who insist on such rigor in the doing of science abandon it entirely when discussing the results of their analog studies. The gulf between findings of an irrelevant or trivial analog study and the interesting important problem which initially prompted the research is, to paraphrase Breger, bridged by words alone. Thus, human neurotic behavior is explained by examining the results of experiments of the behavior of rats in a shock box. No bridging

experiments are performed to determine the relationship (if any) of a rat's response upon being shocked to human neurotic behavior. Instead, words like anxiety, which at the inception of the study were so rigorously and operationally defined, are now "used in the loosest of senses" (Breger, 1969, p. 39). Breger amplifies:

In fact, what the behaviorist does is to assert: "My approach is rigorous and scientific, I define my terms in a narrow operational sense; when I say anxiety, I mean the response that the animal makes in this experimental situation." Having established his claim to scientific status in this fashion, he then allows a host of meanings associated with terms such as anxiety or neurosis --meanings which his own definitions initially ruled out--to re-emerge and to bridge the gap from narrow experiment to broad application. (p. 39)

While analog research has thus become widespread in psychology, the appropriateness of the analog in any given study is seldom questioned, much less tested.

This dissertation has investigated the paper-people analog, and has found it wanting. The finding that judges behave differently depending upon whether the people they are judging are real or paper makes the use of a paper-people analog inappropriate in the study of any questions involving personal interaction of the judge and the judged.

One should not get the mistaken impression from the above discussion that analog research is of no value. Analog designs under certain circumstances can be appropriate and, when they are, valuable. For example, if one is interested in determining the effects of various numbers of cues on a judge's reliability, the use of paper-people profiles may be a most convenient, and appropriate design. Analog designs may also be employed indirectly to evaluate or investigate other psychological

phenomena. In this dissertation the paper-people analog played an indispensable role in assessing the incremental validity of the interview. That the interview contributed little incremental validity even makes the use of paper people attractive for purposes of prediction. But, when researchers employ paper people to investigate the effects of variables on an interviewer's judgments, and then draw conclusions about the interviewer's behavior in real interviews, those researchers are guilty of the behavior so eloquently criticized by Breger.

Of course, this dissertation has investigated the paper-people analog in only one situation. To generalize the results of this study much beyond the scope of this situation would be yet another instance of bridging the gulf between data and reality with words alone. And yet, results of this study provide a clear demonstration of the point which Breger was making: "Methodological restrictions should not obstruct the study of meaningful human data" (pp. 42-43). Researchers interested in the behavior of interviewers in the interview situation should study interviewers in the interview situation.

References

- Asher, J. J. Reliability of a novel format for the selection interview. Psychological Reports, 1970, 26, 451-456.
- Blakeney, R. N., & MacNaughton, J. F. Effects of temporal placement of unfavorable information on decision making during the selection interview. Journal of Applied Psychology, 1971, 55, 138-142.
- Breger, L. The ideology of behaviorism. In L. Breger (Ed.), Clinical-cognitive psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969.
- Brunswik, E. The conceptual framework of psychology. Chicago: The University of Chicago Press, 1952.
- Carlson, R. E. Effect of interview information in altering valid impressions. Journal of Applied Psychology, 1971, 55, 66-72.
- Carrier, N. A. Evaluating the introductory psychology course. Reading, Massachusetts: Addison-Wesley Publishing Company, 1966.
- Clover, W. H. Research on the selection interview since 1969. Unpublished manuscript, Bowling Green State University, 1975.
- Cook, R. L., & Stewart, T. R. A comparison of seven methods for obtaining subjective descriptions of judgmental policy. Organizational Behavior and Human Performance, 1975, 13, 31-45.
- Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.
- Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.
- Dawes, R. M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-188.

- Drake, L. R., Kaplan, H. R., & Stone, R. A. Organizational performance as a function of recruitment criteria and effectiveness. Personnel Journal, 1973, 52, 885-891.
- Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 1971, 6, 1-27.
- Fero, D. D. A lens model analysis of the effects of amount of information and mechanical decision making aid on clinical judgment and confidence. Unpublished doctoral dissertation, Bowling Green State University, 1975.
- Goldberg, L. R. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.
- Gorman, C. D. An examination of the use of hypothetical rates in judgment research. Unpublished manuscript, Bowling Green State University, 1975.
- Grant, D. L., & Bray, D. W. Contributions of the interview to assessment of management potential. Journal of Applied Psychology, 1969, 53, 24-34.
- Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.
- Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinman, D. Social judgment theory (Report No. 176). Boulder: University of Colorado, Institute of Behavioral Science, November, 1974.
- Hammond, K. R., & Summers, D. A. Cognitive control. Psychological Review, 1972, 79, 58-67.

- Hayes, J. R. Human data processing limits in decision making. In E. Bennett (Ed.), Information system science and engineering. Proceedings of the First Congress on the Information Systems Sciences. New York: McGraw-Hill, 1964.
- Hoffman, P. J., & Blanchard, W. A. A study of the effects of varying amounts of predictor information on judgment. Oregon Research Institute Research Bulletin, 1961.
- Mayfield, E. C. The selection interview--a re-evaluation of published research. Personnel Psychology, 1964, 17, 239-260.
- Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.
- Oskamp, S. Overconfidence in case-study judgments. Journal of Consulting Psychology, 1965, 29, 261-265.
- Slovic, P. Analyzing the expert judge: A descriptive study of a stock-broker's decision processes. Journal of Applied Psychology, 1969, 53, 255-263.
- Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.
- Tucker, L. R. A suggested alternative formulation in the developments by Hursh, Hammond, and Hursh, and by Hammond, Hursh, and Todd. Psychological Review, 1964, 71, 528-530.
- Ulrich, L., & Trumbo, D. The selection interview since 1949. Psychological Bulletin, 1965, 63, 100-116.
- Wright, O. R. Summary of research on the selection interview since 1964. Personnel Psychology, 1969, 22, 391-413.

APPENDIX

Instructions

Following these instructions you will find 25 multiple-choice questions about psychology. Answer each question by marking the appropriate response on the IBM answer sheet. Be sure to code your social security number in the blocks labeled "IDENTIFICATION NUMBER" on the IBM answer sheet. Use pencil only. Erase any errors or stray marks thoroughly.

Questions

1. If a child is petted and attended to every time he has a temper tantrum, a learning theorist would expect
 - a) a high rate of temper tantrums
 - b) a low rate of temper tantrums
 - c) the parents to extinguish their petting behavior
 - d) the parents to generalize their attending behaviors
2. In the above example, the attention and petting would be called
 - a) CS
 - b) discriminative stimulus
 - c) reinforcement
 - d) other-actualizing
3. Which of these statements would be least likely made by a scientific psychologist? Joe's behavior is peculiar; therefore,
 - a) Joe may be brain-damaged
 - b) Joe may be a neurotic
 - c) Joe may have learned inappropriate ways of responding
 - d) Joe may have been subjected to severe situational stress
4. You say "Hi, Joe," to someone who, on second glance, turns out to be a complete stranger. This is an example of
 - a) response generalization
 - b) serial position effect
 - c) competing response
 - d) avoidance conditioning
 - e) stimulus generalization
5. You find upon returning home from work that Bosco, your new cocker spaniel puppy, has wet the carpet. You decide to punish him now for his breach of etiquette. You do this time after time. Bosco never becomes housebroken. Which of the following conditions of learning has not been met?
 - a) reinforcement
 - b) repetition
 - c) set
 - d) closeness in time of response and its consequences

Mark questions 6 and 7 in accordance with the following key:

- (a) The problem as stated is capable of empirical solution, or the problem requires some reformulation in order to be tested
 - (b) The problem is by its very nature nonempirical and thus incapable of solution on the basis of factual evidence.
6. Are amoebas capable of learning to move toward a source of light, although they have no natural tendencies to do this?
 7. Is ability to memorize correlated with social adaptability?
 8. Mr. Smith's work suffers because he is constantly thinking about an unpleasant experience which resulted from a foolish but not malicious blunder he made. This is an example of
 - a) phobia
 - b) regression
 - c) delusion
 - d) obsession
 - e) repression
 9. Which of the following most readily lends itself to experimental investigation?
 - a) the effect of lighting on efficiency
 - b) the meaning of dreams
 - c) the effect of home environment on mental illness
 - d) the behavior of the queen bee
 - e) public opinion

10. Billy, aged 6, was full after finishing an enormous Fourth of July dinner including his favorite dessert, deep-dish apple pie. However, when he heard the familiar tinkle of the ice cream man's warning bell as his truck turned the corner, Billy ran to his Dad crying, "May I have an ice-cream sandwich?" This is an example of which of the following?
 - a) cues in the environment which can motivate behavior
 - b) secondary reinforcement
 - c) biological drives which activate behavior are not stable
 - d) regression
 - e) irrational drives
11. In piecework, a person is paid for a certain amount of work completed. A psychologist interested in operant conditioning would say that piece work resembles a
 - a) variable-ratio schedule
 - b) fixed-ratio schedule
 - c) variable-interval schedule
 - d) fixed-interval schedule
12. On the Stanford-Binet intelligence test, older children usually have a higher _____ than younger children.
 - a) MA
 - b) IQ
 - c) both a and b
 - d) neither a nor b
13. Compared to a neurotic, a psychotic is generally
 - a) less in touch with reality
 - b) more anxious
 - c) better able to cope with the environment
 - d) physically as well as psychically ill
14. A schizophrenic
 - a) has several personalities that alternately express themselves
 - b) develops hysterical symptoms
 - c) is usually withdrawn and experiences disturbances of thought
 - d) has several nonintegrated personalities that express themselves simultaneously
15. Partial as compared to continuous reinforcement produces
 - a) greater rate of learning
 - b) greater resistance to extinction
 - c) both a and b
 - d) none of the above
16. For which treatment of phobias is there scientific evidence of success?
 - a) desensitization
 - b) psychoanalysis
 - c) group therapy
 - d) client-centered therapy
17. If a previously neutral stimulus acquires the property of being able to elicit a response, and this acquisition is due to a series of trials in which the neutral stimulus was paired with a stimulus which already elicited that response, we would call this
 - a) response incrementation
 - b) classical conditioning
 - c) instrumental conditioning
 - d) generalization
 - e) discrimination
18. The tendency of the body to maintain constant internal conditions is called
 - a) operant conditioning
 - b) self-actualization
 - c) a drive state
 - d) homeostasis
19. Which characteristic of Christ makes Him an attractive person for the paranoid schizophrenic to identify with?
 - a) he was persecuted
 - b) he projected his sins onto others
 - c) he was furtively aggressive
 - d) none of these
20. The retinal region characterized by the best visual acuity is the
 - a) fovea
 - b) cornea
 - c) lens
 - d) choroid

21. Current evidence suggests that the causes of psychoses are mostly
 - a) early experience
 - b) interpersonal
 - c) unknown
 - d) hormonal dysfunctions
22. Lasting failure to cope with situational anxiety is equivalent to a
 - a) personality disorder
 - b) psychosis
 - c) conduct disorder
 - d) neurosis
23. A patient with this disorder may maintain a single body position for minutes or hours.
 - a) simple schizophrenia
 - b) catatonic schizophrenia
 - c) hebephrenic schizophrenia
 - d) paranoid schizophrenia
24. Entertainers have taught animals to perform remarkable tricks in a very short period of time through the use of
 - a) stimulus generalization
 - b) successive approximations, or shaping
 - c) classical conditioning
 - d) intermittent reinforcement

Question 25 concerns the following: An experiment was designed to determine whether noise would affect the amount of food rats would eat in a given period. Two groups of twenty male rats, matched according to age, weight, and health, were chosen. For four weeks one group of rats was subjected to loud noise while feeding and the other group was fed under regular conditions. The amount of food given each rat was carefully weighed before each feeding and the amount uneaten was weighed after each feeding.

25. The measure taken by the experimenter which reflected the behavior of interest was:
 - a) the amount of time the rats took to eat
 - b) the loud noise
 - c) the amount of food eaten
 - d) the emotionality of the rats
 - e) the weight of the rats

Instructions

Below you will find 15 statements about psychology. Please indicate the extent to which you agree or disagree with each statement by marking the appropriate response on the IBM answer sheet. Use the following key:

A=strongly agree; B=moderately agree; C=neither agree nor disagree; D=moderately disagree;

E=strongly disagree

Be sure to code your social security number in the blocks labeled "IDENTIFICATION NUMBER" on the IBM answer sheet. Use pencil only. Erase any errors or stray marks thoroughly.

Statements

1. Psychology is just common sense dressed up in technical language.
2. In psychology there is too much concern with theory and too little concern with actual studies of people's behavior.
3. Psychology may sound all right in the classroom, but it cannot really be applied to problems of everyday life.
4. More money should be spent on psychological research.
5. It is impossible for psychologists to predict the behavior of persons because the causes of human behavior are too complex.
6. Psychology is a valuable tool to aid the progress of modern civilization.
7. The good that psychology accomplishes more than makes up for its mistakes.
8. Psychologists may talk a lot about the scientific method, but psychology by its very nature can never be a science.
9. Psychology may contribute to the understanding of the unique individual, but it can never help us to understand what people in general are like.
10. Psychologists are too concerned with experimental evidence for their theories and not sufficiently concerned with whether these theories make good common sense.
11. I would be pleased to have a child of mine take up psychology as a life's work.
12. Psychological "explanations" of human behavior are just matters of opinion.
13. Psychologists spend too much time collecting "facts" and too little time developing a theory which would really help us understand people.
14. Psychological experiments are of negligible use, since the results cannot be generalized beyond the particular group of people used in the study.
15. Psychological theories about the elimination of social problems are completely unrealistic; they just do not face the facts about human nature.

72
Soc. Sec. No.: XXX XX XXXX

Sex: M

High School Decile: 3

ACT Standard Scores

English: 20

Mathematics: 16

Social Studies: 17

Natural Sciences: 23

I predict that this student will answer _____ percent of the questions correctly on the standardized introductory psychology test.

I predict that this student will make a raw score of _____ on the "Attitude Toward Psychology" scale.

Judge No.: _____

24 July 1975

Subject: Development of an interview format

To: Judges for Doug Gorman's dissertation

From: Doug Gorman

During the first few days of the fall quarter, you will be interviewing 50 to 60 introductory psychology students. The purpose of these interviews will be to obtain information that will aid you in predicting each student's scores on two criterion measures which will be administered at the end of the fall quarter. Those two criterion measures are:

1. an attitude toward psychology scale, and
2. a standardized introductory psychology multiple choice test covering several of the major areas of psychology.

As a first step in the development of a structured interview format, I would like for you to provide me with your ideas concerning the kinds of information you feel might be helpful in predicting the above criteria. Examples of the kinds of information you might want to consider include:

1. Method of instruction in the student's psychology class (personalized or lecture)
2. Size of the student's class (large or small)
3. Qualifications of instructor (PhD or graduate student)
4. Membership or non-membership in Psi Chi
5. Taking other psychology courses (yes or no)
6. Expectations of student with respect to the psychology course

Let me inform you at this time that, in addition to the information obtained from the interview, you will be provided with the following six cues for each interviewee:

1. Sex
2. High school decile (tenth of class in which the student graduated)
3. ACT English score
4. ACT mathematics score
5. ACT social studies score
6. ACT natural sciences score

Please put a list of your ideas in my mailbox as soon as possible, but not later than 8 August 1975. Also, please

check as many of the potential meeting times below as you would be able to attend, and return this memo with your list of ideas. I will try to select one meeting time that will suit everyone, and will let you know.

_____ 18 August, 10-12

_____ 18 August, 2-4

_____ 19 August, 10-12

_____ 19 August, 2-4

_____ 20 August, 10-12

_____ 20 August, 2-4

_____ 21 August, 10-12

_____ 21 August, 2-4

_____ 22 August, 10-12

_____ 22 August, 2-4

Thanks a lot for your help.

Structured Interview Format

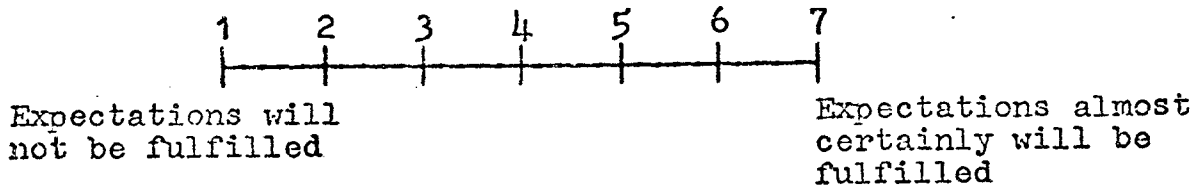
I. Introduction

- A. "Hello. My name is _____. " (Allow interviewee to introduce himself. If he does not, ask him his name, and ask him to sit down).
- B. "I'd like to get some information from you. Most of it will have to do directly or indirectly with psychology and the introductory psychology course you are enrolled in this quarter. Your responses to these questions will in no way affect your grade in the course."

II. Expectations

- A. "I'd like to begin with a brief description of what the introductory psychology course you are taking this quarter will be like. Psychology 201 will be a generally scientifically oriented course dealing with various aspects of behavior. It will familiarize you with several areas in psychology, including learning, sensation, perception, and personality and abnormal behavior. It will not be a self-help course that will teach you how to analyze yourself or other people."

- 1. Is this the kind of course you expected when you registered for Psychology 201?
 - a. If no, what did you expect?
- 2. Do you think you will like a course like the one I just described?
 - a. If yes, why?
 - b. If no, why not?

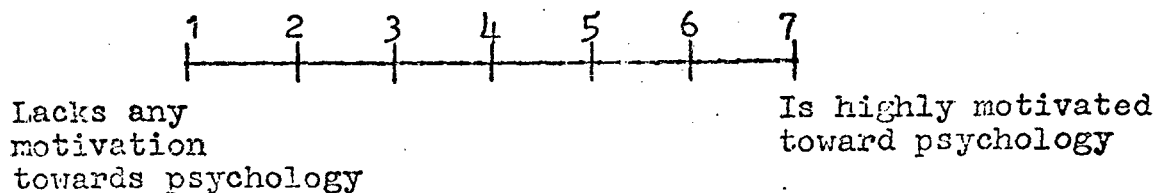


- 3. What grade do you honestly expect to get in this course?



III. Motivation

- A. How much time per week do you anticipate spending on this course?
- B. What kind of study patterns will you employ?
- C. Do you plan to major in psychology?
 - 1. If yes, why?
 - 2. If no, what do you plan to major in?
- D. Is this course an elective or a requirement for your academic program?
 - 1. If an elective, why did you take it?
 - 2. If a requirement, would you have taken this course if you didn't have to?
 - a. If yes, why?
 - b. If no, why not?
- E. Do you read "Psychology Today" magazine?
- F. Do you read, or have you read, any other psychology-oriented books or publications?

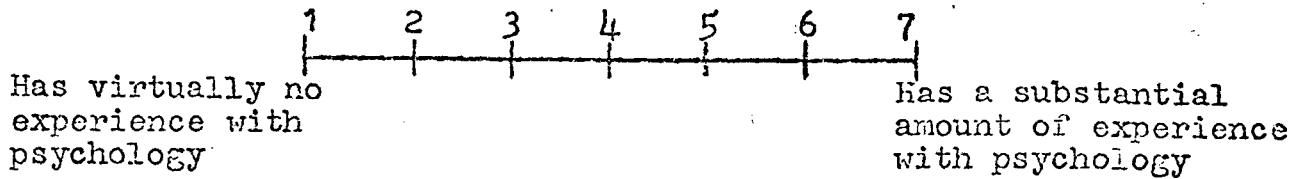


IV. Experience

- A. Did you have a high school psychology course?
 - 1. If yes,
 - a. What was your grade in the course?
 - b. What kinds of things were covered in the course?
- B. Did your high school have a school psychologist or social worker?
 - 1. If yes, did you ever have any contact with him?

C. Have you ever known any (other) professional psychologists?

1. If yes, tell me about them.



V. General Adjustment and Maturity

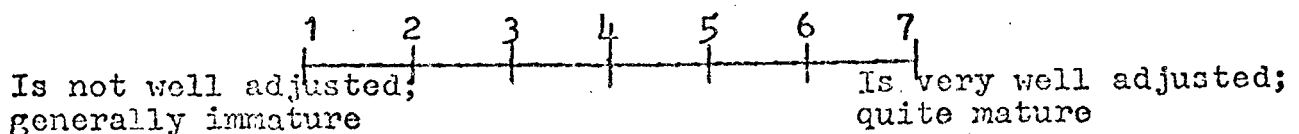
A. If interviewee is a freshman, ask these questions.

1. Are you beginning to feel more comfortable in the "university environment"?
2. How well do you think you will be able to adjust to "college life"?
3. Do you plan to participate in any extracurricular activities at the university?
 - a. If yes, which ones?

B. If interviewee is a sophomore or higher, ask these questions.

1. Do you feel comfortable in the "university environment"?
2. How well do you feel you have adjusted to "college life"?
3. Do you participate in any extracurricular activities at the university?
 - a. If yes, which ones?

C. Before rating the individual on this dimension, take into account your personal contact with him throughout the interview. In addition to his answers to the questions in this section of the interview, assess his general adjustment and maturity by considering how the individual came across as a person. Was he reasonably self-confident and at ease? Is he mature enough to handle the social and academic pressures of "college life"?



VI. Conclusion

- A. "That concludes the interview. Thank you very much for coming in. If you'll stop by the table in the lobby, Doug Gorman will stamp your experiment card and give you a sheet of paper which will explain what this experiment is all about. Thanks again."
- B. After the interviewee leaves, consider all material (both from the interview and from the profile sheet), and make your predictions at the bottom of the profile sheet.
- C. Return the complete folder, including the interview form with your ratings, and profile sheet with your predictions to Doug at the desk in the lobby.

17 September 1975

Subject: Instructions and background information for
conducting interviews

To: Judges for Doug Gorman's dissertation

From: Doug Gorman

This package of materials is divided into four parts. The first section provides you with general instructions for conducting the interviews of Psychology 201 subjects. The second section provides you with a description of each objective cue that you will have available for each interviewee. The third section contains copies of the two criterion measures on which you will be predicting each interviewee's scores, along with statistical descriptions of these criterion measures. Finally, the last section contains a copy of the structured interview you will be conducting with each interviewee.

Study these materials carefully over the next week so that you will be familiar with your task. Remember that we will meet briefly, to practice the interview and to resolve any last-minute problems, on Monday, 22 September at 10:00 AM in the graduate student lounge.

Thanks again for your help.

General Instructions to Judges

1. Try to be in the third floor lobby 5-10 minutes before your scheduled interview period.
2. You will pick up an interviewee at the desk in the third floor lobby. Doug will give you the interviewee's folder.
3. Whether the interviewee is a freshman or an upperclassman will be noted on the folder (F or U). Ask the questions in section V of the interview accordingly.
4. Each folder will contain the interviewee's objective cue profile and an interview form for that interviewee. Look at the cue profile first, before beginning the interview.
5. Follow the interview format explicitly when conducting the interview.
6. Mark every rating scale within the interview as you come to it. Circle the appropriate number on the scale. Do not rate between whole numbers on the scales.
7. Make your predictions on the profile sheet after the interviewee has left the room. Be sure to include your judge number at the bottom of the profile sheet.
8. Bring the interviewee's folder, including your ratings and predictions, to Doug in the lobby.

Description of Objective Cues

There are six objective cues provided for each interviewee. The cues' values can be found on the cue profile sheet in each interviewee's folder. A sample profile sheet is provided on the next page of this package. The following is a description of the six cues.

1. Sex: M=male; F=female.
2. High school decile: In which tenth of the class did the interviewee graduate? A "1" indicates the interviewee graduated in the top 10% of his class; a "2" indicates he graduated in the second 10% of his class; etc.
3. ACT scores: For each of the four areas covered in the ACT test, the interviewee's standard score is given. The following data are provided for the population of students taking Psychology 201 in the fall quarter.

<u>Area</u>	<u>Mean</u>	<u>S.D.</u>	<u>Range of Possible Scores</u>
English	19.8	4.4	1-33
Mathematics	21.2	6.8	1-36
Social Studies	20.1	6.5	1-34
Natural Sciences	22.4	5.6	1-35

Criterion Measures

On the pages immediately following this one you will find copies of each of the two criterion measures. One measure is a 15-item attitude scale, and one measure is a 25-item multiple-choice psychology examination. The following data were gathered during a pretest of these measures on students enrolled in Psychology 201 during the summer quarter.

Attitude Scale (N=76)

Mean=56.3

S.D.=8.5

Range of possible scores=15-75 in steps of 1.

Psychology Examination (N=53)

Mean=61%

S.D.=20%

Range of possible scores=0-100 in steps of 4.

Instructions to Prospective Subjects

On the following pages you will find a list of social security numbers arranged in ascending order. If your number appears on the list, you have been selected to participate in a psychology experiment for which you will receive $\frac{1}{2}$ -hour of credit. You must have two hours of credit for participation in experiments as part of the course requirement in Psychology 201.

This experiment requires you to participate in an interview with a graduate-student interviewer. If your social security number is on the attached list, and if you desire to participate in this experiment, so indicate by taking one experiment card from those available at the front of the room immediately after class today. Each card has a date and a time during which the experiment will be in progress. Take a card with a date and time that you are sure you will be able to meet. Be in the lobby on the third floor of the Psychology Building on the day and time indicated on your experiment card. The researcher, Doug Gorman, will be there to instruct you further.

Thank you very much for your help.

Debriefing Information for Interviewing Experiment

You have just participated in a brief interview with a graduate-student interviewer. The purpose of this experiment is to determine how accurately a judge (the graduate student) can predict future behavior of individuals (like you) based upon a brief conversation (or interview) with them.

At the end of the quarter, you will take a final examination in Psychology 201. You will also be asked to complete an attitude survey. It is your scores on these measures which the interviewer you have just met will be attempting to accurately predict.

Your experiment card (or one similar to it) has been stamped with the name "Doug Gorman." Hand this card in to your Psychology 201 instructor to receive $\frac{1}{2}$ -hour of credit for your participation. Thank you again for your help.

General Instructions

1. On the following several pages you will find cue descriptions, descriptions of the criterion measures, and sample forms identical to those you used in the interview sessions of a couple of weeks ago. Study these carefully, and feel free to refer to them at any time throughout your accomplishment of the task.
2. If you made any notes during the interview sessions of a couple of weeks ago, do not refer to them at any time during your completion of this task. Do not discuss the task with any of the other participating judges until after everyone has finished it.
3. Following the introductory material, this notebook contains cue profiles representing students enrolled in introductory psychology this quarter. Your task is to study each cue profile, and make the predictions asked for on the profile sheet.
4. Begin at the front of this notebook, and work straight through to the back. Do not turn back to predictions once you have made them. Do not thumb through the notebook prior to beginning the task.
5. The last few pages of this notebook ask for some additional information from you. Instructions are included at the top of each page. Complete these pages after all predictions have been made.

Subjective Weighting I

You have just completed making predictions for over 75 cases represented by six cues each. For each of the two criterion measures (the examination and the attitude scale), take 100 points and distribute them among the six cues which were present on each profile, so as to indicate the relative weight you feel you assigned to each cue in making your predictions.

<u>Cue</u>	Weight in predicting examination <u>scores</u>	Weight in predicting attitude scale <u>scores</u>
Sex	_____	_____
High School Decile	_____	_____
ACT English	_____	_____
ACT Mathematics	_____	_____
ACT Social Studies	_____	_____
ACT Natural Sciences	_____	_____
Total	100	100

Subjective Weighting II

About two weeks ago, you interviewed and made predictions concerning over 50 introductory psychology students' future performance on two criterion measures (an examination and an attitude scale). Those predictions were based on several cues. For each of the two criterion measures, take 100 points and distribute them among the several cues listed below, so as to indicate the relative weight you feel you assigned to each cue in making your predictions.

<u>Cue</u>	Weight in predicting examination <u>scores</u>	Weight in predicting attitude scale <u>scores</u>
Sex	_____	_____
High School Decile	_____	_____
ACT English	_____	_____
ACT Mathematics	_____	_____
ACT Social Studies	_____	_____
ACT Natural Sciences	_____	_____
Class (F or U)	_____	_____
Expectations	_____	_____
Grade expected	_____	_____
Motivation	_____	_____
Experience	_____	_____
Gen. Adjust. & Mat.	_____	_____
Total	100	100

General Questions

Please answer the following questions. Feel free to expand and provide any comments as you see fit.

1. Were you more confident or less confident of your predictions of examination scores when you had an opportunity to interview the student than when you had to rely only on the objective cue profile? Why?

2. Were you more confident or less confident of your predictions of attitude scale scores when you had an opportunity to interview the student than when you had to rely only on the objective cue profile? Why?

3. Which procedure (objective cue profile only, or objective cue profile plus interview) do you believe will be a more valid procedure for predicting exam scores, as far as your judgments are concerned?

Do you think this will be generally true for all the judges who participated in the study? Why or why not?

4. Which procedure (objective cue profile only, or objective cue profile plus interview) do you believe will be a more valid procedure for predicting attitude scale scores, as far as your judgments are concerned?

Do you think this will be generally true for all the judges who participated in the study? Why or why not?

5. Do you believe the presence of the student had any substantial effect on the accuracy of your prediction of the exam score? If yes, was that effect generally positive or negative?
6. Do you believe the presence of the student had any substantial effect on the accuracy of your prediction of the attitude scale score? If yes, was that effect generally positive or negative?

7. Make any comments you may have concerning this study in the space provided below. Use extra sheets if necessary.

INTERCORRELATIONS OF OBJECTIVE CUES,
TOTAL SAMPLE (N=461)

Cue	Sex ^a	Decile	English	Math	Social Studies
Sex	----				
Decile	-.34	----			
English	.23	-.40	---		
Math	-.09	-.43	.37	---	
Social Studies	-.13	-.36	.51	.44	---
Natural Sciences	-.15	-.33	.45	.57	.64

a. Male = 1, Female = 2

CROSS-VALIDATED MULTIPLE Rs FOR EXAMINATION
JUDGMENTS IN TWO EXPERIMENTAL TREATMENTS

Judge Number	Treatment	Multiple R
1	Interview	.96
	Paper-People	.97
2	Interview	.92
	Paper-People	.90
3	Interview	.84
	Paper-People	.87
4	Interview	.68
	Paper-People	.87
5	Interview	.95
	Paper-People	.98
6	Interview	.77
	Paper-People	.92
7	Interview	.81
	Paper-People	.97
8	Interview	.78
	Paper-People	.93

CROSS-VALIDATED MULTIPLE R_s FOR ATTITUDE SCALE
JUDGMENTS IN TWO EXPERIMENTAL TREATMENTS

Judge Number	Treatment	Multiple R
1	Interview	.74
	Paper-People	.77
2	Interview	.79
	Paper-People	.88
3	Interview	.61
	Paper-People	.82
4	Interview	.61
	Paper-People	.82
5	Interview	.71
	Paper-People	.97
6	Interview	.77
	Paper-People	.71
7	Interview	.89
	Paper-People	.97
8	Interview	.16
	Paper-People	.93

CORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA,
JUDGE NUMBER ONE (N = 50)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.14																
English	-.04	-.37															
Math	-.29	-.45	.15														
Social Studies	-.30	-.22	.56	.41													
Natural Sciences	-.22	-.20	.56	.46	.66												
Class	-.14	-.04	-.23	.06	.03	-.45											
Expectation	.1	-.04	.10	.11	.22	.10	.12										
Grade	.06	.33	-.26	-.37	-.31	-.33	-.02	-.11									
Motivation	-.20	.19	.30	-.15	.22	.27	-.34	.20	-.09								
Experience	.04	.32	.07	-.30	.05	.13	-.15	.11	.08	.50							
Maturity	-.07	-.12	-.03	.14	-.15	.04	.06	-.12	-.05	-.01	.05						
Exam with Interview	-.24	-.33	.63	.47	.66	.77	-.35	.17	-.35	.49	.20	.26					
Attitude/Interview	-.09	.24	.29	-.03	.23	.34	-.42	.10	-.01	.76	.52	.24	.55				
Exam/Paper	-.20	-.43	.59	.66	.79	.79	-.10	.14	-.50	.08	-.06	.03	.67	.13			
Attitude/Paper	-.20	-.25	.48	.31	.87	.65	-.08	.32	-.25	.20	.02	-.17	.54	.14	.71		
Exam Criteria	-.09	.08	.35	.09	.36	.43	-.34	.07	-.31	.52	.33	.11	.50	.56	.30	.32	
Attitude Criteria	.16	.01	.27	.09	.11	.22	-.21	.15	-.32	.28	.28	-.09	.29	.34	.13	.02	.39

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA,
JUDGE NUMBER TWO (N = 49)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.16																
English	.23	-.50															
Math	-.19	-.55	.52														
Social Studies	-.14	-.50	.47	.57													
Natural Sciences	-.26	-.48	.48	.55	.69												
Class	-.07	-.16	-.01	.32	-.02	.01											
Expectation	.01	.15	.10	-.08	-.02	-.05	-.26										
Grade	.27	.33	-.30	-.34	-.27	-.22	-.09	-.02									
Motivation	-.18	-.04	.13	.01	.17	.17	-.21	.37	-.41								
Experience	-.03	.13	.18	-.11	-.18	-.09	-.31	.30	-.25	.36							
Maturity	-.05	-.52	.41	.36	.35	.51	-.00	.08	-.50	.55	.28						
Exam with Interview	-.18	-.57	.54	.50	.55	.59	.01	.15	-.58	.42	.28	.79					
Attitude/Interview	.01	-.40	.47	.32	.37	.32	-.05	.21	-.62	.47	.45	.72	.84				
Exam/Paper	-.15	-.75	.59	.63	.84	.67	-.02	-.06	-.32	.17	-.12	.52	.60	.35			
Att/Paper	-.03	-.70	.61	.54	.83	.57	-.01	-.07	-.30	.17	-.13	.51	.62	.41	.93		
Exam Criteria	-.18	-.51	.50	.46	.65	.55	.03	.14	-.41	.14	-.04	.41	.61	.44	.67	.60	
Attitude Criteria	.17	-.12	.29	.14	.16	-.05	-.03	.47	-.01	.10	.15	.00	.04	.11	.14	.11	.10

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA,
JUDGE NUMBER THREE (N = 47)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.41																
English	.20	-.35															
Math	.09	-.59	.41														
Social Studies	-.27	-.32	.46	.44													
Natural Sciences	-.01	-.37	.54	.63	.55												
Class	-.16	-.32	.33	.23	.32	.16											
Expectation	.15	-.09	.17	.09	-.04	.11	-.07										
Grade	-.14	.18	-.10	-.24	-.01	-.20	-.03	-.36									
Motivation	.23	.10	-.17	-.01	-.15	-.07	-.41	.32	-.19								
Experience	.22	.05	-.04	.03	.14	-.12	-.21	.29	-.09	.56							
Maturity	.14	-.07	.30	.18	.12	.12	.12	.57	-.38	.37	.36						
Exam with Interview	.30	-.53	.55	.70	.58	.70	.02	.19	-.25	.17	.26	.24					
Attitude/Interview	.48	-.32	.39	.45	.19	.38	-.13	.43	-.36	.37	.42	.37	.74				
Exam/Paper	.07	-.59	.52	.73	.60	.71	.36	.90	-.25	.03	.07	.24	.75	.47			
Att/Paper	.23	-.72	.53	.71	.59	.61	.32	.22	-.28	.05	.12	.27	.74	.47	.83		
Exam Criteria	.20	-.15	.53	.28	.30	.39	.09	.13	-.08	.08	.16	.16	.42	.38	.27	.21	
Attitude Criteria	.32	-.06	-.08	.01	-.11	-.19	-.13	-.01	.02	.24	.27	.01	.06	.32	-.07	.04	.21

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA
JUDGE NUMBER FOUR (N = 51)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.35																
English	.37	-.04															
Math	.01	-.36	.26														
Social Studies	-.17	-.30	.49	.22													
Natural Sciences	-.26	-.34	.37	.60	.48												
Class	-.12	.11	-.21	-.02	-.12	-.12											
Expectation	.03	-.11	.15	.07	-.08	.12	.05										
Grade	-.24	.29	-.31	-.12	-.12	-.10	-.07	-.35									
Motivation	.01	.03	.27	-.32	.17	-.05	-.18	.34	-.04								
Experience	.05	-.08	.20	-.10	.18	.28	-.25	.19	-.06	.47							
Maturity	-.15	-.18	.15	.13	.22	.11	.21	.34	-.25	.12	.10						
Exam with Interview	.12	-.38	.48	.29	.39	.35	.02	.34	-.38	.28	.28	.49					
Attitude/Interview	.19	-.27	.51	.02	.25	.27	-.24	.22	-.23	.47	.56	.29	.51				
Exam/Paper	.18	-.71	.56	.45	.65	.50	-.03	.02	-.28	.05	.17	.23	.50	.24			
Attitude/Paper	.24	-.63	.44	.41	.56	.47	.06	.03	-.30	.04	.13	.17	.45	.15	.89		
Exam Criteria	.00	-.42	.37	.33	.35	.43	-.06	.30	-.22	.15	.24	.13	.59	.29	.43	.34	
Attitude Criteria	.02	-.15	.25	.06	.08	.21	-.40	.17	-.28	.08	.22	-.17	-.02	.28	.09	-.01	.35

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA
JUDGE NUMBER FIVE (N = 58)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.31																
English	.07	-.34															
Math	-.14	-.37	.30														
Social Studies	-.22	-.43	.40	.53													
Natural Sciences	-.27	-.25	.24	.52	.63												
Class	-.20	-.02	-.01	-.18	.09	-.31											
Expectation	-.06	-.36	.20	.12	.22	.27	.05										
Grade	.01	.38	-.14	-.15	-.22	-.09	.10	-.32									
Motivation	.17	-.25	.09	-.03	.12	.13	-.25	.53	-.27								
Experience	.14	-.04	-.04	-.11	-.13	-.20	.25	.13	-.20	.19							
Maturity	-.11	-.44	.32	.32	.46	.36	-.05	.42	-.35	.52	.30						
Exam with Interview	-.03	-.65	.50	.56	.81	.57	-.09	.46	-.32	.36	-.03	.72					
Attitude/Interview	.21	-.22	.20	.20	.24	.24	-.25	.46	-.18	.60	.23	.54	.47				
Exam/Paper	.00	-.72	.42	.56	.87	.57	.03	.37	-.37	.29	-.08	.53	.86	.32			
Attitude/Paper	-.09	-.73	.47	.55	.91	.55	.15	.33	-.21	.17	-.08	.52	.84	.32	.89		
Exam Criteria	.16	-.57	.30	.44	.47	.40	-.14	.35	-.33	.14	-.04	.29	.53	.33	.60	.57	
Attitude Criteria	.02	-.03	-.05	.08	-.07	.15	-.11	-.20	.14	-.09	.10	-.09	-.09	.07	-.03	-.04	.07

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA
JUDGE NUMBER SIX (N = 44)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.52																
English	.18	-.39															
Math	-.06	-.37	.57														
Social Studies	-.03	-.29	.49	.55													
Natural Sciences	-.00	-.43	.61	.75	.68												
Class	-.07	-.05	.06	.07	.12	.03											
Expectation	.29	-.27	.16	-.11	-.19	.08	-.16										
Grade	-.05	.24	-.33	-.20	-.38	-.40	-.02	-.02									
Motivation	.47	-.18	.06	-.08	-.09	.08	-.34	.56	.11								
Experience	.17	-.09	.26	.02	.16	.11	-.00	.27	-.15	.40							
Maturity	.45	-.43	.24	.01	.23	.09	.23	-.15	-.08	.05	.08						
Exam with Interview	.57	-.70	.56	.46	.50	.57	.07	.19	-.22	.28	.18	.57					
Attitude/Interview	.51	-.62	.46	.18	.31	.38	.09	.33	-.17	.36	.29	.57	.79				
Exam/Paper	.14	-.60	.79	.77	.70	.83	.17	.05	-.37	.04	.17	.33	.70	.48			
Attitude/Paper	.08	-.40	.56	.71	.69	.70	-.00	.03	-.26	.20	.32	.16	.49	.32	.78		
Exam Criteria	-.03	-.21	.33	.38	.30	.51	.08	.14	-.51	.00	.10	-.13	.17	.16	.42	.30	
Attitude Criteria	.08	-.29	.12	.08	.25	.21	.19	.01	-.15	.08	-.19	.03	.15	.27	.19	.09	.18

INTERCORRELATIONS BETWEEN CUES, JUDGMENTS, AND CRITERIA
JUDGE NUMBER SEVEN (N = 59)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.43																
English	.37	-.46															
Math	-.12	-.45	.37														
Social Studies	.03	-.53	.57	.59													
Natural Sciences	-.16	-.41	.42	.68	.78												
Class	-.07	.19	-.20	-.20	-.14	-.18											
Expectation	-.06	-.06	.08	-.07	.13	.15	.09										
Grade	-.17	.29	-.37	-.07	-.31	-.28	.07	-.17									
Motivation	.19	-.12	.37	.04	.14	.05	-.41	-.02	-.11								
Experience	.17	-.05	.21	-.12	-.13	-.09	-.08	.43	-.26	.36							
Maturity	-.07	-.13	.10	.04	.20	.16	.16	.05	-.03	.30	.18						
Exam with Interview	.27	-.55	.66	.42	.56	.52	-.14	.23	-.37	.50	.36	.40					
Attitude with Interview	.42	-.51	.48	.11	.40	.29	-.27	.18	-.40	.56	.33	.38	.77				
Exam Paper	.24	-.63	.84	.65	.78	.67	-.30	.16	-.31	.29	.09	.06	.70	.49			
Attitude Paper	.15	-.57	.66	.61	.94	.71	-.19	.13	-.32	.16	-.04	.11	.57	.39	.86		
Exam Criteria	-.10	-.32	.36	.26	.48	.45	.13	.31	-.15	.16	.15	.16	.40	.31	.41	.45	
Attitude Criteria	.17	-.25	.20	.11	.17	.19	-.16	.29	-.10	.15	.24	.12	.09	.08	.24	.18	.09

INTERCORRELATIONS BETWEEN CUES, JUDGEMENTS, AND CRITERIA,
JUDGE NUMBER EIGHT (N = 42)

	Sex	Decile	English	Math	Social Studies	Natural Sciences	Class	Expectation	Grade	Motivation	Experience	Maturity	Exam/Int	Att/Int	Exam/Paper	Att/Paper	Exam Criteria
Decile	-.41																
English	.48	-.41															
Math	-.06	-.45	.22														
Social Studies	.03	-.30	.46	.39													
Natural Sciences	-.03	-.38	.41	.56	.65												
Class	-.21	-.22	-.11	.19	.21	.16											
Expectation	.24	-.04	.11	.00	-.01	-.11	-.27										
Grade	-.03	.29	-.21	-.34	-.33	-.28	-.30	-.19									
Motivation	.21	-.11	.05	.19	.11	.05	.06	.36	-.34								
Experience	-.04	.06	-.04	.07	.05	.00	-.12	.55	-.10	.48							
Maturity	.08	-.31	.25	.46	.52	.35	.30	.19	-.47	.33	.17						
Exam with Interview	.26	-.59	.37	.45	.51	.50	.25	.32	-.42	.42	.22	.50					
Attitude/Interview	.13	-.28	.20	.35	.25	.21	.12	.43	-.54	.48	.23	.49	.65				
Exam/Paper	.32	-.84	.60	.67	.47	.51	.18	.03	-.44	.12	-.05	.42	.56	.33			
Attitude/Paper	.60	-.88	.53	.42	.34	.39	.19	-.01	-.28	.12	-.14	.37	.51	.29	.81		
Exam Criteria	.21	-.50	.20	.33	.47	.50	.07	.04	-.15	.22	.05	.24	.44	.02	.47	.49	
Attitude Criteria	-.01	-.07	.04	.02	.24	.17	.04	.04	-.03	.06	.06	.23	.34	.36	.01	.03	.06

R_e VALUES FOR EACH CRITERION WITH SIX OBJECTIVE CUES
 BASED ON REDUCED NUMBER OF CASES FOR EACH JUDGE

Judge No.	No. Cases	R_e Exam	R_e Attitude
1	50	.51	.41
2	49	.72	.40
3	47	.59	.43
4	51	.55	.32
5	58	.66	.27
6	44	.52	.38
7	59	.57	.30
8	42	.68	.28
Mean	--	.61	.35

MEANS AND STANDARD DEVIATIONS FOR THE OBJECTIVE AND INTERVIEW CUES

Group	No. Cases	Sex ^a		Decile		English		Math		Social Studies		Natural Science		Class ^b		Expect.		Grade ^c		Moti- vation		Experi- ence		Maturity	
		$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Judge 1	56	1.7	.45	2.8	1.74	20.3	4.07	21.2	6.45	20.9	5.50	22.5	5.15	1.4	.50	4.9	1.10	1.8	.47	4.5	1.43	3.4	1.89	4.7	1.01
Judge 2	54	1.8	.39	2.5	1.70	20.7	4.41	22.2	6.67	22.3	5.91	24.1	4.99	1.4	.50	4.6	.96	1.9	.61	4.0	1.10	3.3	1.37	4.4	1.23
Judge 3	50	1.6	.49	2.8	1.57	20.5	4.25	22.7	5.63	21.2	6.22	22.3	5.54	1.5	.50	4.4	1.05	1.8	.55	4.3	1.27	2.8	1.58	4.4	1.03
Judge 4	55	1.7	.46	3.1	2.14	19.8	4.18	20.1	7.54	21.0	5.72	23.0	4.90	1.3	.47	4.0	1.09	1.8	.54	4.0	1.27	3.7	1.48	4.3	1.13
Judge 5	61	1.7	.46	2.8	1.80	19.8	3.85	22.4	6.33	20.1	5.67	22.4	5.30	1.3	.47	4.6	1.13	1.8	.45	3.8	1.27	2.7	1.04	4.4	1.15
Judge 6	55	1.7	.45	2.9	2.02	19.8	4.39	21.3	6.82	19.3	6.87	22.5	5.08	1.4	.49	5.8	1.42	1.8	.56	3.9	1.33	2.7	1.87	5.5	1.23
Judge 7	61	1.7	.47	2.8	1.93	20.2	4.43	22.9	6.19	20.6	6.27	23.8	5.23	1.3	.48	4.9	1.39	1.6	.53	4.7	1.20	3.0	1.74	5.1	1.17
Judge 8	48	1.6	.48	3.3	2.01	20.1	4.44	22.4	5.61	19.9	5.77	23.0	5.09	1.4	.49	4.6	.87	1.7	.54	3.6	1.21	2.6	1.30	4.8	.95

a. Male=1, Female=2.

b. Freshman=1, Upperclassman=2.

c. A=1, B=2, C=3, D=4, F=5.

MEANS AND STANDARD DEVIATIONS
FOR THE CRITERION MEASURES

Group	No. Cases	Exam Mean	Exam Std. Dev.	Attitude Mean	Attitude Std. Dev.
Judge 1	50	63.8	11.28	55.3	7.35
Judge 2	49	65.9	12.73	54.5	5.85
Judge 3	47	66.0	14.08	56.8	7.74
Judge 4	51	66.4	12.55	55.0	6.65
Judge 5	58	63.0	14.40	54.0	6.92
Judge 6	44	62.1	13.60	52.5	7.38
Judge 7	59	70.1	11.92	56.7	6.79
Judge 8	42	64.1	13.56	55.0	7.47
Total Sample	400	65.3	13.14	55.0	7.07

INTER-RATER RELIABILITIES FOR THE PAPER-PEOPLE TREATMENT
 BASED ON 25 PROFILES COMMON TO ALL NOTEBOOKS,
 EXAMINATION JUDGMENTS

Judge Number	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7
2	.83						
3	.75	.81					
4	.76	.87	.78				
5	.81	.91	.90	.85			
6	.85	.88	.89	.91	.89		
7	.84	.83	.87	.82	.89	.93	
8	.85	.88	.75	.83	.84	.87	.81

INTER-RATER RELIABILITIES FOR THE PAPER-PEOPLE TREATMENT
BASED ON 25 PROFILES COMMON TO ALL NOTEBOOKS,
ATTITUDE SCALE JUDGMENTS

Judge Number	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7
2	.60						
3	.72	.76					
4	.70	.84	.68				
5	.73	.80	.70	.89			
6	.63	.82	.74	.72	.79		
7	.80	.77	.85	.75	.83	.74	
8	.42	.76	.65	.78	.62	.64	.62