UNDERSTANDING THE IMPACT OF PRE-INTERVIEW INFORMATION ON THE

RELIABILITY, VALIDITY, ACCURACY AND

DIFFERENTIAL VALIDITY OF EMPLOYMENT INTERVIEW DECISIONS:

COMPARISONS ACROSS INTERVIEW QUESTION TYPE,

RATING SCALE AND SCORING PROTOCOLS

By

PETER ALEXANDER HAUSDORF, M.A.

A Dissertation

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

Faculty of Business

McMaster University

UNDERSTANDING EMPLOYMENT INTERVIEW DECISIONS

DOCTOR OF PHILOSOPHY (1997)   McMASTER UNIVERSITY
(Business)   Hamilton, Ontario

TITLE: Understanding the impact of pre-interview information on the reliability,
            validity, accuracy and differential validity of employment interview
            decisions: Comparisons across interview question type, rating scales and
            scoring protocols.

AUTHOR: Peter Alexander Hausdorf,   B.Sc. (McMaster University)
                                        M.A. (University of Guelph)

SUPERVISOR: Professor W. H. Wiesner

NUMBER OF PAGES: xiii, 154

ABSTRACT

The impact of pre-interview impressions on the reliability, validity, accuracy and bias of interview decisions was examined in the context of different interview types. Although it is generally well accepted that structuring the interview improves its reliability and validity, little research has been conducted with respect to resistance to interviewer bias.

Dipboye (1982) proposed that pre-interview impressions can impact on post interview decisions through the interviewer's behaviour and cognitive processes. This study used videotaped interviews to assess the impact of a negative/inaccurate pre-interview impression on subsequent interview ratings via the cognitive process. Predictor data consisted of pre- and post interview ratings and criterion data consisted of supervisor ratings of 'on-the-job' performance. Three types of interview formats were manipulated: question type (job relevant versus general), rating scale (graphic versus behaviourally anchored), and combination of ratings (clinical versus mechanical).

There were considerable differences in the interview results when the pre-interview information about a candidate was positive (accurate) versus negative (inaccurate). Specifically, when positive (accurate) information preceded the interview the greatest contributor to interview reliability and validity was the use of job relevant questions. However, when the interview was preceded by negative (inaccurate) information about a

candidate the use of job relevant questions, behaviourally anchored rating scales and a mechanical combination of ratings were all necessary to eliminate any impact of bias on the reliability and validity of the interviewer's decision. These results indicate that structuring the interview can minimize the potentially negative effects of interviewer biases on the reliability and validity of interview decisions.

ACKNOWLEDGEMENTS

When I started this dissertation four years ago I had no idea that it would take this long nor that it would place considerable demands on my family, my committee and myself. A considerable amount of time and effort has been invested by Dr. Willi Wiesner throughout this process from an academic and technical role to a primarily motivational one. To Willi I am eternally grateful for making sure that I completed the dissertation within the context of many competing demands.

From a similar perspective, thanks goes to my committee members (Dr. Rick Hackett, Dr. Naresh Agarwal and Dr. Ken Deal) who persisted through the process and were able to follow the information, ideas and revisions over long waiting periods. I am particularly indebted to Rick for his detail orientation with respect to every aspect of the dissertation.

Last but not least, thanks to my wife Donna for her encouragement and support throughout this process. She was my sounding board, data entry person, filer and motivator all rolled into one person. This would not have been accomplished without her.

TABLE OF CONTENTS

CONTENTS

Chapter

CONTENTS

CONTENTS

Chapter

CONTENTS

Chapter

Appendix

CONTENTS

TABLES

FIGURES

CHAPTER 1:

INTRODUCTION

The employment interview continues to be the most widely used selection

instrument in North America (Rowe, Williams, and Day 1994). Despite the critical

importance of this process both for organizations (i.e., to identify the best applicants in a

fair and nonbiased manner) and for the applicants (i.e., to find a good company to work

for and to be treated fairly) we do not completely understand the decision making process

and contextual influences that impact on interviewers' decisions.

A considerable body of literature exists on the employment interview including

extensive reviews (Wagner 1949; Mayfield 1964; Ulrich and Trumbo 1965; Wright 1969;

Schmitt 1976; Arvey 1979; Arvey and Campion 1982). A recent review of interview

research by Harris (1989) recommended that:

- more theories need to be developed to understand interview
  processes and outcomes,
- more innovative research designs are needed in research (e.g., videotapes
  of actual interviews), and
- more research is needed on a variety of practical issues relevant to
  interviews in general and the structured interview in particular (e.g.,
  incremental validity, cost, adverse impact).

The purpose of this thesis is to demonstrate the impact of pre-interview information on the reliability, validity, accuracy and bias of interview decisions and to assess how interview structure (e.g., job relevant questions, behaviourally anchored rating scales and a mechanical combination of ratings) moderates that impact. This research will further develop our understanding of the employment interview by attempting to integrate two streams of previous research on the interview which are typically referred to as process and outcome oriented research. The process perspective focuses on pre-interview information as it affects interviewer decision making. The outcome perspective focuses on the impact of pre-interview information and interviewer decision making on hiring decision outcomes (e.g., decision validity). Thus the thesis draws upon the recommendations made by Harris (1989). In order to set the stage for the study, several issues are discussed. These are: interview types, evaluation criteria, and the interview literature.

The employment interview represents a complex social and psychological process. Discussing the different types of interview will help the reader gain an appreciation for the variety of interviewing methods that are available which have contributed to the complexity of the interview research literature. Reviewing the evaluation criteria serves two purposes. First, it provides the necessary information for understanding why some interview researchers advocate for one type of interview over another. Second, it identifies the tools with which the interview is evaluated. The presentation of theory and

empirical findings are necessary in order to clearly present the logic for the hypotheses

presented in this thesis.

Interview Types: How many different types of interviews exist?

In the literature there are two broad categories of interview which are simply

referred to as structured and unstructured interviews. Although this classification implies

that these categories are dichotomous, most interviews exist along a continuum from

unstructured to structured. However, for the purposes of convenience and simplicity the

thesis will use the structured/unstructured distinction.

Structured Interviews

Dipboye and Gaugler (1993) describe structured interviews as having the

following features:

- standardized questioning of applicants,
- questions/ratings based on job analysis data,
- interviewers provided with information about job requirements,
- interviewers received no ancillary data prior to the interview,
- interviewers required to take notes during the interview,
- use well-defined rating scales,
- interviewee responses recorded during the interview with evaluation occurring after,
- ratings are decomposed into component dimensions (e.g., separate ratings of education and social skills),
- ratings are combined mechanically,
- multiple interviewers are used, and
- interviewers received comprehensive training.

It is important to note that not all structured interviews contain every characteristic. As mentioned previously, structured interviews tend to contain more of these aspects than unstructured interviews. A minimum for defining structured interviews comes from Pursell, Campion and Gaylord (1980, 908) in which structure was defined as: "a series of job-related questions with predetermined answers that are consistently applied across all interviews for a particular job" with ratings scored for each question that are combined for an overall interview rating.

Although there are many possible variations of structured interview characteristics as well as many different ways of acquiring information in the structured interview, several approaches have become well known in the interviewing literature. These are:

- Patterned Behaviour Description Interview (Janz 1982),
- Situational Interview (Latham, Saari, Pursell and Campion 1980), and
- Comprehensive Structured Interview (Pursell, Campion and Gaylord, 1980).

Patterned Behaviour Description Interview (PBDI)

The PBDI (also known as the BDI), (Janz, Hellervik and Gilmore 1986) collects interviewee descriptions of past general behaviour that relates to potential job behaviour. It is based on the notion that past behaviour is a good predictor of future behaviour (Wernimont and Campbell 1968). Using this methodology questions are created based on

job analysis data which identifies effective and ineffective job behaviours (also known as critical incidents; Flanagan 1954) as identified by experienced job incumbents. These behaviours are then converted into performance dimensions.

Based on the critical incidents, questions are developed to tap into past experiences of applicants and their responses are scored based on the performance dimensions. After the interview has been developed, it is reviewed by supervisors and incumbents for clarity, specificity, and comprehensiveness.

In addition to the questions, common probes are used (e.g., what were the circumstances? what led to the event?, when did it happen?, what was the outcome?). Essentially, the questions are designed to encourage applicants to recall a previous experience in as much detail as possible so that the interviewer can provide ratings of the applicants' past performance that will predict their future performance in similar situations.

The following is an example of a PBDI question with corresponding performance dimensions. This question was used to assess an applicant's ability to arrive on time and orientation to driving safely for a transit operator (bus driver) position.

Tell me about a time when you were driving to a meeting or engagement and found yourself running late?

Probes: What were the circumstances?
           Where were you going?
           What did you do?
           What was the outcome?

Performance Dimensions:
1   Drove in an unsafe manner (e.g., ran red lights, greatly exceeded the speed limit) and arrived late.
2   -
3   Drove safely and arrived late.
4   -
5   Adjusted driving behaviour while driving safely (e.g., drove a little over the speed limit), arrived on time.

Situational Interview (SI)

The SI has a different conceptual basis than the PBDI. It is based on the idea that future behaviours are positively related to present intentions (Locke's goal setting theory 1968). In other words, if intentions do in fact, lead to behaviour then this type of interview can provide accurate information about an applicant's future job behaviour.

Typically, SI questions are developed by collecting job analysis data through the critical incident technique. Behavioural observation scales (BOS's) are then developed for each question and the final list of questions are reviewed by supervisors (Latham et al. 1980). An example of a typical SI question with corresponding BOS is:

Your spouse and two teenage children are sick in bed with a cold. There are no relatives or friends available to look in on them. Your shift starts in 3 hours. What would you do in this situation?

BOS:

1	I'd stay home
2	-
3	I'd phone my supervisor and explain my situation
4	-
5	Since they only have colds, I'd come to work

The primary difference between the PBDI and SI pertains to the time orientation with PBDI's focusing on the past and SI's focusing on the future. As a result of the difference in time orientation, PBDI questions usually require probes whereas SI questions do not because with the PBDI the interviewee inevitably does not provide all of the relevant information that the interviewer needs to provide a rating. Both interviews utilize a variety of different rating scales.

Comprehensive Structured Interview (CSI)

The comprehensive structured interview covers a broader range of questions than either the PBDI or SI with areas covering situations, job knowledge, job simulations and worker requirements. The situational questions are similar to SI questions. A typical job knowledge question would ask the interviewee to describe the steps involved in some

aspect of the job (e.g., assembly/disassembly and repair of machine) whereas a job simulation question would ask the interviewee to demonstrate a task required for the job (e.g., reading an instructional procedure). A worker requirements question assesses skills the worker would need to be successful in the job (e.g., ask the worker how he/she feels about working in high places to assess his/her fear of heights). Similar to the PBDI and SI, the CSI uses structured scoring guides (Harris 1989).

In summary, these are three of the more common types of structured interviews. There are many variations incorporating different characteristics of structured interviews which emphasize the continuous nature of the structure concept. Although the concept of structured interviews has been around for some time (see McMurry 1947; Maas 1965) their extensive use in research on the interview has only emerged in the past 5 years. Moreover, unstructured interviews maintain wide popularity whereas the use of structured interviews is marginal.

Unstructured Interviews

Wiesner and Cronshaw (1988) defined unstructured interviews as: "following a free format without predetermined questions and/or no rating scales". In addition, unstructured interviews involve the interviewer making a global subjective rating of the interviewee at the end of the interview rather than combining scores based on individual

questions. A sample unstructured interview question is "Tell me about your strengths and weaknesses" which, when considered along with other questions the interviewer asks, contributes to the interviewer's overall suitability rating for the applicant.

Evaluation Criteria: How do we assess the interview as a selection device?

The employment interview is a selection device. Therefore, it should be held accountable to criteria that are used to evaluate all selection devices. According to the Guidelines for Educational and Psychological Testing established by the Canadian Psychological Association (CPA 1987), selection devices should be reliable, valid and nonbiased, therefore the interview should be reliable, valid and nonbiased. The guidelines have defined and set specific criteria for establishing reliability, validity and avoiding bias.

Reliability

Interview scores may vary unsystematically across different interview questions (internal consistency), interview occasions (test-retest reliability) or across different interviewers (inter-rater reliability) (Nunnally 1967). These differences are identified as errors of measurement and are described as measures of reliability. The guidelines state that proper test evaluation (and thus interview evaluation) should consist of identifying sources of measurement error, size of the error, reliability, and generalizability of results

across different circumstances. Moreover, these estimates of reliability should consider

the relevant sources of error and the expected aggregation of test scores (i.e., individuals

or groups of test takers) (CPA 1987). Reliability is crucial because it is considered a

necessary (although not sufficient) condition for validity (Nunnally 1967). In other words,

an instrument's validity is limited by its reliability.

Generalizability theory represents a recent re-conceptualization of reliability which

fits extremely well with interview research. Whereas the traditional concept of reliability

separates variance into its true and error components, generalizability theory isolates the

error associated with different sources (Shavelson, Webb and Rowley 1989). For

example, in an interviewer decision study sources of error could be due to the interviewer,

the interviewee or their interaction. This approach to reliability uses analysis of variance

to partial out the effects of different sources of error.

A generalizability study can be placed into two categories: a generalizability study

(G study) and a decision study (D study). The G study focuses on the development of a

measurement procedure and the D study applies that procedure (Shavelson, Webb and

Rowley 1989). Figure 1 presents a simple example of the generalizability calculations

involved in a decision study in which all interviewees (in this case two) are rated by all

interviewers (in this case four). (For simplicity any variance from interactions were not

separated from the residual variance.) The data represents an overall suitability score from one (unsuitable) to five (hire immediately).

Figure 1: An Example of Generalizability Theory

Interviewee	Interviewer			
	1	2	3	4
A	5	5	5	5
B	4	2	3	1

$$\rho^2_\iota = \frac{MS_{ee} - MS_{res}}{MS_{ee} + (n_r - 1)MS_{res}}$$

$$\rho^2_\iota = \frac{12.5 - .83}{12.5 + (4 - 1).83} = .78$$

	SS	df	MS	
Total	17.5	7		$\Sigma X^2 - (\Sigma X)^2/N$
Interviewee	12.5	1	12.5	$\Sigma X_e^2/n_e - (\Sigma X)^2/N = MS_{ee}$
Interviewer	2.5	3	0.83	$\Sigma X_r^2/n_r - (\Sigma X)^2/N = MS_{er}$
Residual	2.5	3	0.83	

In the example there is considerable agreement across interviewers with respect to Interviewee A but not Interviewee B. Moreover, all interviewers agree that Interviewee A is more qualified. The generalizability coefficient is calculated using the mean squares

from the two factor ANOVA and produces a value of .78 which represents considerable agreement across the interviewers.

Based on Classical Test Theory (Nunnally 1967) an examinee's observed score is represented by a true score and error score. Since the error component is assumed to be random, when one calculates the mean across multiple observations the effect of the error is minimized. Any deviations from the mean can be due to individual variance on the examinee, error or a combination of the two. When multiple raters are involved, as is often the case in interviews, any differences across interviewers adds another source of error variance. Generalizability theory, therefore, goes beyond classical test theory to portion out the variance between examinees, raters and error (Crocker and Algina 1986).

Validity

According to the guidelines, validity is the most important consideration in evaluation. Validity is focuses on the appropriateness, meaningfulness and usefulness of the specific inferences made from scores (CPA 1987). In other words, is the instrument measuring what it was designed to measure? Do the results it produces make sense? Are they useful? Validating the inferences made from scores (on tests or from interviews) consists of ascertaining construct-related, content-related, criterion-related validity evidence (CPA 1987). Given the unitary concept of validity proposed by Binning and

Barrett (1989), the critical determinant for validity is the collection of evidence that supports the inferences made from the interview. That is, validity is measured using the correlation between predictor scores (in this case interview scores) and job performance ratings (in this case supervisory ratings). To the extent that interview scores are positively and significantly correlated with job performance ratings, there is support for the validity of the test.

Accuracy

Whereas reliability and validity measure the strength of a relation between two sets of scores, accuracy measures the strength and kind of relation between scores. This concept has been popular in the performance appraisal literature in which the accuracy of managerial performance ratings have been evaluated in comparison to ratings from experts (Sulsky and Balzer 1988). The accuracy of a particular score is determined in relation to the expert rating (which is considered to be the true score). Sulsky and Balzer (1988) reviewed the many different types of accuracy indices available. The index chosen for this study was developed by Cronbach (1955) which is calculated by subtracting the true score from the actual score for each interviewee, squaring this value and averaging across all interviewers. Although this measure is similar to a difference score, it is distinct in that it does not imply that the difference represents a construct but simply a variation. This index

of accuracy is more precise than the correlation coefficient because it focuses on how different the scores are as opposed to simply how they covary with each other.

Bias

The issue of bias is complex. Jensen (1980, 375) defined bias as "systematic errors in the predictive or construct validity of test scores of individuals that are associated with the individual's group membership". In addition, Jensen (1980, 377) identified the concept of situational bias which he defined as "any influences in a testing situation, independent from the test itself, that may bias test scores". For example, any particular biases on the part of the interviewer would produce a bias in interview scores that would be independent of the interview questions. Since it is very difficult to identify an interviewer's potential biases, typically the measurement of bias has focused on statistical differences between applicants who differ on noticeable characteristics (e.g., gender). In other words, if one finds differences in selection decisions across applicants with noticeable differences (e.g., gender), the conclusion is that these outcomes are the result of bias.

There are two primary approaches to the measurement of bias. These are differential validity and differential prediction. According to Boehm (1972), differential validity exists if one or both validity coefficients are significantly different from zero and they differ significantly for one or more subgroups. Differential prediction is defined by

Cleary (1968, 115) as "if an instrument is biased for members of a subgroup for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup". Differential prediction exists if there are significant differences in the slopes, intercepts and standard errors of regression equations across subgroups (Bartlett, Bobko, Mosier, and Hannan 1978).

From a human rights perspective, bias can result in adverse impact. If, for example, an interviewer has a bias against female applicants, this bias could deflate interview scores for women which would reduce the proportion of women hired relative to men. Arvey (1979) reviewed the legal and psychological aspects of discrimination in the employment interview and concluded that the mechanisms of bias are not well understood and that more research is necessary.

CHAPTER 2:

THEORETICAL FRAMEWORK FOR INTERVIEW RESEARCH

The development of interview theories is a fairly recent phenomenon. Arvey and

Campion (1982) provided one of the first frameworks that integrated the diverse aspects

of the interview process. Prior to their work, much of the field focused on individual

effects.

There is a lack of an overriding framework or theory on interviewing (Hakel 1989)

which has contributed historically to the fragmentation of interview research (Eder and

Ferris 1989). With different interview researchers separated into two distinct camps and

focusing on different issues, this fragmentation will continue. For example, unstructured

interview researchers (of which Robert Dipboye's work is typical) tend to focus on the

interview as a cognitive and behavioural process (e.g., the effect of pre-interview biases on

the interview process, the impact of pre-interview information on interview decisions, the

impact of application information on interviewer recall). In contrast, structured interview

researchers (of which Willi Wiesner's work is typical) tend to focus on interview decision

outcomes (e.g., reliability and validity in relation to an external criterion). Recent theories

have attempted to integrate the findings from both streams of research.

Figure 2 on the next page represents a comprehensive framework for understanding interviewer decision making and is an amalgamation of four interview theories: interviewer information processing and decision making (Dipboye and Gaugler 1993), employment interview judgement from an interactionist perspective (Eder and Buckley 1988), sampling model of information processing in personnel decisions (Motowidlo 1986) and representation of construct and empirical validity of the interview (Cronshaw and Wiesner 1989). In addition, interview structure was added as it was under-represented in the individual models.

This framework can be divided into six categories: interview preparation, situational factors, interviewer characteristics, applicant characteristics, interview process and interview outcomes. Research findings will be presented in the context of these categories. Since collecting interview data requires a considerable amount of time and effort (i.e., interviews are time consuming), a complete test of this framework has not, as yet, been conducted. Despite this fact, independent studies have identified many effects that fit within these categories and although the interactions between these effects cannot be identified at this time, past research can provide support for independent aspects of the model.

Figure 2: Theoretical Framework for Interview Research
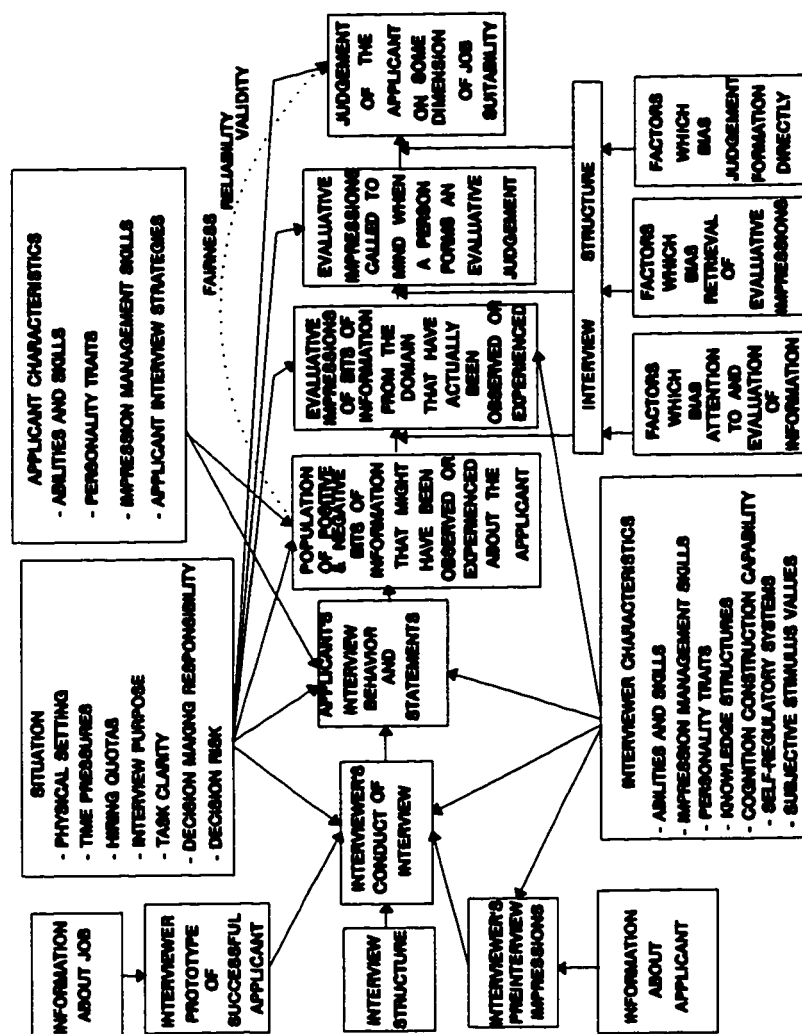
Interview Preparation

In actual organizations, most interviewers review applicant and job materials before the interview (Dalessio and Silverhart 1994). Therefore, interview preparation can include collecting job analysis data, creating a prototype of the ideal applicant, reviewing information about each applicant, or developing potential interview questions. This is an important stage in relation to the interview because it is here that the interviewer can develop his/her pre-interview impression which may impact on the interview process and its outcome (Dipboye 1982). The following aspects of interview preparation will be discussed: information about the job, interviewer prototypes of successful applicants, information about the applicant, interviewers' pre-interview impressions and interview structure.

Information About the Job

Wiesner and Cronshaw (1988) conducted a meta-analysis[1] of interview studies and tested whether the type of job analysis used to develop the interview had a moderating effect on the validity of the interview decision. There were three types: 'formal' based on a

---

[1] - Meta-analysis is a quantitative procedure for reviewing a body of literature. It provides the researcher with effect size estimates and allows the researcher to correct for artifactual sources of variance (such as sampling error, criterion unreliability, and range restriction). For an excellent review see Hunter and Schmidt's book titled Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (California: Sage, 1990).

standard methodology, 'armchair' which was defined as an informal assessment of job

content, and 'unknown' in which the job analysis technique was not described in the

research. Wiesner and Cronshaw (1988) found that job analysis type moderated the

validity of structured interviews with formal job analysis producing the highest validity

(.48 uncorrected, .87 corrected for restriction of range and criterion unreliability) and the

unknown job analysis type producing the lowest validity (.31 uncorrected, .56 corrected

for range restriction and criterion unreliability). Therefore, based on these results, job

analysis information contributes to the validity of interview decisions for structured

interviews, although the mechanism by which this is achieved is not clear.

Interviewer Prototype of Successful Applicant

In addition to contributing to the development of the interview, information about

the job can impact on the interviewer's prototype[2] of the successful applicant.

Furthermore, prototypes can vary significantly across interviewers. Mayfield and Carlson

---

[2] - This refers to a cognitive schema or reference point that all applicants are compared to in the decision making process. See the book titled Social Cognition: The Ontario Symposium (Volume 1) edited by Higgins, Herman and Zanna (New Jersey: Erlbaum & Associates, 1981) for an excellent review.

(1966) conducted several studies with "paper people"[3] and found that there was high agreement across interviewers regarding the favourability of some items of information (e.g., earned all of his college expenses) whereas other items resulted in considerable disagreement (i.e., managers disagreed on whether the information was favorable or unfavorable for the position of life insurance agent). For example, one item with a high rate of disagreement was "Is presently active in eight outside [community] groups".

Mayfield and Carlson (1966) concluded that interviewers held common and specific stereotypes of the ideal applicant. London and Hakel (1974) studied the impact of interviewer expectations, favourability of information, and interviewee order on the decisions made by 80 introductory Psychology students. They found that interviewers who expected to find high calibre applicants gave higher ratings to applicants.

Kinicki, Lockwood, Hom and Griffeth (1990) found that interviewers based their hiring decisions on different impressions. This evidence suggested that the interviewers had different prototypes of the ideal applicant which directed their attention to different applicant characteristics. More recently, Stevens and Kristof (1995) found that applicants

---

[3] - These are hypothetical job candidates who are typically described on paper and presented to study participants in terms of resume or interview information.

who engage in self-promotion and ingratiation during the employment interview receive more positive interviewer evaluations.

Information About the Applicant

Typical real life interviews are always preceded by the interviewer reviewing applicant qualifications which are determined from resumes and application forms. This information typically includes the following: information about an applicant's home address, last employer, previous work experience, education, and names of individuals who can provide references (Stone and Meltz 1988). Therefore, interviewers enter the interview with considerable knowledge of applicants and possibly with a mental ranking of the "best" candidates. This knowledge combined with the interviewers' prototype of the successful applicant contributes to the interviewer's pre-interview impression.

Interviewer's Pre-Interview Impressions

Several studies have found support for the effect of pre-interview impressions on post-interview ratings of job suitability (Dipboye, Fontenelle and Garner 1984; Dipboye, Stramler and Fontenelle 1984; Hoff-Macan and Dipboye 1990). Dipboye (1982) developed a process model of the interview which encompasses events before and after the actual interview itself. In organizations, interviewers rarely conduct interviews

without some prior information about applicants from their resume, employment tests or even previous screening interviews, and thus they often have preconceived ideas about candidates before the actual interview takes place.

Dipboye argues that post-interview impressions are influenced to a great extent by these pre-interview effects. There has been some support for the positive relationship between pre- and post-interview impressions across the following industries: manufacturing, transportation, public utilities, and retail selling companies (Springbett 1958); the Canadian army (Springbett 1958); a stock brokerage firm (Phillips and Dipboye 1989), and universities (Macan and Dipboye 1988). These studies aggregated correlations across interviewers and found support for the positive relationship between pre- and post-interview ratings. A recent study by Dalessio and Silverhart (1994) found a positive relationship between biodata scores and interviewer ratings which were both positively related to job success outcomes.

Dipboye (1982) describes the pre-interview impression effect as a "behavioural confirmation of expectations". In other words, interviewers have certain expectations about the applicant which they confirm in the interview through the type of questions and manner in which they ask them. For example, Macan and Dipboye (1988) and Liden, Martin and Parsons (1993) found that interviewers with positive attitudes towards

applicants demonstrated more signs of approval both verbally and nonverbally to those applicants.

Dipboye (1982) suggests two mechanisms whereby pre-interview impressions can bias the conduct of the interview. Firstly, positive behaviour of the interviewer towards the applicant motivates the applicant to create a positive impression. Secondly, the interviewer spends most of his/her time discussing the position or engaging in small talk (since he/she is already confident of the applicant's abilities) with the result that nothing negative about the applicant emerges from the interview (Binning, Goldstein, Garcia and Scattaregia 1988). Reciprocal interaction analysis has been proposed by Liden and Parsons (1989) as a method to describe these mechanisms and develop our understanding of the interview process.

A third mechanism identified by Dipboye (1982) and others (Hamilton, Katz and Leirer 1980; Carlston 1980) is the influence of pre-interview information on the interviewer's processing of information. This third mechanism, as explained later, will be one of the foci of this study.

In summary, pre-interview impression effects can influence the outcome of the interview by changing the interaction between the interviewer and interviewee during the interview, by biasing the interviewer's processing of information, or some combination of the two. It should be noted at this time that these effects and the studies mentioned here

were all conducted using unstructured interviews and thus the impact of pre-interview

effects on structured interviews is as yet unknown.

Interview Structure

The typical requirement for interview structure was previously defined in this

paper as: "a series of job-related questions with predetermined answers that are

consistently applied across all interviews for a particular job" (from Pursell et al. 1980,

908). Extrapolating from Dipboye's (1982) self-fulfilling prophecy model, in which the

interviewer's behaviour determines the outcome of the interview by influencing the

interviewee through verbal and nonverbal gestures, interview structure can have an impact

on interview decisions in two ways.

First, it ensures that all interviewers ask the same questions, that these questions

are job-related, and that they are evaluated using the same rating scales. This

standardization of the interview reduces the opportunities for interviewers to ask their

own questions and discourages interviewers from rating applicants on other dimensions

than those identified as important for the job. In other words, interview structure may

eliminate the <u>direct</u> effects of each interviewer's prototype of a successful applicant and

his/her pre-interview impressions by standardizing questions and rating scales. By

extension therefore, there should be less bias in structured interviews than unstructured interviews.

Second, interview structure works indirectly to reduce (and possibly eliminate) the effects of interviewer biases on the storage, retrieval and evaluation of interviewee information. A good analogy for interview structure is that it first provides a clearer view of the target (i.e., the potentially successful performer) and then helps the interviewer to hit the target by focusing him/her on the relevant information. Schwab and Heneman (1969) found higher levels of inter-rater agreement with structured interviews. Maurer and Fay (1988) found that SI questions improved interview rating consistency over interviewer training. Finally, Motowidlo, Carter, Dunnette, Tippins, Werner, Burnett and Vaughan (1992) found that interviewer accuracy depended on the amount of relevant behavioural information about applicants with more information producing greater accuracy.

Several recent meta-analyses (Huffcutt and Arthur 1994; Marchese and Muchinsky 1993; Wiesner and Cronshaw 1988; Wright, Lichtenfels, and Pursell 1989; McDaniel, Whetzel, Schmidt, Hunter, Maurer and Russell 1986) have found structured interviews to have higher validities than unstructured interviews against job performance criteria. Huffcutt and Arthur (1994) found that interview validity increases with interview structure. Mean effect size across all interviews with job performance was .37 which

ranged from .20 (no formal structure) to .57 (same questions, no probes and benchmarked answers). Marchese and Muchinsky (1993) produced a mean corrected validity of .38 with job performance across 31 studies. Wright, Lichtenfels and Pursell (1989) reviewed 13 studies and found the structured interview to have mean validity of .39 (corrected for unreliability). These results support both the SI (Latham and Saari 1984; Weekley and Gier 1987) and PBDI (Janz 1982; Orpen 1985) interviews. A more recent study (Green, Alter and Carr 1993) using behaviour description questions and behavioural scoring anchors for patrol sergeants produced similar results to the meta-analyses with an inter-rater reliability of .65 and validity coefficients of .42 and .44 (for consensus and panel average scores).

Wiesner and Cronshaw (1988) reviewed a large number of international studies of the interview and collected 151 validity coefficients to analyze (relating the interview score with a work related criterion). On the basis of their study, structured interviews produced a mean effect size of .63 with individual interviewers and .60 with board interviews (i.e., at least two interviewers) (corrected for direct range restriction and criterion unreliability). Wiesner and Cronshaw (1988) also found that there was a considerable amount of unaccounted-for variance in the structured interviews (73% for individual and 39% for board interviews). The unexplained variance in the meta-analysis

points to the existence of other factors that impact on interview validity beyond interview structure.

The most compelling argument for increasing research on the structured interview over the unstructured interview is based on the findings of Wiesner and Cronshaw (1988). They were able to account for all the variability in unstructured interview validity with statistical artifacts. This suggests that moderators of unstructured interview validities have minimal impact and thus unstructured interview research with the purpose of identifying factors that impact on validity is not a meaningful endeavor. Moreover, Wiesner and Cronshaw (1988) were unable to account for all of the structured interview validity variance which suggests that moderators do exist and that further research to identify the factors that moderate structured interview validity is warranted.

One of the first studies to assess factors that moderated structured interview validity was Wiesner and Oppenheimer (1990) who studied the effects of note taking versus rating scales on structured interview decisions. Using videotapes, 94 senior undergraduate business and MBA students were asked to rate applicants on the basis of overall suitability and likelihood of hire. There were four conditions in the study: observe only, note taking, numeric scales, and scoring guides. Interviewers using scoring guides made more accurate predictions of job performance. Moreover, consistent with previous

research (Peterson and Pitz 1986), Wiesner and Oppenheimer (1990) found that statistical predictors outperformed subjective predictors.[4]

In addition, Wiesner and Oppenheimer (1990) had subjects identify relevant and nonrelevant information they recalled from the interviews. The Numeric Scales and Scoring Guides groups recalled more non-relevant and relevant information than the Observe-only and Note-taking groups, and both of the former groups had higher validities than the latter. Wiesner and Oppenheimer (1990) suggest that the scales and scoring guides kept the incorporation of non-relevant information from impacting on the final interview scores thus increasing the validity of the interview for those groups. This study did not determine which aspects of the non-relevant information was used in the final interview ratings for the Observe-only and Note-taking groups, nor if scoring guides would improve the validity of unstructured interviews.

Wiesner (1989) explored the effects of job relevance, timing and rating scales on the validity of structured interviews. He found that job relevant questions, specific scoring guides and ratings made immediately after each question improved the reliability and

---

[4] - Subjective predictors, also known as clinical predictors, involve an overall judgement which is made based on interviewer's mental process of combining all of the information collected across the interview questions. Statistical predictors, also known as mechanical predictors, combine the ratings from each question additively or as an average (with or without weights) to produce an overall rating.

validity of the structured interview. Wiesner (1989) found interview validity to increase as job relevance increased, rating scale increased, and timing of ratings decreased.

Wiesner (1989) found that 72.7% of the variance in job performance was accounted for by the interview. This percentage may be inflated because the research design reduced error variance by limiting the number of interviewees who were observed. Even if we accept this percentage, 27.3% of variance in structured interview validity was still unaccounted for. Pulakos, Schmitt, Whitney and Smith (1996) found that structuring the interview (including past behaviour questions and behaviourally anchored rating scales) and extensive interviewer training eliminated all individual interviewer variability in their ratings.

In conclusion, the general process of collecting job and applicant information prior to the interview creates circumstances that produce interviewer conceptualizations of the "perfect" applicant and preliminary expectations about how well each applicant fits with the prototype. However, these factors may only impact on the unstructured interview process and structuring the interview may nullify any of their moderating effects. There are also situational factors that affect the interview as well as individual differences across interviewers and applicants.

Situational Factors

Situational factors pertain to the context in which the interviews are conducted but not the interview process itself. Therefore, these factors include the following: the environment in which the interview is conducted (physical, time, and external pressures), the interview's purpose, task clarity, decision making responsibility, and decision risk. All of these factors can impact on the decisions made during the interview, but are external to the interview process itself.

The Interviewing Environment

Dipboye and Gaugler (1993) suggested that time pressures, physical setting and hiring quotas can influence interviewers' pre-interview impressions, how interviewers conduct sessions, and the cognitive decision making processes that produce the final decision. There is some support in the decision making literature for these relationships. For example, Carlson, Thayer, Mayfield and Peterson (1971) found that inexperienced interviewers were more likely to accept a less qualified applicant than experienced interviewers when quotas were stressed.

The physical environment can impact on interviewer decision making in several ways. First, distractions (such as noises or interruptions) can interrupt information recording or integration (Siegel and Steele 1980; Schuh 1978). Second, an interview

room that is not conducive to the interview process can reduce the amount of information applicants provide. Also, any potential items that may be distracting in the interview room (e.g., a golf trophy in one's office) can provide for non-job-related information to impact on the interviewer's decision. Wright (1974) found that time pressure impacts on interviewer ratings as well.

Contrast effects occur when subsequent applicant ratings are inflated or deflated because of previous applicants who were extremely positive or negative. Evidence for this effect has been mixed in the literature with some studies finding strong support (Carlson 1967; Rowe 1967; Cesare, Dalessio, and Tannenbaum 1988; Heneman, Schwab, Huett and Ford 1975) and other studies finding support with minimal effects (Hakel, Ohnesorge, and Dunnette 1970) or no support at all (Landy and Bates 1973; Hakel, Dobmeyer, and Dunnette 1970). One study revealed that the impact of contrast effects was influenced by the number of good or bad preceding applicants (Wexley, Yukl, Kovacs, and Sanders 1972).

In addition to the empirical support for the influence of these situational factors, one can conceptualize what occurs when an organization must hire an applicant from a current applicant pool. For example, some positions may be extremely difficult to fill and after a long period of searching unsuccessfully an organization may be willing to lower its standards in order to find a qualified applicant. Therefore, if the interviewer knows that

he/she must hire one of the applicants, this will impact on his/her impressions of the applicants, especially ones that are almost qualified.

Interview Purpose

Interviews have a variety of uses in organizations. They are used for recruitment, selection, promotion and performance appraisal, however little research has been conducted studying the impact that interview purpose has on decision outcome. As Eder and Buckley (1988) suggest, interviewers engage in different behaviours when conducting recruitment/selection interviews as compared to simple selection interviews. In the former, interviewers must present themselves and the organization in a positive manner while at the same time attempting to evaluate the applicant. In the latter, interviewers simply have to collect information to evaluate the applicants. Clearly, these two scenarios are very different both from the demands placed on the interviewers and the opportunities for the applicants to influence the decision making process (e.g., in a recruitment/selection interview, applicants have a greater opportunity to ask questions about the job).

Tullar, Mullins and Caldwell (1979) found that interviewers focused differentially on recruitment versus selection based on the quality of applicants. With low quality applicants they focused on selection and with high quality applicants they focused on

recruitment. As a result, the validity of interviewer decisions varied based on interviewer's perceptions of applicant quality.

Task Clarity

Peters, O'Connor and Eulberg (1985) identified two components that comprise task clarity: the array and clarity of task demands and the extent of interviewer preparation and training. The array and clarity of task demands pertains to the information interviewers receive in order to accomplish the task. For example, interviewers who are provided with little job information or guidelines about qualifications would have a more difficult time discriminating between applicants due to a lack of task clarity. Osburn, Timmrick and Bigby (1981) suggested that discriminability is positively influenced by the specificity of evaluation criteria. Research suggests that task clarity does influence the reliability of decisions made after employment interviews (Langdale and Weitz 1973).

One can conceive of interview structure as improving the task clarity for interviewers as they provide a list of specific questions and rating scales for evaluation. Also, to the extent that the interviewer's decision is influenced by his/her prototype of the successful candidate as research suggests (Sydiaha 1961; Bolster and Springbett 1961; Hakel, Hollman and Dunnette 1970), interview validity should increase as the interviewer's prototype and the evaluation criteria converge (the mechanism being improved task

clarity) (Wyer and Srull 1981). One issue that emerges from this is the question of whether the interviewer's prototype or schema can be changed, and if presenting job information, evaluation criteria or structuring the interview can facilitate this change.

Finally, job complexity can have an impact on the decision making process because it increases the need for task clarity, especially if the interviewer is given little or no guidance about the job and evaluation criteria. Tullar, Mullins, and Caldwell (1979) found that interview length and job complexity interacted to reduce interview validity. In other words, with complex jobs interviewers took more time in the interview and strayed away from job-relevant questions which reduced the validity of the interview decisions.

The other issue with respect to task clarity is the extent of interviewer preparation and training. Structuring the interview, in addition to increasing task clarity, also facilitates interviewer preparation, although it provides no guarantee that the interviewer will be prepared. Research indicates that some training reduces interviewers' rating errors (Wexley, Sanders, and Yukl 1973; Latham, Wexley, and Pursell 1975; Fay and Latham 1982) and can improve validity (Dougherty, Ebert and Callender 1986).

Decision Making Responsibility

Decision making responsibility refers to the extent that an interviewer expects his/her opinion to be used in the final decision and his/her accountability. For example,

individual interviewers have more decision making responsibility than group interviewers, especially if all group members' opinions are weighted equally. Moore and Lee (1974) found no difference in interview judgement between individual and group interviews. In contrast, Rozelle and Baxter (1981) found that interviewers who were told their decisions would be discussed in a staff meeting and that applicants could review their reports provided descriptions that more accurately reflected applicants than those who were told their ratings would be confidential.

Decision Risk

Decision risk is defined as the interviewer's perceived cost (personal and/or organizational) of making a mistake. For example, personal decision risk increases if the interviewer is likely to be the job candidate's supervisor (Dipboye and Gaugler 1993). Although risk has not been studied directly, one study defined risk in terms of accountability and found that differences in risk can impact on interviewer decisions. Gordon, Rozelle, and Baxter (1988) found that subjects in high accountability conditions produced more extreme ratings than those in low accountability conditions.

In summary, there are many factors to consider when attempting to understand interviewers' decisions about job applicants. Although these factors have been identified independently in research, their combined effect has yet to be studied as well as their

impact on all aspects of the interview process from pre-interview preparation to interview outcomes.

Interviewer Characteristics

Interviewer characteristics refer to the demographic profiles of interviewers which can serve as a proxy for their biases. For example, an older more experienced interviewer (with age used as a proxy) may favour applicants who are older and more experienced as well. Furnham and Burbeck (1989) found that interviewers with more job experience rated applicants more strictly. Interviewers can differ in many ways that may impact on decision accuracy.

Leonard (1976) studied situations in which evaluation criteria were multi-dimensional. He found that interviewers who were able to incorporate more interviewee attributes into their decision making, as opposed to those who relied on one decision factor, were also more likely to be influenced by the "similar to me" effect. Interviewers also differ on how they weight different types of information (Hakel, Dobmeyer and Dunnette 1970), and in the types of information processing errors that they make (Hakel 1982; Ferguson and Fletcher 1989).

Interviewers evaluate applicants more favourably when they are similar (e.g., race, age) (Ghee, Rozelle and Baxter 1986; Lin, Dobbins, and Farr 1992; Prewett-Livingston,

Feild, Veres III, and Lewis 1996), when they like the applicant (Dipboye, Fromkin, and

Wiback 1975; Rand and Wexley 1975), and when they have a positive pre-interview

impression (Huguenard, Sager, and Ferguson 1970).

Applicant Characteristics

Applicant characteristics provide interviewers with information that they may use

in their hiring decisions via stereotypes. For example, women are considered to be more

compassionate than men and thus an interviewer may feel that a woman is better suited to

the job of social worker than a man.

The outcomes of research studying the effects of age on interviewer evaluations of

applicants have been mixed. In several studies, older applicants received lower job

suitability ratings than did younger applicants from undergraduates (Rosen and Jerdee

1976; Avolio and Barrett 1987) and employment managers (Haefner 1977). In contrast,

Arvey, Miller, Gould and Burch (1987) found that older applicants received higher scores

than did younger applicants. More recently, two studies did not find any age effects (Lin,

Dobbins and Farr 1992; Blankenship and Cesare 1993).

Similarly, the findings for applicant sex have been mixed with some studies

revealing higher job suitability ratings for men (Dipboye, Fromkin, and Wiback 1975;

Dipboye, Arvey and Terpstra 1977; Haefner 1977; Zikmund, Hitt and Pickens 1978), and

other studies finding positive effects for women (Elliott 1981; Parsons and Liden 1984), although the differences may be due to gender stereotypes about the job (Shaw 1972; Cash, Gillen, and Burns 1977). Often the effect of gender is negligible (Arvey 1979; McDonald and Hakel 1985).

The same pattern has been found for race effects with majority white males being rated more favourably in some studies (Wexley and Nemeroff 1974; Haefner 1977; Rand and Wexley 1975) and unfavourably in other studies (Parsons and Liden 1984; Singer and Eder 1989) when compared with minority groups. The long length of time between these studies (7 years) may indicate that there has been an attitudinal shift from earlier studies which now favours minority applicants. Similar to the impact of gender, these effects are often negligible as well (Arvey 1979; McDonald and Hakel 1985).

In addition, studies have examined the effects of different applicant characteristics on job suitability such as, applicant attractiveness (Riggio and Throckmorton 1988; Heilman and Saruwatari 1979), and the masculinity of dresses for women (Forsythe, Drake, and Cox 1985). The following interaction effects have also been found: for applicant age and sex of the interviewer (Raza and Carpenter 1987), applicant age and interviewer accountability (Gordon, Rozelle, and Baxter 1988), applicant attractiveness and job type (Beehr and Gilmore 1982), and applicant attractiveness and gender (Shaw 1972).

Overall, the above results suggest that it is still unclear as to how applicant characteristics impact on interviewer suitability ratings. As the findings by Shaw (1972), Cash, Gillen, and Burns (1977), and Singer and Eder (1989) suggest, a deeper understanding of the effects of applicant characteristics may require an understanding of interviewers' job stereotypes as well as other stereotypes (i.e., gender, age, race). Moreover, Singer and Sewell (1989) found that there was less age discrimination when interviewers had more information about the applicant.

Interview Process

The focus of this thesis is on the cognitive decision making process that occurs during the interview and the impact this process has on the reliability, validity, accuracy and bias of the interview decision. The finding that structured interviews produce more accurate decisions in terms of validity (Huffcutt and Arthur 1994, Marchese and Muchinsky 1993; Wiesner and Cronshaw 1988; Wright, Lichtenfels and Pursell 1989; and McDaniel et al. 1986) creates an opportunity for us to understand the cognitive processes that occur in both structured and unstructured interviews. In other words, how does interview structure affect the cognitive decision making process of the interviewer to produce more accurate decisions?

Studies have found both primacy and recency effects of information on single evaluative judgements made by interviewers (Farr 1973; Farr and York 1975; Peters and Terborg 1975). Belec and Rowe (1983) found that applicants were more likely to receive favourable ratings when positive information appeared first in the interview. This finding was consistent even if negative information came after the positive information. In addition, Johns (1975) found that the presentation order of information affected the competency of interviewer decisions.

Dipboye (1982) suggests that these effects occur because of self-fulfilling prophecies. In other words, interviewers establish an impression before or early in the interview and then attempt to find evidence to support it. A recent study of corporate interviewers by Dougherty, Turban and Callender (1994) found support for the interviewers' use of confirming strategies in the interview. Rowe (1984) suggests that these effects occur because of interviewer attributions. This hypothesis was supported empirically by Tucker and Rowe (1979). Srull and Wyer (1980) found that the encoding of information affected later judgements, therefore if information was encoded in a positive way (i.e., positive impression) then this had a greater impact on the rating of suitability.

In addition, many studies have supported the finding that positive non-verbal behaviours (e.g., eye contact, appropriate pauses) can influence interviewer suitability

ratings (Anderson 1960; Keenan and Wedderburn 1975; Amalfitano and Kalt 1977; Imada and Hakel 1977; Young and Beier 1977; Tessler and Sushelsky 1978; McGovern and Tinsley 1978; Parsons and Liden 1984; Gifford, Ng and Wilkinson 1985; Forbes and Jackson 1980; and Anderson and Shackleton 1990).

Some studies have shown that negative information carries greater weight in interviewer decision making than positive information (Springbett 1958; Hollman 1972; London and Hakel 1974). Constantin (1976) found that unfavourable information had a greater negative impact on interviewer ratings when it was perceived to be job relevant by interviewers (as opposed to not relevant) but favourable information had an impact all the time (i.e., independent of its perceived job relevance).

Alternatively, Snyder and Cantor (1979) and Snyder and Swann (1978) found support for the use of confirming strategies or hypothesis testing in the interview but this was not replicated inside the laboratory (McDonald and Hakel 1985) or outside (Sackett 1982).

Other interview process factors have been found to impact positively on interviewer suitability ratings such as: applicant assertiveness (Dipboye and Wiley 1977); applicant verbal behaviour (Parsons and Liden 1984; Rasmussen 1984) and applicant impression management (Giacalone and Rosenfeld 1986; Gilmore and Ferris 1989; Fletcher 1989).

Interview Outcomes

The outcome of the interview is ultimately a decision pertaining to a candidate's suitability for a particular job (in laboratory studies) or to hire a candidate (in field studies). One conclusion that can be drawn from this literature is that the interview itself involves many complex behaviours and cognitive processes interacting in many ways that have yet to be completely measured or understood. Since the key outcomes are the products of interviewers' decisions and the accuracy of those decisions, these need to be understood in the context of the employment interview.

CHAPTER 3:

A MODEL FOR THE INTERVIEW

A model is proposed for the interview which considers pre-interview information, interviewer behaviours, cognitive processes and subsequent decision outcomes. Based on the literature reviewed here, the accuracy of interviewer judgements of applicant suitability depend on several things. First, the pre-interview impression can impact on the decision by impacting on the interviewer's behaviour and cognitive information processing. Second, the opportunity for the interviewer to sample job relevant behaviour of applicants, which depends on the interview process, is important for decision accuracy. Third, how information is encoded and retrieved by each interviewer can influence the decision. The encoding and retrieval of information is influenced by the type of interview used and the rating format.

Since the pre-interview impression is developed before the interview, it can impact on the final decision by controlling the interviewer's use of questions, verbal responses and nonverbal cues to influence the applicant's behaviour and subsequent sampling of relevant job behaviours for that applicant. This is referred to as the self-fulfilling prophecy by

Dipboye (1982) and will be referred to hereafter as the behavioural effect. The pre-interview impression can also impact on the interviewer's perception of the applicant which may bias the encoding and retrieval of information with subsequent impact on the hiring decision. This will be referred to as the cognitive effect and it is the focus of the present study.

This model can provide an explanation for many of the findings in the interview literature. Since all interviews begin with the interviewer reviewing job and applicant information prior to the interview, interviewer biases (i.e., age, gender, job stereotypes) can impact on all decisions via both behavioural and cognitive paths. Structured interviews improve the accuracy of decisions by considerably reducing the behavioural path and improve the sampling of job relevant behaviour (with job relevant questions). Rating scales improve decision accuracy by improving the sampling of job relevant behaviour. The mechanical combination of ratings improves decision accuracy by eliminating the retrieval effect of any biases. Therefore, although the interviewer may still have a bias towards the applicant, the process of summing up the results to each question removes the interviewer (and hence his/her biases) from influencing the final interview rating.

Therefore, interviews that are structured, use rating scales and mechanically combine ratings should have the highest reliability and validity and least amount of bias

because they reduce the effects of biases that are not related to job performance. They achieve this by reducing the cognitive processing required by the interviewer and interviewer variability in the rating process. The only place in which biases can have an impact is in the encoding of information to provide the rating, with the impact of the bias being reduced for the behaviourally anchored scales as compared to the graphic rating scales.

Rationale for the Current Study

The focus of this thesis is the impact of pre-interview information on interview decision accuracy and how interview structure moderates this relationship. Previous research has consistently demonstrated the superior validity of structured interviews (Huffcutt and Arthur 1994; Marchese and Muchinsky 1993; Wright, Lichtenfels and Pursell 1989; Wiesner and Cronshaw 1988; McDaniel et al. 1986), however little research (other than Wiesner and Oppenheimer 1993; Wiesner and Oppenheimer 1990; Wiesner 1988) has explored the reasons for the greater accuracy of structured interviews. Moreover, these few studies have not addressed the impact of interview structure on bias in interview decisions. A recent meta-analysis found that reliability was one factor that accounted for the high validity of structured interviews (Conway, Jako, and Goodman 1995).

In addition, the singular focus on the interviewer's cognitive decision making process is necessary because this process represents the interaction of all related variables (e.g., pre-interview impressions, interview type, applicant characteristics, interviewer characteristics, situational constraints) that influence the interviewers decision. The essential goal of the interview is to identify applicant knowledge, skills and abilities (KSAs) that relate to the job (as determined from job analysis information). It follows that a greater amount of KSA information should increase the accuracy of the interview decision (Wernimont and Campbell 1968). The best approach for developing our understanding of the interview is to start with the interviewer's cognitive decision making processes and control for outside factors by conducting a laboratory study.

Hypotheses

The following hypotheses have been drawn from the interview, measurement and cognition literature. Previous research (Springbett 1958; Tucker and Rowe 1979; and Dipboye, Stramler and Fontenelle 1984) found a positive relationship between interviewer's pre-interview evaluations and their post-interview decisions.

Dipboye (1982) described this phenomenon as the "self-fulfilling prophecy" which refers to interviewers behaving, observing, and interpreting information differently during the interview based on their pre-interview impression (i.e., positive or negative). For

example, an interviewer who considered an applicant to be well qualified for a position (based on a resume and application form) interacted more positively (e.g., smiled more) and perceived that applicant more favourably (e.g., ignored negative information) during the interview. Therefore, Hypothesis 1a can be stated as:

H1a: The pre-interview rating will be positively correlated with the post interview rating.

Based on the theory presented by Dipboye (1982) the pre-interview impression is carried over into the post-interview decision. Therefore, to the extent that this impression is removed from the interview ratings (i.e., by using job relevant questions and a mechanical combination of ratings) then it should have less of an impact on the post-interview decision. Therefore, Hypothesis 1b can be stated as:

H1b: The correlation between the pre-interview impression and the final interview decision will be lower for the job relevant interview with a mechanical combination of ratings than the general interview with one overall rating.

The corollary to H1a is that a positive pre-interview rating will lead to a positive post-interview rating and a negative pre-interview rating will lead to a negative post-interview rating. Therefore, Hypothesis 2 is:

H2:   Interviews preceded by positive pre-interview information, will have higher (i.e.,

more positive) average post-interview ratings than interviews preceded by negative

pre-interview information.

Hypotheses 3 through 5 have four components to them (a, b, c, and d) which refer

to reliability, validity, accuracy and bias respectively. For this study, reliability is measured

using generalizability coefficients (Shavelson, Webb and Rowley 1989) and validity is

measured using point-biserial correlation coefficients. Accuracy is measured using

Cronbach's (1955) index of accuracy and bias is measured using differential validity

(Boehm 1972).

For the present study, differential validity refers to significantly different validity

coefficients between applicants with neutral versus negative pre-interview information.

The negative information, which was not related to the job, represents a potential bias that

if used in the selection decision, would reduce the validity of that decision.

Previous meta-analyses have consistently demonstrated that job relevant questions

increase the reliability and validity of interviews (Conway, Jako, and Goodman 1995;

Huffcutt and Arthur 1994; Marchese and Muchinsky 1993; Wright, Lichtenfels and

Pursell 1989; Wiesner and Cronshaw 1988; McDaniel et al. 1986). Furthermore,

Kataoka, Latham and Whyte (1995) found that interview structure (i.e., situational questions with rating scales) minimized the impact of being provided with an overall assessment of the applicant (as either "good" or "poor" and referred to as anchoring) as it affects interviewer decisions. Therefore, the following hypothesis can be made:

H3: Interview decisions based on a common set of job relevant questions will be more reliable (3a), valid (3b), accurate (3c) and less biased (3d) when compared to those with a common set of general questions.

In addition, recent work has demonstrated that interview decisions are more reliable and valid when a specific rating scale is used to evaluate interviewee responses as compared to simple observation or note taking (Wiesner and Oppenheimer 1990; Wiesner 1988). Vance, Kuhnert and Farr (1978) found greater reliability with behavioural scales over graphic rating scales. Moreover, Kataoka, Latham and Whyte (1995) concluded that the behavioural rating scales used with the situational interview reduced the impact of anchoring effects. Therefore, hypothesis 4 is:

H4: Interview decisions based on a common set of behavioural rating scales for scoring each question will be more reliable (4a), valid (4b), accurate (4c), and less biased (4d) when compared to those with no rating scales.

Research has been conducted in both clinical and industrial psychology (Sawyer 1966; Einhorn 1972; Peterson and Pitz 1986; Wiesner and Oppenheimer 1990) supporting the mechanical combination of individual ratings (i.e., simply summing up or averaging the ratings on individual interview questions to produce a total score for each interviewee) over the clinical combination (i.e., an overall judgement made by the interviewer after the interview). Therefore, hypothesis 5 can be expressed as follows:

H5: Interview decisions based on a common set of question ratings combined mechanically will be more reliable (5a), valid (5b), accurate (5c) and less biased (5d) than those with a common set of question ratings combined clinically.

CHAPTER 4:

METHOD

This study focuses on the impact of question type (job relevant versus general questions), rating process (observe only versus rating scale), combination of ratings (clinical versus mechanical) and accuracy of pre-interview impression (accurate/positive or inaccurate/negative) on the subsequent reliability, validity, accuracy and bias of interview ratings.

All participants observed videotapes of actual interviews of three social workers who were working as social welfare caseworkers in the social services department of a mid-sized regional government and volunteered for the study (Wiesner 1988). Although these were actual interviews, the interviewees already held positions with the organization. These videotapes and job performance questionnaires were developed by Wiesner (1988). The criterion data were obtained in a field setting (i.e., performance ratings) for comparison to the predictor data collected in this study (i.e., interview ratings of applicant suitability).

In comparison with actual interviews, "interviewers" in this study would approximate members of an interview panel who observe the interview but do not ask

questions. Therefore, they can have no impact on how interviewees respond but they can be influenced by their responses.

Although this design appears to be somewhat unorthodox it has several benefits over traditional designs (i.e., actual interviews). First, it provides an opportunity to assess the impact of the identified variables on interviewers' decisions while minimizing the number of interviews that need to be conducted. In order to conduct this study with real interviews under realistic circumstances (i.e., a maximum of 4 panel members per interview) a very large number of interviewees would be required as well as a large number of interviews (or the sample would have to dramatically decreased). Therefore, the loss of realism is offset by the increase in feasibility and statistical power. Second, the use of videotapes allows for more stringent control of the many extraneous variables that can impact on interviewer decisions. Third, the role of interviewers as observers provides the opportunity to assess the impact of the cognitive decision making process on interviewer decisions independent of the interviewers' ability to influence interviewees'.

A total of 217 senior level (3rd and 4th year) undergraduate Commerce students from a medium sized Canadian university observed the three applicant videos. The participants were randomly assigned to one of the eight conditions and presented with the interviews using job relevant questions or the interviews using general questions. In every condition, the subjects received a package with instructions that incorporated the pre-

interview impression and rating process. (Pre-interview and interview materials are

presented in Appendices A through J.) All of the possible comparisons are presented in

Table 1.

TABLE 1.-- Differences across conditions

Condition	Pre-interview impression	Type of questions	Rating process	Combination of ratings
1.	Accurate/Positive	Job Relevant	Observe-Only	N/A
2a.	Accurate/Positive	Job Relevant	Rating Scale	Clinical
2b.	Accurate/Positive	Job Relevant	Rating Scale	Mechanical
3.	Inaccurate/Negative	Job Relevant	Observe-Only	N/A
4a.	Inaccurate/Negative	Job Relevant	Rating Scale	Clinical
4b.	Inaccurate/Negative	Job Relevant	Rating Scale	Mechanical
5.	Accurate/Positive	General	Observe-Only	N/A
6a.	Accurate/Positive	General	Rating Scale	Clinical
6b.	Accurate/Positive	General	Rating Scale	Mechanical
7.	Inaccurate/Negative	General	Observe-Only	N/A
8a.	Inaccurate/Negative	General	Rating Scale	Clinical
8b.	Inaccurate/Negative	General	Rating Scale	Mechanical

The videotaped applicants were presented in random order to eliminate any

potential contrast effects. Each session took approximately 50 minutes to complete;

subjects were required to view all of the interviews and complete corresponding rating

forms. Overall, the design of the study can be considered similar to a concurrent validation design with the exception that there were only 3 interviewees in this study and many interviewers.

Material Preparation

Measures

The structured interviews and job performance rating forms were developed based on a thorough job analysis using the Position Analysis Questionnaire (PAQ) (McCormick, Jeanneret and Meecham 1972) and Critical Incidents Technique (Flanagan 1954). In addition, the PAQ items were reviewed with incumbents and supervisors to identify those considered to be most critical for the job (Wiesner 1988). The job relevant questions for the Patterned Behaviour Description Interview (PBDI) (Janz, Hellervik and Gilmore 1986) and the five point rating scales were developed from the job analysis data. The general questions were developed from lists of popular interview questions (e.g., what is your greatest accomplishment?).

Behaviourally anchored ratings scales (BARS) (Smith and Kendall 1963) which provide descriptions of job behaviours with different ratings on a five point scale were used to measure job performance (Wiesner 1988). A total of 26 behaviours, categorized as Interpersonal Relations (9 items), Organizational Skills and Behaviours (10 items), and Behavioural Predispositions (7 items) were rated by supervisors. These were combined

and divided by 5 to produce an overall job performance score (ranging from 1 to 5) which was used as the criterion for this study. (For a more detailed description of the development of interviews and performance measures refer to Wiesner 1988.)

The reliability of the job performance appraisal was determined based on evaluations from the current and previous supervisors of nine job incumbents. Cronbach's alpha for the ratings was high (.976, n=18). For the evaluations of the three incumbents used for this study, the inter-rater reliability of the job performance ratings was .730 (based on generalizability analysis), which was considered acceptable. However, due to concerns expressed by supervisors that there were differences between current and previous job performance ratings, only the current job performance ratings were used (Wiesner 1989). The average job performance ratings were high (4.02, S.D.= 0.52, Range = 2.6 to 5.0) indicating that the ratings were severely restricted in range. The mean job performance ratings for the three interviewees were 3.59, 4.11 and 4.71 out of 5 (Wiesner 1988).

Videotaped Interviews

Three videotapes (representing employees from below average, average and superior job performance categories) of welfare caseworkers were used in the study. The caseworkers voluntarily participated in the interviews and were told to assume that they were moving to a new geographical location and had to apply for a similar job to the one

they had. All three caseworkers were female. No information or rehearsal preceded the interviews. The interviews were conducted by a female graduate student at the University of Waterloo who had previous experience in a social services department and had taken an interviewing course. Prior to conducting the interviews she rehearsed her role with the experimenter until mastery was achieved. Both the job relevant and general questions were represented in the interview script from which the interviewer read in order. Therefore, in order to provide separate videos with job relevant and general questions, they were edited and new tapes were created with questions of only one type across the 3 applicants (Wiesner 1988).

Present Study

The manipulation for the present study occurred prior to each interview in the pre-interview impression stage of the design. Different subject groups received information that was consistent or inconsistent with the reported job performance of the interviewee (i.e., reference check information that either did or did not correspond to her job performance). This information was varied in order to assess its effect across question type, rating process, combination of ratings, interview reliability and validity. The negative pre-interview information which was not directly job related represented bias. Different undergraduate classes participated in the study with members randomly receiving one of

the eight packages. An overview of the study with corresponding groups and time

estimates is presented in Table 2.

TABLE 2.-- Description of subject tasks for each stage of the study

Stage	Tasks	Time (min)
1.	Complete consent form	5
2.	Read instructions, job description and rating scales	7
3.	Review interview questions	3
4.	Read resume and reference check	5
5.	Watch video and provide appropriate ratings	21
6.	Provide overall ratings	2
7.	Complete demographics profile	3
8.	Debrief	4
	Total	**50**

Note: Steps 4, 5 and 6 are repeated for each applicant.

Pre-Interview Materials

In Stage 1, prior to the viewing of the videotapes, subjects received a consent form

briefing them on the purpose of the study which was to determine how accurate

interviewers are in predicting job performance. The form stated that the videotapes

contained material edited from actual employment interviews and subjects' predictions

would be compared to job performance ratings. This was done to ensure voluntary

participation, to explain the edited breaks in the interview, and to encourage subjects to

focus on the task and provide accurate ratings.

Pre-Interview Impression

After subjects signed the consent forms they received one of the 8 pre-information

packages which included instructions, job descriptions, resumes, reference checks and

rating scales (Stage 2). Eight packages were developed to represent the combination of

pre-interview information (2), question type (2), and rating options (2). Four of the

conditions contained resumes and reference checks that were consistent with job

performance (positive for two applicants and negative for one), and the other four

contained resumes and reference checks that were inconsistent with job performance for

two applicants and consistent for one applicant. The applicant that had the same

information presented across all conditions was used as a control. These packages are presented in Appendices B through I.

The three packages were pilot tested with a sample of college students to ensure that the negative information clearly demonstrated an effect on evaluations of applicants prior to the interview. The mean rating (on a 5 point scale) for the package with the negative information was 3.19 (S.D. = .68) and for the other two packages with no negative information was 3.93 (S.D. = .68) and 4.15 (S.D. = .60). This difference was significant (F=17.05, p<.001, N=27).

Presentation of Videotapes

Participants were provided with an overview of the procedure (instructions are presented in Appendix A) and one of the eight packages. Once all of the pre-interview materials had been reviewed and subjects had read over the resume and reference check information for the first applicant and the interview questions and rating scales (Stage 4), then the videotape for the first applicant was presented. This process was repeated for each applicant. The 3 applicants were presented in random order.

The order of videos and the selection of conditions was random. All individuals were randomly assigned to a group. Groups 1,3,5 and 7 were asked to simply observe the interview without taking any notes. Groups 2, 4, 6 and 8 were asked to provide ratings on the forms that they were given (Stage 5). After each interview, the videotape was stopped

and subjects provided their overall ratings of applicant suitability (Stage 6). After all of the interviews had been observed and all ratings were provided, the subjects were asked to fill in their demographics (Stage 7).

The data were collected over a 10 day period after which all subjects were debriefed (Stage 8). In addition, they were told the true purpose of the study and asked not to disclose it to other students in case additional students would be required.

CHAPTER 5:

RESULTS

The data were analyzed using SPSS for Windows. The data file contained the following variables: identification number, experimental condition, pre-interview rating, clinical post-interview rating, mechanical post-interview rating, and job performance rating. With each interviewer providing ratings on all the three applicants, the maximum number of data points could be as high as 651 (i.e., 217 interviewers multiplied by 3).

Sound Quality

As this was the first administration of the videotapes to a larger group (i.e., Wiesner 1989 presented the videos to groups of 5) the sound quality was measured by asking participants to estimate the percentage of responses they heard from each applicant. The mean percentage of responses heard was 71.5% (standard deviation = 19.1) for Applicant #1, 79.8% (standard deviation = 14.7) for Applicant #2, and 85.7% (standard deviation = 12.2) for Applicant #3. The difference in sound quality was significant across all applicants (between Applicant #1 and #2 $t = 8.8$, $df = 215$, $p < .001$; #2 and #3 $t = -6.21$,

df = 215, p<.001; #1 and #3 t = -11.1, df = 215, p<.001). Furthermore, sound quality was confounded with job performance as they were significantly correlated (.35, N=648, p<.001).

Since it was likely that poor sound quality would result in less accurate ratings, all participants who reported hearing less than 75% were removed from the analysis. The majority of these participants were removed because of the poor sound quality only for applicant #2, therefore it was also decided that applicant #2 be removed from the study which allowed more participants to remain in the analysis. For the reduced sample, the mean percentage of responses heard across both applicants was 88.2 (standard deviation = 8.0). The mean percentage of responses heard for Applicant #1 was 86.7 (standard deviation = 8.4) and for Applicant #3 was 89.7 (standard deviation = 7.4). The difference in sound quality was still significant between these two applicants, but the difference was considerably smaller. The correlation between sound quality and job performance was reduced as well (.19, N=300, p<.001). The implications of this finding will be addressed in the discussion.

Sample Demographics

A total of 217 Commerce Undergraduate students voluntarily participated in the study to receive a bonus mark in their course for participation. The videotapes were

presented during the 50 minute class with different classes viewing one of the eight

randomly sorted videos. The sample was approximately evenly distributed across the 8

conditions ranging from a low of 25 for Condition 6 to a high of 31 for Condition 4. After

the removal of any participants who could not hear at least 75% of the applicant responses

the total sample size was reduced to 150 participants ranging from 15 in Condition 7 to 25

in Condition 3. These numbers were considered acceptable for the analysis. The

breakdown of participants is presented in Table 3.

TABLE 3.-- Sample demographics by condition

#	Condition	Original N	Reduced N	% Reduction
1.	Accurate/Job Relevant/Observe Only	28	22	21.43%
2.	Accurate/Job Relevant/Rating	28	19	32.14%
3.	Inaccurate/Job Relevant/Observe Only	29	25	13.79%
4.	Inaccurate/Job Relevant/Rating	31	17	45.16%
5.	Accurate/General/Observe Only	26	17	34.62%
6.	Accurate/General/Rating	25	19	24.00%
7.	Inaccurate/General/Observe Only	26	15	42.31%
8.	Inaccurate/General/Rating	26	16	38.31%
	Total	**219**	**150**	**31.50%**

Note: Only participants who heard less than 75% of the responses from each
applicant were removed. The data for Applicant #2 were removed as well.

Gender

The sample was approximately equally divided by gender with 59.3% of the sample being male and 40.7% being female. Gender differed significantly between conditions (Chi-Square = 14.98, df = 7, p<.05). The largest difference was for Condition 3 which had 5 males (20.0%) and 20 females (80.0%). For male interviewers the correlation between interview ratings and job performance collapsed across all conditions was .27 for clinically combined ratings (N = 178 based on two interview ratings per participant, p<.001) and .30 (N = 66, p<.01) for mechanically combined ratings. For female interviewers the correlation between interview ratings and job collapsed across all conditions was .31 for clinically combined ratings (N = 122, p<.001) and .39 (N = 76, p<.001) for mechanically combined ratings.

Program

Of the 150 participants, 86.7% were Commerce students, 7.3% were registered in Labour Studies, 4.7% were Engineering students, and the remainder 1.4% were equally represented by Geography and Masters in Business Administration. These differences were not significant across conditions (Chi-Square = 35.59, df = 28, p=.15).

Program Year

Of the 150 participants, 52.0% were in the second year of their program, 24.7% were in their third year, and 22.0% were in their fourth year. The other 1.7% were equally represented by first and fifth year students. Since the videotapes were presented during students' class time it was expected that there would be differences in program year across conditions. There were significant differences across conditions ($F = 2.66$, $df = 7$, $p<.05$).

Prior Experience in Social Work

Participant experience with the Social Work profession was hypothesized as having an impact on interviewer ratings and therefore data were collected on participants previous experience as Social Workers. Only four subjects (2.7%) identified that they had been involved in social work and their experience did not differ significantly between conditions (Chi-Square $= 11.01$, $df = 7$, $p=.14$).

Prior Experience as an Interviewee

There was considerable variability in the number of occasions participants had been interviewed for a job. These data ranged from 0 (5.3%) to 40 (0.7%). The majority of participants (77.3%) had been involved in 5 or fewer interviews. The impact of

interviewee experience was tested using a One Way ANOVA with the finding that the results were not significant across conditions (F = 1.20, df = 7, p=.31).

Prior Experience as an Interviewer

Most of the participants (78.0%) had no experience as an interviewer, although 16% reported that they had been an interviewer on 3 or fewer occasions. The complement of the sample (6%) stated that they had been an interviewer on 4 or more occasions. The impact of interviewer experience was tested using a One Way ANOVA with the finding that the results were not significant across conditions (F = 1.31, df = 7, p=.25).

Impact of Background Variables on Applicant Selection

A logistic regression was calculated to assess the impact of the background characteristics on the decision to hire one applicant over the other (i.e., Applicant #1 versus Applicant #3). The variables entered into the regression were: year of program, social work experience, interviewer experience, interviewee experience, and gender. None of these variables were significantly related to the choice of applicant.

Choice

A logistic regression was calculated to assess the factors that led to the decision to chose one applicant over the other (i.e., Applicant #1 versus Applicant # 3). Choice is an important concept in selection research because ultimately every hiring decision is a choice between multiple applicants. In this study the job performance data made it possible to assess the "right" choice. The background variables were entered in the equation first, followed by the use of job relevant questions, rating scales and the accuracy of pre-interview information.

The use of job relevant information and the accuracy of pre-interview information significantly predicted the choice of applicant. Applicant #3 (the strongest performer on the job) was selected more often than Applicant #1 when the interview contained job relevant information (B = 1.04, Wald = 8.19, Sig. = 0.004) and accurate pre-interview information preceded the interview (B = 0.80, Wald = 4.84, Sig. = 0.028).

Results of Hypotheses

Hypothesis 1a, which postulated that the pre-interview rating would be positively correlated with the post interview rating, was supported. The correlation between pre-

interview ratings and post-interview ratings was .38 (N=300, p<.001).[5] This finding is consistent with the theory proposed by Dipboye (1982).

Hypothesis 1b, which predicted that the correlation between the pre-interview impression and the final interview decision would be lower for the job relevant interview with a mechanical combination of ratings than the general interview with one overall rating, was supported. The pre-interview impression was significantly correlated with the post-interview rating for the general interview followed by an overall rating (.62, N=30, p<.001) but not for the job relevant interview with mechanical combination of ratings (-.23, N=34, p=.10). In addition, these were significantly different from each other (z = -3.25, p<.001). Therefore, when general questions are used and the interviewer makes an overall evaluation after the interview, the pre-interview impression seems to play a considerable role in the post-interview rating. In contrast, the impact of the pre-interview impression is eliminated when the interview has job relevant questions with rating scales that are mechanically combined to produce an overall score for the interviewee.

---

[5] With each "interviewer" providing ratings on two applicants, the total number of data points could be as high as 300 (150 participants multiplied by 2). The use of multiple ratings is similar to the use of multiple supervisors providing job performance ratings on several employees. For example, Latham and Saari (1984) reported that 4 supervisors provided performance assessments on 29 job incumbents but did not report any corrections for potential rater effects. Although the potential exists for a rater effect to be introduced, the small number of assessments provided by each "interviewer" over the large number of participants (150) minimizes any potential impact (personal communication with Maryann Fraboni, Ph.D.). In addition the two ratings were randomly separated and correlations between pre-interview and post-interview ratings did not change significantly (.37, N=150, p=.45).

Hypothesis 2, which predicted that interviews preceded by positive pre-interview information, would have higher (i.e., more positive) average post-interview ratings than interviews preceded by negative pre-interview information, was supported. The pre-interview information for Applicant #1 was positive across all conditions whereas the pre-interview information varied across four conditions for Applicant #3. There were significant differences between average post interview ratings (from 1=poor to 5=excellent) for Applicant #3 when positive versus negative information preceded the interview (4.19 versus 3.47 on a 5 point scale, t=-6.20, p<.001). In contrast, there were no differences across the same conditions for Applicant #1 when only positive information preceded the interview (3.95 versus 3.83, t=1.17, p=.244). Therefore, the negative information had its intended lowering effect on post interview scores.

Reliability

Inter-rater reliability was calculated based on generalizability theory (Crocker and Algina 1986). Generalizability analysis uses data obtained from a two factor (or repeated measures) ANOVA with examinees and raters as independent variables. Essentially it utilizes the mean squares value for examinees and error. The generalizability coefficient is influenced by the number of raters with the coefficient generally increasing as the

number of raters increases. Therefore, the results presented in Table 4 are based on one,

two and multiple raters to demonstrate the potential differences in generalizability.

For illustrative purposes, the results have been grouped based on the accuracy of

pre-interview information. Overall, it is evident that the generalizability coefficients

increase as one moves from one rater to multiple raters which is expected based on the

theory (Crocker and Algina 1986). Moreover, when the pre-interview information was

accurate, the generalizability coefficient was quite high with multiple raters, and there

were only minor differences in generalizability between conditions. However, when the

TABLE 4.-- Generalizability coefficients across conditions

Condition	1 Rater	2 Raters	Multiple Raters (N in brackets)
Accurate/Job Relevant/Mechanical Rating	0.24	0.39	0.84 (19)
Accurate/Job Relevant/Clinical Rating	0.37	0.54	0.92 (19)
Accurate/Job Relevant/Observe Only	0.37	0.53	0.93 (22)
Accurate/General/Mechanical Rating	0.26	0.34	0.82 (19)
Accurate/General/Clinical Rating	0.19	0.32	0.82 (19)
Accurate/General/Observe Only	0.28	0.44	0.87 (17)
Inaccurate/Job Relevant/Mechanical Rating	0.43	0.60	0.93 (17)
Inaccurate/Job Relevant/Clinical Rating	0.05	0.10	0.48 (17)
Inaccurate/Job Relevant/Observe Only	0.09	0.17	0.71 (25)
Inaccurate/General/Mechanical Rating	0.11	0.19	0.66 (16)
Inaccurate/General/Clinical Rating	0.04	0.08	0.40 (16)
Inaccurate/General/Observe Only	0.00	0.01	0.06 (15)

pre-interview information was inaccurate, there were considerable differences between conditions.

Hypothesis 3a, which stated that interviews with a common set of job relevant questions would be more reliable when compared to those with a common set of general questions, was supported. The weighted average generalizability coefficient for job relevant questions was .80 (119) as compared to .63 (102) for general questions. The difference was significant ($z = 2.62$, $p < .01$) based on conversion to Fisher Z (Guilford 1956).[6]

Hypothesis 4a, which indicated that interview decisions with a common set of behavioural rating scales for scoring each question would be more reliable when compared to those with no rating scales, was not supported. The generalizability coefficient for interviews with rating scales (i.e., providing a rating after each question) was .74 (142). The coefficient for interviews which did not use any scales for interview questions (i.e., observe only and provide an overall rating after the interview) was .68 (79) ($z = 1.13$, $p =$

---

[6] There is no commonly accepted statistical test of significance for comparing generalizability coefficients (based on personal communication with Dennis Doverspike and Richard Shavelson). Therefore, the use of conversion to Fisher Z and comparing Z values may not be appropriate if the distribution of generalizability coefficients is not normal. Despite the lack of evidence to support the normality of generalizability coefficients, Doverspike, Carlisi, Barrett, and Alexander (1983) reported confidence intervals for generalizability coefficients with the caution that the normality of the distribution may not have been met. The approach chosen in this thesis is similar in that Z values were calculated in the absence of information about normality but with the same caution.

0.13). The generalizability coefficient for the "observe only" interviews, which is similar to the unstructured interview used on other studies, was consistent with that research, particularly Wiesner and Cronshaw (1988).

Hypothesis 5a , which predicted that interview decisions based on common set of ratings combined mechanically would be more reliable when compared to those with a common set of ratings combined clinically, was supported. The generalizability coefficient was .82 (71) for interviews with ratings combined mechanically and .67 (71) for ratings combined clinically ($z = 2.17$, $p < .05$).

Validity

The criterion-related validity of post-interview decisions was calculated using point biserial correlations for post-interview ratings and job performance across each condition. Job performance was measured using an average of supervisor ratings on 5 point scales across 26 behaviours representing interpersonal relations, organizational skills and behaviours and behavioural predispositions. These were determined to be important for the Social Worker position based on job analysis information (Wiesner 1988).

The overall validity coefficient for "observe only", clinical and mechanical ratings was .29 (N=158, $p<.001$), .28 (N=142, $p<.001$), and .35 (N=142, $p<.001$), respectively. The differences were calculated using the conversion to Fisher Z (Guilford 1956). The

differences were not significant for "observe only" and clinical ratings (z = 0.09, p = 0.46) but were significant for "observe only" and mechanical ratings (z = -2.02, p < .05) and clinical ratings and mechanical ratings (z = -2.06, p < .05). The correlations by condition are presented in Table 5.

TABLE 5.-- Validity coefficients across conditions

Condition	Pre-Interview/ Post-Interview	Pre-Interview/ Performance	Post-Interview/ Performance	N
Accurate				
Job Relevant/Mechanical Rating	0.39*	0.45**	0.53**	38
Job Relevant/Clinical Rating	0.42**	0.45**	0.49**	38
Job Relevant/Observe Only	0.46**	0.21	0.45**	44
General/Mechanical Rating	0.38*	0.25	0.28*	38
General/Clinical Rating	0.38*	0.25	0.32*	38
General/Observe Only	0.48**	0.18	0.41*	34
Inaccurate				
Job Relevant/Mechanical Rating	-0.23	-0.50**	0.49**	34
Job Relevant/Clinical Rating	0.19	-0.50**	0.24	34
Job Relevant/Observe Only	0.39**	-0.23	0.26	50
General/Mechanical Rating	0.15	-0.48**	0.25	32
General/Clinical Rating	0.15	-0.48**	0.11	32
General/Observe Only	0.62***	-0.30	-0.02	30

Note: * = p<.05, ** = p<.01, *** = p<.001

Hypothesis 3b, which postulated that interview decisions based on a common set of job relevant questions will be more valid when compared to those with a common set of general questions, was supported. The correlation between post-interview ratings and performance for the four job relevant interviews was .44 (N=188, p<.001). The correlation between post-interview ratings and performance for the four general interviews was .24 (N=204, p<.01). The difference was significant (z = 2.23, p < .05).

Hypothesis 4b, which proposed that interview decisions based on a common set of behavioural rating scales for scoring each question would be more valid when compared to those with no rating scales, was not supported. The correlation between post-interview ratings and performance for the interviews with rating scales was .32 (N=284, p<.001). The correlation between post-interview ratings and performance for the interviews without rating scales for each question (i.e., observe only) was .29 (N=158, p<.001; z = 0.33, p = .38).

Hypothesis 5b, which stated that interview decisions based on a common set of ratings combined mechanically would be more valid when compared to those with a common set of ratings combined clinically, was supported. The correlation between post-interview ratings for interviews with mechanically combined ratings was .35 (N=142, p<.001). The correlation between post-interview ratings and performance for clinically

combined ratings was .28 (N=142, p<.001). The difference was significant (z = 2.059, p < .05).

<div align="center">Accuracy</div>

Accuracy was measured using the calculation developed by Cronbach (1955) which is based on the squared difference between the true score and actual score averaged across participants. For the present study, the job performance rating represents the true score and the interview rating the actual score. The accuracy differential reflects the squared difference between the interview rating and job performance rating (both were based on a 5 point scale). Therefore, an accuracy differential of 0 represents the most accurate interview rating.

Hypothesis 3c, which predicted that interview decisions based on a common set of job relevant questions would be more accurate when compared to those with a common set of general questions, was supported. Accuracy differentials were analyzed using One Way Repeated Measures ANOVAs to assess the hypotheses. Based on the Repeated Measures ANOVA (Between Subjects Effect F = 10.26, df = 298,1, p<.01), the average difference between post-interview ratings and performance was significantly less for interview decisions based on job relevant questions (1.02, N=166) than interview decisions based on general questions (1.64, N=134).

TABLE 6.-- Accuracy differentials across conditions

Condition	Applicant #1	Applicant #3	N
Positive/Accurate/Job Relevant/Mechanical Rating	0.39	0.65	18
Positive/Accurate/Job Relevant/Clinical Rating	0.61	0.62	19
Positive/Accurate/Job Relevant/Observe Only	1.01	0.54	22
Positive/Accurate/General/Mechanical Rating	1.33	1.45	19
Positive/Accurate/General/Clinical Rating	1.34	1.41	19
Positive/Accurate/General/Observe Only	2.26	1.15	17
Negative/Inaccurate/Job Relevant/Mechanical Rating	0.56	0.86	17
Negative/Inaccurate/Job Relevant/Clinical Rating	0.78	1.34	17
Negative/Inaccurate/Job Relevant/Observe Only	1.23	1.88	25
Negative/Inaccurate/General/Mechanical Rating	1.17	2.02	16
Negative/Inaccurate/General/Clinical Rating	1.01	2.25	16
Negative/Inaccurate/General/Observe Only	0.90	2.33	15

Note: All the pre-interview information for Applicant #1 was accurate.
Only Applicant #3 had both accurate and inaccurate information.

Hypothesis 4c, which predicted that interview decisions based on a common set of

behavioural rating scales would be more accurate when compared to those with no rating

scales, was not supported. The difference in average ratings between mechanical or

clinical ratings and performance (1.61, N=150) and observation only ratings and

performance (2.03, N=150) was not significant (Between Subjects Effect F = 1.24, df = 298,1, p=.25).

Hypothesis 5c, which predicted that interview decisions based on a common set of ratings combined mechanically would be more accurate when compared to those with a common set of ratings combined clinically, was not supported. The difference between post-interview ratings and performance was not significantly different (Between Subjects Effects F = 1.02, df = 140,1, p=.31) for mechanically combined ratings (1.48, N=72) when compared to clinically combined ratings (1.73, N=72).

Differential Validity

Since bias has been operationalized as differences in validity coefficients across conditions, differential validity was calculated for each condition comparing the validities for accurate versus inaccurate pre-interview information.

Hypothesis 3d, which postulated that interview decisions based on a common set of job relevant questions would be less biased by negative (inaccurate) pre-interview information than interview decisions based on a common set of general questions, was not supported. The gap between post-interview ratings and job performance (based on the accuracy differential) for interviews with job relevant questions was not significantly different for positive/accurate versus negative/inaccurate pre-interview information

(positive/accurate = .49, N=120, p<.01; negative/inaccurate = .32, N=118, p<.05; z = 1.56, p=.06). The gap between post-interview ratings and job performance for interviews with general questions was also not significantly different for accurate versus inaccurate pre-interview information (positive/accurate = .33, N=110, p<.05; negative/inaccurate = .12, N=94, p=.11; z = 1.56, p=.06). Therefore, there was no evidence to support differential validity with either interview type (job relevant or general questions).

Although it should be noted that both tests produced results that were close to significant (i.e.,, p=.06), the support for the reduction of bias by question type (job relevant versus general) depended on a non-significant difference between accurate/positive and inaccurate/negative pre-interview information for job relevant questions and a significant difference for general questions. Therefore, even if both tests were significant at p>.05, this would not support the hypothesis that question type moderated the impact of pre-interview information on post interview decisions.

Hypothesis 4d, which proposed that interview decisions based on a common set of behavioural rating scales for scoring each question would be less biased when compared to those with no rating scales, was supported. The gap between post-interview ratings and job performance for interviews with ratings scales was not significantly different for positive/accurate versus negative/inaccurate pre-interview information (positive/accurate = .40, N=152, p<.01; negative/inaccurate = .27, N=132, p<.05; z = 1.23, p=.11). In

contrast, the gap between post-interview ratings and job performance was significant for interviews with observation only during the interview and an overall rating after (positive/accurate = .47, N=78, p<.05; negative/inaccurate = .16, N=80, p=.11; z = 2.18, p<.05). It appears that the presence of rating scales eliminated differential validity.

Hypothesis 5d, which predicted that interview decisions based on a common set of ratings combined mechanically would be less biased when compared to those with a common set of ratings combined clinically, was supported. The gap between post-interview ratings and job performance for interviews with a mechanical combination of ratings were not significantly different for positive/accurate versus negative/inaccurate pre-interview information (positive/accurate = .41, N=76, p<.01; negative/inaccurate = .34, N=66, p<.05; z = 0.48, p=.32). In contrast, the gap between post-interview ratings and job performance was significant for interviews with a clinical combination of ratings (positive/accurate = .41, N=76, p<.01; negative/inaccurate = .17, N=66, p=.11; z = 1.65, p<.05). It appears that the mechanical combination of ratings eliminated the differential validity.

TABLE 7.-- Differential validity across conditions

Condition	Accurate/ Positive	Inaccurate/ Negative	Differential Validity
Job relevant questions	.49**	.32*	z = 1.56
General questions	.33*	.12	z = 1.56
Rating scales for each question	.40**	.27*	z = 1.23
Observe only (overall rating after interview)	.47**	.16	z = 2.18*
Mechanical combination of ratings	.41**	.34*	z = 0.48
Clinical combination of ratings	.41**	.17	z = 1.65*

Note: * = p<.05, ** = p<.01, *** = p<.001

In summary, there was support for three of the four indices (all except differential validity) with respect to the differences between interviews with job relevant questions versus those with general questions. The use of rating scales for each question versus observation only followed by an overall rating after the interview reduced differential validity but had no impact on the other indices. The combination of ratings produced a similar pattern with mechanical combination resulting in differences for reliability, validity and bias but not accuracy. Overall, the strongest interview type was the interview with job relevant questions with rating scales for each question and a mechanical combination

TABLE 8.-- Summary of results

Issue	Hypothesis	Issue	Supported (Y/N)
Reliability	3a	Job Relevant	Yes
	4a	Rating Scale	No
	5a	Combination	Yes
Validity	3b	Job Relevant	Yes
	4b	Rating Scale	No
	5b	Combination	Yes
Accuracy	3c	Job Relevant	Yes
	4c	Rating Scale	No
	5c	Combination	No
Differential Validity	3d	Job Relevant	No
	4d	Rating Scale	Yes
	5d	Combination	Yes

of ratings to produce a final score for the interview. This interview produced a strong

reliability (.84), the highest validity (.53, N=38, P<.01), the lowest accuracy differential

(.39), and no differential validity.

CHAPTER 6:

DISCUSSION

The purpose of this study was to study the impact of positive/accurate and negative/inaccurate pre-interview impressions on interview decisions across different interview formats. This study integrates two streams of interview research that has historically been conducted independently. These two streams are typically referred to as process research and outcome research. This thesis makes four primary contributions to interview research:

- facilitates our understanding of the interview in the broader selection context which includes resume and reference check information,

- facilitates our understanding of how biases established prior to the interview can influence the reliability, validity and accuracy of interview decisions and how different interview formats can reduce this impact,

- develops our understanding of the cognitive processes involved in the interviewer's selection decision, and

- integrates two previously separate streams of interview research.

Results in the Context of Outcome Research

Reliability

The generalizability coefficient measures the amount of variance accounted for in the ratings across examinees, raters and error and thus it is the most appropriate index of reliability in this context. Although the use of the generalizability coefficient in research is fairly recent, its theoretical basis is strongly rooted in analysis of variance.

The generalizability coefficient, similar to other indices of reliability, is influenced by the number of raters who provided the rating. In the present study the number of raters was considerably high which explains the high reliabilities across both structured and unstructured interviews. However, when one calculates the generalizability coefficient based on 1 or 2 raters, which is typical of the number of raters in actual interviews, the reliability drops considerably for both types of interviews. This finding may be attributed to the following three issues: lack of knowledge the students have about the job of Social Caseworker, lack of time to sufficiently review the study materials, and the large number of participants in each session.

Four meta-analyses have reviewed interview reliability across either one or both types of interviews (structured and unstructured). Table 9 presents the results of this research in comparison to the current study. Two of these studies measured the inter-rater reliability of the unstructured interview.

Table 9.-- Comparisons to prior meta-analyses on interview reliability

Interview Type	Study	1	2	3	4	Current Study (2 raters)	Current Study (>15 raters)
Unstructured			.61		.69	.24	.49
Job Related/Unstructured		.75			.75	.34	.77
Job Related/Structured		.84	.82	.70	.92	.41	.80

Note: Study 1 - McDaniel, M.A., Whetzel, D. L., Schmidt, F. L., Hunter, J. E., Maurer, S., and Russell, J. 1986, Study 2 - Wiesner, W. H., and Cronshaw, S. F. 1988, Study 3 - Wright, P. M., Lichtenfels, P. A., and Pursell, E. D. 1989, Study 4 - Conway, J. M., Jako, R. A., and Goodman, D. F. 1995, Current Study is based on interview data for accurate/positive pre-interview information only.

In the two meta-analyses that reported reliability coefficients for unstructured interviews (Wiesner and Cronshaw 1988; and Conway, Jako and Goodman 1995), the average reliability was .69 and .61 respectively. In the current study unstructured interview reliability was .24 with two raters and .49 with over fifteen raters. The large differences in reliability for the unstructured interview between the current study with two raters and the meta-analyses suggest that some aspect of the current study reduced the reliability of interview ratings for the unstructured interview.

A similar pattern was found for the structured interview reliability. The four

meta-analyses (McDaniel, M.A., Whetzel, D. L., Schmidt, F. L., Hunter, J. E., Maurer, S.,

and Russell, J. 1986; Wiesner, W. H., and Cronshaw, S. F. 1988; Wright, P. M.,

Lichtenfels, P. A., and Pursell, E. D. 1989; and Conway, J. M., Jako, R. A., and

Goodman, D. F. 1995) found average reliability coefficients ranging from .70 to .92,

whereas the reliability for the current study (with two raters) was .41. The reliability

improved to .80 with fifteen raters or more. Further research needs to explore why these

differences were found as they may have implications for the external validity of this type

of research using students who are inexperienced interviewers. Fortunately, despite the

absolute differences in reliability, the pattern of results in the current study are similar to

those in the meta-analyses (i.e., job related/structured interviews are more reliable than

unstructured interviews).

There were differences in interview reliability based on the accuracy of the pre-

interview information. The reliability of the interview was higher when accurate pre-

interview information preceded the interview. This suggests that the interviewers were

influenced by the pre-interview information. This information increased the interview's

reliability when it was consistent with the interview but reduced the reliability when it was

inconsistent. This finding suggests that pre-interview information can impact on post-

interview decisions. Furthermore, different interviewers dealt with the incongruous

information in different ways which reduced the reliability to a greater extent in the inaccurate condition.

Although there was an overall improvement in reliability with job relevant questions as opposed to general questions, rating scales and the combination of ratings did not have the hypothesized impact. For example, the "observe only" conditions were expected to be the lowest in reliability but produced high generalizability coefficients in several conditions (except when the pre-interview information was negative/inaccurate). This suggests rating scales and a mechanical combination of ratings improve the reliability of the interview primarily when they are used in conjunction with job relevant questions.

Overall, the pattern of results found here are consistent with the body of literature on the interview. The combined use of job relevant questions, rating scales and a mechanical combination of ratings increased interview reliability as measured by the generalizability coefficient. This finding is consistent with that of Wiesner (1988) and can be explained by the improved measurement properties of interviews with job relevant questions, rating scales and a mechanical combination of ratings.

Validity

In this study, validity refers to criterion-related validity or the correlation between predictor scores and criterion scores. The predictor consisted of ratings made based on

interview information presented in videotapes. The criterion was based on current supervisor assessments of applicant job performance (i.e., applicants who were interviewed were already working on the job).

This study can be considered a postdictive validation study in which the criterion was measured first and the predictor measured later. Although, each aspect of validity (i.e., construct, content and criterion-related) are equally important, determining the criterion-related validity is necessary when a predictor has adverse impact. In other words, when a test has been shown to have adverse impact, the employer must demonstrate that there is relationship between predictor scores and criterion scores (i.e., the predictor has criterion-related validity). Furthermore, the correlation coefficient is the primary unit of measurement for differential validity which is an indicator of bias.

In general, these results are consistent with those of meta-analyses of the interview (Wiesner and Cronshaw 1988; McDaniel, Whetzel, Schmidt, Hunter, Maurer, and Russell 1986; Wright, Lichtenfels and Pursell 1989; Marchese and Muchinsky 1993; Huffcutt and Arthur 1994). The results of the meta-analyses ranged from .11 to .17 for the unstructured interview and .29 to .45 for the structured interview. For the current study, similar results were found for both the unstructured interview (.21) and the structured interview (.44). These correlations were based on all the possible interviewer ratings

(over 15 interviewers) which mitigated the differences in reliability discussed in the

previous section.

Table 10.-- Comparisons to prior meta-analyses on interview validity

Interview Type	Study	1	2	3	4	5	Current Study
Unstructured		.16	.17			.11	.21
Job Related/Unstructured		.18				.20	.35
Job Related/Structured		.30	.34	.29	.45	.34	.44

Note: Study 1 - McDaniel, M.A., Whetzel, D. L., Schmidt, F. L., Hunter, J. E., Maurer, S., and Russell, J. 1986, Study 2 - Wiesner, W. H., and Cronshaw, S. F. 1988, Study 3 - Wright, P. M., Lichtenfels, P. A., and Pursell, E. D. 1989, Study 4 - Marchese, M. C., and Muchinsky, P. M. 1993; Study 5 - Huffcutt, A. I., and Arthur, W. A. 1994.

Accuracy

The validity coefficient measures the strength of a relation between two sets of

scores, accuracy measures the strength and the kind of relation between scores (Gordon

1970). Therefore, with the availability of job performance scores and only two applicants,

it is more reasonable to use accuracy than validity as an indicator of how well the

interview predicts performance. Many different measures of accuracy exist and have been used in performance appraisal research for some time (Sulsky and Balzer 1988).

Historically, accuracy scores have not been used in interview research, therefore it is not possible to compare these results to those of prior studies. There was considerable consistency in the pattern of results between the validity and accuracy criteria. For example, the interviews with the highest validity coefficients also had the highest accuracy (i.e., lowest accuracy differential between post-interview and job performance ratings).

Differential Validity

The differential validity analysis demonstrated the impact of using scales (either behaviourally anchored or graphic) on reducing the differential validity associated with a negative/inaccurate pre-interview impression. Interestingly, question type did not produce the expected pattern of differential validity on its own; it was nonsignificant for both job relevant questions and general questions. The use of rating scales and a mechanical combination of ratings reduced impact of the negative/inaccurate pre-interview impression (i.e., eliminated any differential validity). The most resistant interview type to the negative pre-interview impression was the interview that combined job relevant questions, behaviourally anchored rating scales and a mechanical combination of ratings.

Additional Findings

One additional finding deserves mention which was the significant correlation between sound quality and job performance. This finding was surprising given that different individuals provided the interview ratings (i.e., students in the study) and job performance ratings (i.e., interviewee's supervisors) and as a result there should not have been a significant relationship between them. It may be that the sound quality measure reflected the speech patterns of the applicants (i.e., volume, tone, eye contact with the camera) that relate to the interpersonal component of the job performance evaluation.

In summary, the four indices clearly show that job relevant questions improve the reliability, validity, and accuracy of interview decisions. Also, in the presence of inaccurate pre-interview information the interview with job relevant questions, rating scales and a mechanical combination of ratings reduces the differential validity associated with this information.

Outcomes in the Context of Process Research

Overall, the impact of the positive/accurate pre-interview impression is consistent with the body of literature on the interview. Interestingly, the generalizability coefficient was significantly lower when the interview was preceded by negative/inaccurate information.

The impact of negative/inaccurate information on the reliability of interview decisions across interview formats has not been addressed prior to this study. This finding suggests that any negative information the interviewer receives prior to the interview can bias the interview decision. In most real life hiring situations, interviewers receive resume information and information from other sources (i.e., headhunters, past managers) which can create a positive or negative bias prior to the interview.

Pre- and post-interview ratings were highly correlated when the pre-interview information was accurate (ranging from .38 to .48). The positive relationship between pre- and post-interview ratings found in this study is consistent with the research conducted by Dipboye and others (Dipboye, Fontenelle and Garner 1984; Dipboye, Stramler and Fontenelle 1984; Hoff-Macan and Dipboye 1990). This study also provides support for the impact of pre-interview information on the interviewer's processing of information (Dipboye 1982; Hamilton, Katz and Leirer 1980; Carlston 1980) and it is the first study to consider the impact of information processing independent of the interviewer's behaviour.

In the negative/inaccurate pre-interview information condition, the pre- and post-interview ratings were only significantly correlated when the interviewer provided a global rating after the interview was concluded (the "observe only" groups). The use of rating scales reduced the correlation between pre- and post-interview ratings. This finding

suggests that any negative/inaccurate pre-interview biases have the greatest impact when the interviewer provides a global rating at the end of the interview. In addition, the impact can occur across interviews with job relevant questions or general questions. Therefore, future research on pre- and post-interview ratings should consider the type of interview and rating scales used.

Implications for Bias

One of the purposes of this thesis was to explore the impact of bias on the reliability, validity, accuracy and differential validity of interview decisions. Bias was operationalized as a negative/inaccurate pre-interview impression that was established by placing a negative comment in a reference check prior to the interview. Therefore, this inaccurate information produced a negative pre-interview impression which had an impact on subsequent interview ratings. This manipulation was designed to mimic the impact that interviewer biases could have on their ratings of interviewees. For example, if an interviewer felt that only men should become police officers, then he may evaluate female applicants more severely. Also, he may not be aware of the impact of this bias on his interview ratings. Future interview research should incorporate interviewer biases to understand their impact on interview validity.

This study suggests that interviews which use job relevant questions, rating scales and a mechanical combination of ratings will have less bias in the final interview decision. Therefore, using this type of interview will provide more opportunity for applicants across all designated groups to demonstrate their abilities in the interview. In other words, the structured interview can reduce any adverse impact that may exist in unstructured interviews.

These results are consistent with a recent study by Kataoka, Latham and Whyte (1995) which found that anchoring (identifying an applicant as "good" or "bad" prior to the interview) had an impact on subsequent interview ratings and that this impact was reduced when Pattern Behaviour Description Interviews and Situational Interviews were used.

New Findings

Since this is the first study to assess the impact of inaccurate pre-interview information across unstructured and structured interview formats there are some new findings here that have implications for both process and outcome oriented interview research. The first new finding is that pre-interview biases can have an impact on interviewer decisions via the information processing mechanism alone. The use of videotapes in this study allowed the analysis of the impact of bias on interviewer decisions,

independent of the interviewer's behaviour. This finding suggests that even in standardized interviews, which require interviewers to ask the same questions, any preconceived biases may have an impact.

The second new finding is that pre-interview biases can impact on interviews with both job relevant and general type questions. Only the interview with job relevant questions, ratings scales scored during the interview, and a mechanical combination of ratings was resistant to the effects of inaccurate pre-interview information based on reliability, validity and accuracy information.

The third new finding is the improved reliability, validity and accuracy of the unstructured interview when the interviewer receives accurate pre-interview information. This result may provide some insight into why interviewers continue to use the unstructured interview when researchers question their validity. Typically, interviewers in a real setting receive multiple pieces of information about an applicant including their resume, application form, possibly performance information, as well as information from the interview itself. In contrast, typical experimental interview research provides interviewers with only information from the interview.

Therefore, it may not be possible to compare the findings of experimental interview research with actual interviews because interviewers have less information in the former situation. Since reliability sets the upper limit of validity, the improvement in the

reliability of interview decisions resulting from accurate pre-interview information, can

partially explain the improvement in the validity of the unstructured interview in this study.

Limitations of this Study

The primary limitation of this study is external validity. As with any experimental

study the external validity is always traded off for internal validity. In the present study,

the use of undergraduate students with little interviewing experience or knowledge of the

Social Caseworker job limits the generalizability of the results. In addition, the use of

videotaped interviews of current job incumbents rather than applicants limits the study's

generalizability since actual interviews involve an interaction between the interviewer and

interviewee.

In defence of research using students, Ilgen (1986) points out that there are good

reasons to conduct a study with students. One reason is to conduct laboratory research to

study issues that cannot be studied in the field. In order to fully understand the complex

issues surrounding the interview decision, it was felt that this study should be conducted in

the laboratory prior to being tested in actual hiring situations.

In other words, the rationale for this type of study is based on the complexity of

the interview process itself. In actual interviews it is difficult to tease apart all of the

effects that impact on the interviewers' decision. Furthermore, actual interviewers bring

many different perspectives into the interview, based on their knowledge of different jobs and stereotypes that have been supported in past interviews they have conducted. The present design was chosen to reduce the potential impact of these variables.

In the scope of research, this loss of external validity is more than offset by the gain in internal validity. Also, with the findings presented here, there is reason to look to future research in the field to increase the external validity of the study's findings. Research in the field conducted by Kataoka, Latham, and Whyte (1995) and Latham and Skarlicki (1995) suggest that structuring the interview can reduce the impact of interviewer biases. In addition, Marlowe, Schneider and Nelson (1996) found that even experienced managers had negative biases that impacted on less attractive female job applicants.

CHAPTER 7:

CONCLUSIONS

This study found that negative/inaccurate pre-interview information can have a negative impact on the reliability, validity, accuracy and differential validity of interview ratings. This impact was reduced by the use of job relevant questions, rating scales and a mechanical combination of ratings. Therefore, in order to reduce potential biases in the interview process and subsequent adverse impact, interviews should utilize job relevant questions, rating scales and a mechanical combination of ratings. Future research should attempt to replicate these results with experienced interviewers in more realistic interview situations.

APPENDIX A: PARTICIPANT INSTRUCTIONS

INSTRUCTIONS

My name is Peter Hausdorf and I am a Ph.D. student here at McMaster. Dr. Wiesner and I are conducting a study of decision making in the employment interview. Today you will be placed in the role of an employment interviewer for a social services organization. You will be presented with resumes and reference checks and will observe actual job interviews on videotape for 3 applicants.

The process is as follows:

The resume and reference check will be presented for the first applicant and then you will have a chance to observe the interview. After you have considered the first applicant the resume and reference check will be presented for the second applicant followed by the interview. These steps will be repeated for the third applicant. After you have seen all the applicants you will be asked to fill out some final information.

You should all have a study package which has a consent form on the front. Please read the consent form and sign it if you consent to be involved in the study which will be conducted today. Four different types of packages have been distributed so make sure that you focus on your own package and don't look to others for guidance because they will be doing something different. If you have any questions, ask me. You should make sure that your package is complete. Check the last page for a THANK-YOU and make sure that you have every page in between. Also, in the top right hand corner of the page you should have one of the following codes _____.

The interviews have some distortion in the sound so it may be difficult for you to hear them. I have tried to minimize this problem as much as possible. The best way to deal with it is to focus and concentrate on the applicant's voice as much as possible. Also, you don't have to hear everything that is said as long as you understand the basis of their response.

Turn to page 1 of your booklets now and read the instructions.

APPENDIX B: GENERAL INSTRUCTIONS AND JOB DESCRIPTION

EMPLOYMENT INTERVIEW STUDY

Introduction

Please assume that you are an administrator in a social welfare agency which is responsible for the administration of a welfare benefits program (financial assistance for needy families and individuals). Your agency needs to hire additional welfare case workers and you have been given the task of interviewing applicants for the vacant positions. Even though you will be viewing a videotape of portions of an employment interview, try to put yourself in the position of the interviewer in the tape - as if you were asking the questions. Note that this is an actual employment interview which we were able to videotape. We have selected the most important interview questions for inclusion in this briefer edited version.

Below is a job description for the position of social worker - public welfare. Please read it carefully, making sure you understand what the job entails and the qualities a candidate should possess in order to be able to do this job. Then turn to the instructions on the next page.

Job Description: SOCIAL WORKER - PUBLIC WELFARE

Administers welfare and compensation laws and regulations, performing any combination of the following duties:

Conducts home visits and interviews applicants for assistance in order to determine their eligibility for welfare benefits. Investigates health and medical records, and obtains details of saving deposits, cash-on-hand, and surrender values of insurance policies and earnings. Maintains liaison with welfare agencies and refers applicants for assistance to appropriate sources. Discusses status of particular cases with relevant officials to determine course of action. Calculates welfare budgets, issues relief vouchers and keeps accurate client files. Examines municipal welfare and assistance files and accounts to determine eligibility of recipients and to verify amounts paid. Works independently with minimum supervision and interactions with co-workers. Handles numerous, simultaneous demands and ensures that deadlines are met. Performs routine visits and reports, as assigned, on a daily basis as well as occasional non-routine duties. Ensures that knowledge of legislation and procedures affecting welfare benefits remains current by reading updates, memos and reports. Provides detailed and accurate explanation of regulations to welfare clientele and diffuses client frustrations. Encourages client independence by discussing options and

opportunities for client development. With the client and making appropriate referrals. Conducts self in a professional manner at all times.

For each candidate you will find actual resumes and brief notes made during a reference check from a previous employer. Previous research has shown that resume and reference information can help improve the accuracy of interview decisions.

PLEASE READ THE RESUME AND REFERENCE CHECKS FOR CANDIDATE #1 NOW

APPENDIX C: APPLICANT RESUMES

CANDIDATE #1   **Dianne Anderson**
824 Main Street
Guelph, Ontario
(519) 824-9745

EDUCATION:

Sep 1986 -   Carlton University
Apr 1990   Bachelor of Social Work

Sep 1982 -   Ottawa Valley High School
Jun 1986   S.S.H.G.D.

EXPERIENCE:

May 1990 -   **Children's Aid Society**
present   Guelph, Ontario
             Counseling Assistant:
             Assisted senior social worker in providing assessment, classification and
             referral of clients and their families involved in the program. Conducted
             presentations to local schools and community centres about CAS services.
             Maintained client files. Facilitated weekly self help program under
             supervision.

May 1989 -   **Hamilton-Wentworth Detention Centre, Adult Services**
Aug 1989   Hamilton, Ontario
             Summer Practicum:
             Provided classification and assessment of provincial offenders following
             remand to the facility; counseling; case planning for offenders; reports to
             parole boards, treatment facilities, and other social service professionals;
             participation on various case planning boards; co-facilitator for a group
             therapy program.

Jun 1987 -   **Huntington's Disease Society**
Aug 1987   Kingston, Ontario
             Summer Position
             Provided counseling and support to family members of clients. Participated
             in fund raising events. Conducted presentations at conferences and local
             schools to create awareness about the disease.

RELATED SKILLS:
Nov 1993   CPR Training, Heart and Stroke Foundation
Mar 1990   Christopher Leadership Course

REFERENCES:
                Available Upon Request

CANDIDATE #2　　**Terri Taylor**
219-5200 Lakeshore Road
Burlington, Ontario
(905) 639-8240

EDUCATION:

Sep 1984 -　　Laurentian University
Apr 1988　　　Bachelor of Social Work

Sep 1981 -　　Regina College of Applied Arts
Apr 1984　　　Administrative Assistant Diploma

Sep 1977 -　　Calgary Central High School
Jun 1981　　　S.S.H.G.D. Equivalent

EXPERIENCE:

Sep 1988 -　　**Alcoholics Anonymous**
present　　　　Toronto, Ontario
　　　　　　　Program Assistant:
　　　　　　　Provided group counseling. Ensured that clients received support through
　　　　　　　programs. Provided individual counseling to family members. Maintained
　　　　　　　follow up with medical practitioners. Provided progress reports to
　　　　　　　employers.

May 1986 -　　**Covenant House**
Aug 1986　　　Toronto, Ontario
　　　　　　　Summer Position
　　　　　　　Supervised children at the Family Support Clinic while their mothers were
　　　　　　　in group counseling sessions with outreach workers. Also, participated in
　　　　　　　fund raising activities.

May 1985 -　　**Whitby Psychiatric Hospital**
Aug 1985　　　Whitby, Ontario
　　　　　　　Summer Practicum:
　　　　　　　Provided individualized reports on patients identified for outplacement
　　　　　　　services. Implemented programs developed by other disciplines (i.e.,
　　　　　　　Psychiatry, Psychology) and provided reports. Conducted presentations at
　　　　　　　conferences.

RELATED SKILLS:

May 1993	Group Counseling Workshop, Alcoholics Anonymous
Dec 1993	Understanding Addiction Course, Clarke Institute
Feb 1991	CPR Course, St. John's Ambulance

REFERENCES:

Available Upon Request

CANDIDATE #3   **Wilma Jones**
                     256 Arthur Road
                     Waterloo, Ontario
                     (519) 823-3582

EDUCATION:

Sep 1985 -   McMaster University
Apr 1989   Bachelor of Social Work

Sep 1981 -   Smith River High School
Jun 1985   S.S.H.G.D.

EXPERIENCE:

May 1989 -   **Canada Employment Centre**
present   Oshawa, Ontario
               <u>Placement Officer:</u>
               Assisted employers in finding competent employees; Helped unemployed
               individuals find work; career counseling; presentations to employers,
               students, and other organizations; workshops for teaching effective work
               and job search skills; and general office duties.

May 1988 -   **Mississauga General Hospital**
Aug 1988   Mississauga, Ontario
               <u>Summer Practicum:</u>
               Provided support and counseling to cancer patients and family members.
               Participated in weekly staff meetings as the representative from social
               services. Attended conferences and conducted presentations.

May 1987 -   **South Shore Transition House**
Aug 1987   Terrace Bay, Ontario
               <u>Summer Intern</u>
               Assisted staff in continuation of programs for victims of spousal abuse over
               the summer. Provided individual counseling and preliminary assessments
               with clients. Participated in case management conferences.

RELATED SKILLS:
Feb 1993   CPR Training, Heart Foundation
Oct 1992   Racial Awareness Seminar, CEIC
Mar 1992   First Aid, Red Cross Society

REFERENCES:
             Available Upon Request

APPENDIX D: POSITIVE/ACCURATE PRE-INTERVIEW REFERENCE CHECKS

CANDIDATE #1   **Reference Check**

Applicant's Name: Dianne Anderson

Supervisor: Margaret Kline

Organization: Children's Aid Society

Dianne has been with us for 4 years. During this time she developed her assessment, classification and referral skills with respect to clients who were involved with the centre. Her performance was excellent and she demonstrated a keen interest in her work and vocation. She participated in a group therapy program which was very successful. Therefore, I can recommend her for the position of Social Services Worker with your organization.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	**2**	**3**	**4**	**5**

Poor   Excellent

TURN TO THE NEXT PAGE NOW

CANDIDATE #2   Reference Check

Applicant's Name: Terri Taylor

Supervisor: Dianne Kelly

Organization: Alcoholics Anonymous

Terri has been with our organization for the past 6 years as a Program Assistant.   During that time she assisted me in keeping the office organized and provided counseling to clients.   Overall her performance was good.   She worked well with the other staff in the office and developed good rapport with clients.   She demonstrated a lot of initiative with presentations.   She also initiated a workshop for families of alcoholics.   The only issue I would raise pertains to her affiliation with an extreme right wing political organization of which I don't approve.   Despite this difference she has been a good employee for us, therefore I can recommend her to you for the position of Social Services Worker.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	2	3	4	5

Poor   Excellent

TURN TO THE NEXT PAGE NOW

CANDIDATE #3   **Reference Check**

Applicant's Name: Wilma Jones

Supervisor: Nadine Wilson

Organization: Canada Employment Centre

Wilma has been employed with our organization for 5 years as a Placement Officer. I worked closely with her on several projects. She is a very hard worker and has been a real asset to our organization. Since her arrival she has improved our success rate through effective implementation of group and individual counseling programs. Her rapport and understanding with unemployed individuals has been remarkable. In addition to providing direct services, Wilma was responsible for liaising with employers and other agencies which she did professionally and courteously. All of her reports were accurate and timely. I can fully recommend her for the position of Social Services Worker and will admit that we will be sad to see her go.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1   **2**   **3**   **4**   **5**

Poor   Excellent

TURN TO THE NEXT PAGE NOW

APPENDIX E: NEGATIVE/INACCURATE APPLICANT REFERENCE CHECKS

CANDIDATE #1   Reference Check

Applicant's Name: Dianne Anderson

Supervisor: Margaret Kline

Organization: Children's Aid Society

Dianne has been with us for 4 years. During this time she developed her assessment, classification and referral skills with respect to clients who were involved with the centre. Her performance was excellent and she demonstrated a keen interest in her work and vocation. She participated in a group therapy program which was very successful. Therefore, I can recommend her for the position of Social Services Worker with your organization.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	**2**	**3**	**4**	**5**

Poor   Excellent

TURN TO THE NEXT PAGE NOW

CANDIDATE #2   Reference Check

Applicant's Name: <u>Terri Taylor</u>

Supervisor: <u>Dianne Kelly</u>

Organization: <u>Alcoholics Anonymous</u>

Terri has been with our organization for the past 6 years as a Program Assistant. During that time I worked closely with her on several projects. She is a very hard worker and has been a real asset to our organization. Since her arrival she has improved our success rate through effective implementation of group and individual counseling programs. Her rapport and understanding with family members has been remarkable. In addition to providing direct services, Terri was responsible for liaising with employers and medical practitioners which she did professionally and courteously. All of her reports were accurate and timely. I can fully recommend her for the position of Social Services Worker and will admit that we will be sad to see her go.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	**2**	**3**	**4**	**5**

Poor   Excellent

TURN TO THE NEXT PAGE NOW

CANDIDATE #3   Reference Check

Applicant's Name: Wilma Jones

Supervisor: Nadine Wilson

Organization: Canada Employment Centre

Wilma has been employed with our organization for 5 years as a Placement Officer. During that time she assisted me in keeping the office organized and helped to find jobs for unemployed individuals. Overall her performance was good. She worked well with the other staff in the office and developed good rapport with clients. She demonstrated a lot of initiative with presentations she made to employers. She also initiated a workshop for teaching effective job search skills. The only issue I would raise pertains to her affiliation with an extreme right wing political organization of which I don't approve. Despite this difference she has been a good employee for us, therefore I can recommend her to you for the position of Social Services Worker.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

```
1           2           3           4           5
|_____|_____|_____|_____|
Poor                                      Excellent
```

TURN TO THE NEXT PAGE NOW

APPENDIX F: "OBSERVE ONLY" INSTRUCTIONS, JOB RELEVANT QUESTIONS
AND RATING SCALE

INSTRUCTIONS

On the next page you will find a list of the interview questions asked during the interview which we videotaped. Please read through the interview questions in order to familiarize yourself with them prior to watching the videotape.

Previous research results suggest that interviewers are most accurate when they wait until the end of the interview to rate the candidate. They are able to pay more attention to the interviewee's actions and to what the interviewee says when they use this approach (i.e., they are less likely to get distracted from what is going on in the interview by paying too much attention to rating forms, etc.). Please follow the videotaped interview by reading the questions from the attached list when they are asked by the interviewer. Wait until the end of the interview to make your ratings.

You will be viewing the taped interview shortly.

PLEASE <u>REVIEW</u> THE QUESTIONS ON THE NEXT PAGE NOW

TURN TO THE NEXT PAGE NOW

QUESTION 1

Tell me about your last supervisor. How well did work together? What is your impression of his or her management abilities? Did you ever disagree with his or her decisions? If so, what did you do?

QUESTION 2

Give me an example of a difficult decision you had to make in your last job. Tell me what you decided and why. In general, when you had to make difficult decisions, did you seek advice from others or did you make the decision on your own?

QUESTION 3

Have you ever been in a situation where a number of people or tasks demanded your attention at the same time? If so, describe the situation and how you handled it.

QUESTION 4

Have you ever been in a situation where you knew that you could not meet the deadline? If so, how often did this occur and tell me about a typical situation and what you did.

QUESTION 5

In your last job, how did you go about ensuring that all your work would be done on time (Did you use any kind of system?)?

PLEASE PREPARE TO WATCH THE VIDEO INTERVIEW FOR CANDIDATE #1 NOW

REMEMBER <u>NOT</u> TO TAKE NOTES DURING THE INTERVIEW

DO NOT PROVIDE THIS RATING UNTIL AFTER THE INTERVIEW.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	2	3	4	5

Poor   Excellent

PLEASE READ THE RESUME AND REFERENCE CHECKS FOR CANDIDATE #2 NOW

TURN TO THE NEXT PAGE NOW

APPENDIX G: "OBSERVE ONLY" INSTRUCTIONS, GENERAL QUESTIONS AND GRAPHIC RATING SCALE

INSTRUCTIONS

On this page you will find a list of the interview questions asked during the interview which we videotaped. Please read through the interview questions in order to familiarize yourself with them prior to watching the videotape.

Previous research results suggest that interviewers are most accurate when they wait until the end of the interview to rate the candidate. They are able to pay more attention to the interviewee's actions and to what the interviewee says when they use this approach (i.e., they are less likely to get distracted from what is going on in the interview by paying too much attention to rating forms, etc.). Please follow the videotaped interview by reading the questions from the attached list when they are asked by the interviewer. Wait until the end of the interview to make your ratings.

You will be viewing the taped interview shortly.

PLEASE REVIEW THE QUESTIONS NOW

INTERVIEW QUESTIONS

1. What do you think the role of social services should be?

2. What types of people do you like most? What types do you like least?

3. How would other people describe you as an individual?

4. What do you consider to be your strengths? What are your weaknesses?

5. What is your greatest accomplishment?

REMEMBER <u>NOT</u> TO TAKE NOTES DURING THE INTERVIEW

PLEASE PREPARE TO WATCH THE VIDEO INTERVIEW FOR CANDIDATE #1 NOW

DO NOT PROVIDE THIS RATING UNTIL AFTER THE INTERVIEW.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	**2**	**3**	**4**	**5**

Poor   Excellent

PLEASE READ THE RESUME AND REFERENCE CHECKS FOR CANDIDATE #2 NOW

TURN TO NEXT PAGE NOW

APPENDIX H: JOB RELEVANT QUESTIONS AND BEHAVIOURALLY
ANCHORED RATING SCALES

INSTRUCTIONS

       On the next three pages you will find a list of the interview questions asked during the interview which we videotaped. Please read through the interview questions in order to familiarize yourself with them prior to watching the videotape. You will notice that each question is accompanied by a rating scale with five "anchors" or sets of descriptions. These descriptions represent the kinds of answers that the applicant might give to the question above the rating scale. Previous research has shown that interviewers using this kind of rating scale tend to give more accurate ratings of the applicant. Please use these descriptions as a guide in helping you determine the most appropriate rating for each answer. Parts of an applicant's answer may match two or more of the descriptions below each scale. If this happens, please rate the answer in accordance with the description which is closest or contains the greatest number of similar behaviours.

Previous research results suggest that interviewers are most accurate when they rate each answer to an interview question as it is given rather than waiting until the end of the interview to rate the candidate. They are less likely to forget or distort relevant information when they use this approach. Please follow the videotaped interview, when it is shown, by reading the questions from the attached list when they are asked by the interviewer. As the applicant answers the question, please rate the applicant's answer on the scale below that question by circling the most appropriate number. Do <u>not</u> wait until the end of the interview to make your ratings. You will not have much time to make your rating before the next question is asked so please use this time to familiarize yourself with the answer descriptions below each scale. You should be able to rate each answer soon after it is given.

You will be viewing the first taped interview shortly. Please remember to <u>circle</u> the most appropriate number as the answer is given.

<div align="center">

PLEASE <u>REVIEW</u> THE QUESTIONS ON THE FOLLOWING PAGES NOW

DO <u>NOT</u> PROVIDE ANY RATINGS UNTIL THE VIDEOTAPE STARTS

</div>

QUESTION 1

Tell me about your last supervisor. How well did work together? What is your impression of his or her management abilities? Did you ever disagree with his or her decisions? If so, what did you do?

1	2	3	4	5
Argued with decisions and ideas; Was critical and berated supervisor openly	Was reluctant to follow supervisor's directions; Was unenthusiastic	Had formal relations; Was aloof and detached; Did only what was necessary	Was conscientious and responsible in trying to follow supervisor's directions	Cooperated; Was concerned with supervisor's needs; Was honest but tactful

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 2

Give me an example of a difficult decision you had to make in your last job. Tell me what you decided and why. In general, when you had to make difficult decisions, did you seek advice from others or did you make the decision on your own?

1	2	3	4	5
Refused to make decision and wanted someone else to take responsibility; Proceeded as if oblivious to procedures and ignored others' suggestions	Asked constant questions; Was hesitant committing to the decision or made an inappropriate decision and did not seek advice	Had good knowledge of procedures and used good judgment in making a timely decision; Usually consults appropriate resources prior to making decision	Made appropriate decision independently; Sought advice in very unusual or sensitive situations; Recognized own limitations	Made appropriate decision, even in sensitive situation or assisted others in making decisions but recognizes own limitations

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 3

Have you ever been in a situation where a number of people or tasks demanded your attention at the same time? If so, describe the situation and how you handled it.

1	2	3	4	5
Became distraught and frustrated under pressure; Blew up at others or blamed others for situation	Ignored other tasks; Left people waiting; Tried to withdraw or escape the situation; Gave up	Used variety of tactics to distract self from the task at hand; delayed action as long as possible	Accepted concurrent demands as part of the job; Remained calm and dealt with one issue at a time	Prioritized and refocused to handle most important or urgent demands first; Assured those waiting they would be looked after first

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 4

Have you ever been in a situation where you knew that you could not meet the deadline? If so, how often did this occur and tell me about a typical situation and what you did.

1	2	3	4	5
Often missed deadlines and created crisis situations as a result; Did not complete work or work was inaccurate; Buried under paperwork; Required close monitoring	Frequently piled up overdue work and tried to work just before deadline; Work was partly completed (details missed); Viewed work as unimportant	Generally reviewed work and met deadlines but occasionally created extra work when the deadline was missed; Needed occasional reminders; Work was usually accurate	Kept track of and met daily, weekly, and monthly deadlines; Followed through on assigned tasks and tied together relevant information	Established a work plan and paced the workload to complete tasks well before the deadline; Kept others apprised when appropriate; Did unassigned work in anticipation of future needs

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 5

In your last job, how did you go about ensuring that all your work would be done on time (Did you use any kind of system?)?

1	2	3	4	5
Work was often overdue; Worked on several tasks at once; Did not use any kind of organizing system; Required frequent reminders	Accumulated work; Worked reactively; Own resources were too disorganized to be useful; Asked others to provide answers or information	Did most work as it arose; Collected and kept resources at hand; Used own resources to provide answers or information; Occasionally needed reminders	Reviewed and prioritized work regularly and did required work; Developed a well-organized system to easily access information; Rarely needed reminders	Anticipated time needed for future tasks; Had spare time; System of organizing was very logical and easy to use; Others could use it to find what they needed

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	2	3	4	5
Poor				Excellent

PLEASE READ THE RESUME AND REFERENCE CHECKS FOR CANDIDATE #2 NOW

TURN TO THE NEXT PAGE NOW

APPENDIX I: GENERAL QUESTIONS AND GRAPHIC RATING SCALES

INTERVIEW QUESTIONS

QUESTION 1
What do you think the role of social services should be?

1	2	3	4	5

Poor   Excellent

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 2
What types of people do you like most? What types do you like least?

1	2	3	4	5

Poor   Excellent

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 3
How would other people describe you as an individual?

1	2	3	4	5

Poor   Excellent

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 4
What do you consider to be your strengths? What are your weaknesses?

1	2	3	4	5

Poor   Excellent

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

QUESTION 5
What is your greatest accomplishment?

| 1 | 2 | 3 | 4 | 5 |

Poor   Excellent

PLEASE CIRCLE YOUR RATING FOR THIS QUESTION NOW.

PLEASE PREPARE TO WATCH THE VIDEO INTERVIEW FOR CANDIDATE #1 NOW

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

DO NOT PROVIDE THIS RATING UNTIL AFTER THE INTERVIEW.

PLEASE CIRCLE YOUR OVERALL RATING OF THIS APPLICANT FOR THE POSITION OF SOCIAL SERVICES WORKER ON THE SCALE BELOW:

1	**2**	**3**	**4**	**5**

Poor   Excellent

PLEASE READ THE RESUME AND REFERENCE CHECKS FOR CANDIDATE #2 NOW

TURN TO NEXT PAGE NOW

APPENDIX J: DEMOGRAPHIC QUESTIONS

PLEASE RANK THE THREE APPLICANTS (WITH 1 BEING THE BEST OF THE THREE, 2 THE MIDDLE AND 3 THE WORST) FOR THE POSITION OF SOCIAL SERVICES WORKER:

Applicant #1 (Dianne):   __

Applicant #2 (Terri) :   __

Applicant #3 (Wilma) :   __

PLEASE EXPLAIN WHY YOU PROVIDED THESE RANKINGS (IN OTHER WORDS WHAT DIFFERENTIATED THE APPLICANTS?):

SOUND QUALITY

PLEASE ESTIMATE IN PERCENT (I.E., FROM 0% TO 100%) HOW MUCH OF EACH APPLICANT'S RESPONSE YOU HEARD AND UNDERSTOOD:

Applicant #1 (Dianne):   __

Applicant #2 (Terri) :   __

Applicant #3 (Wilma) :   __

BACKGROUND INFORMATION

Program of Study: _____

Year in Program: _____

Have you ever worked in the field of Social Work?

        NO__ YES__, If yes for how long? ___ (in years)

How many job interviews have you participated in as an <u>interviewee</u> (estimate if necessary)? ___

How many job interviews have you conducted as the <u>interviewer</u>? __

Please indicate your gender by marking an "X" in the appropriate space.

FEMALE ___ MALE __

        THANK-YOU FOR YOUR PARTICIPATION

REFERENCES

Amalfitano, J. G., and Kalt, N. C. 1977. Effects of eye contact on the evaluation of job applicants. Journal of Employment Counseling 14 : 46-48.

Anderson, C. W. 1960. The relation between speaking times and decision in the employment interview. Journal of Applied Psychology 44(4) : 267-268.

Anderson, N. R. 1991. Decision making in the graduate selection interview: An experimental investigation. Human Relations, 44(4) : 403-417.

Anderson, N. R., and Shackleton, V. 1990. Decision making in the graduate selection interview: A field study. Journal of Occupational Psychology, 63 : 63-76.

Arvey, R. D. 1979. Unfair discrimination in the employment interview: Legal and psychological aspects. Psychological Bulletin, 86(4) : 736-765.

Arvey, R. D., and Campion, J. E. 1982. The employment interview: A summary and review of recent research. Personnel Psychology, 35 : 281-322.

Arvey, R. D., and Faley, R. H. 1988. Fairness in selecting employees. Don Mills, ON: Addison-Wesley.

Arvey, R. D., Miller, H. E., Gould, R., and Burch, P. 1987. Interview validity for selecting sales clerks. Personnel Psychology, 40 : 1-12.

Avolio, B. J., and Barrett, G. V. 1987. Effects of age stereotyping in a simulated interview. Psychology and Aging, 2 : 56-63.

Barr, S. H., and Hitt, M. A. 1986. A comparison of selection decision models in manager versus student samples. Personnel Psychology, 39 : 599-617.

Bartlett, C. J., Bobko, P., Mosier, S. B., and Hannan, R. 1978. Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. Personnel Psychology, 31 : 233-241.

Beehr, T. A., and Gilmore, D. C. 1982. Applicant attractiveness as a perceived job-relevant variable in selection of management trainees. Academy of Management Journal, 25 : 607-617.

Belec, B. E., and Rowe, P. M. 1983. Temporal placement of information, expectancy, causal attributions and overall final judgements in employment decision making. Canadian Journal of Behavioural Science, 15(2) : 107-120.

Berkowitz, L., & Donnerstein, E. 1982. External validity is more than skin deep. American Psychologist, 37 : 245-257.

Bernstien, V., Hakel, M. D., & Harlan, A. 1975. The college student as interviewer: A threat to generalizability? Journal of Applied Psychology, 60(2) : 266-268.

Binning, J. F., & Barrett, G. V. 1989. Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 73 : 193-198.

Binning, J. F., Goldstein, M. A., Garcia, M. F., & Scattaregia, J. H. 1988. Effects of pre-interview impressions on questioning strategies in same- and opposite-sex employment interviews. Journal of Applied Psychology, 73 : 30-37.

Blankenship, M. H., & Cesare, S. J. 1993. Age fairness in the employment interview: A field study. In R. D. Arvey (Chair), Perceptions, Theories, and Issues of Fairness in the Employment Interview. Symposium presented at the One Hundred and First Annual Convention of the American Psychological Association, Toronto, Ontario.

Boehm, V. R. 1972. Negro-white differences in validity of employment and training selection procedures: Summary of research evidence. Journal of Applied Psychology, 56 : 33-39.

Bolster, B. I., & Springbett, B. M. 1961. The reaction of interviewers to favorable and unfavorable information. Journal of Applied Psychology, 45(2) : 97-103.

Campbell, J. P. 1986. Labs, fields and straw issues. In: Generalizing from Laboratory to Field Settings, ed. E. A. Locke, 269-279. Lexington, MA: Lexington Books.

Canadian Psychological Association. 1987. Guidelines for Educational and Psychological Testing. Ottawa, ON: Author.

Carlson, R. E. 1967. Selection interview decision: The effect of interviewer experience, relative quota situation, and applicant sample on interview decision. Personnel Psychology, 20 : 259-280.

Carlson, R. E., Thayer, P. W., Mayfield, E. C., & Peterson, D. A. 1971. Research on the selection interview. Personnel Journal, 50 : 268-275.

Carlston, D. E. 1980. The recall and use of traits and events in social inference processes. Journal of Experimental Social Psychology, 16 : 303-328.

Cash, T. F., Gillen, B., & Burns, D. S. 1977. Sexism and "beautyism" in personnel consultant decision making. Journal of Applied Psychology, 62 : 301-307.

Cesare, S. J., Dalessio, A., & Tannenbaum, R. T. 1988. Contrast effects for black, white, male and female interviewees. Journal of Applied Psychology, 18 : 1261-1273.

Cleary, T. A. 1968. Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 5 : 115-124.

Constantin, S. W. 1976. An investigation of information favorability in the employment interview. Journal of Applied Psychology, 61(6) : 743-749.

Conway, J. M., Jako, R. A., & Goodman, D. F. 1995. A meta-analysis of interrater and internal consistency reliability of selection interviews. Journal of Applied Psychology, 80(5) : 565-579.

Crocker, L., & Algina, J. 1986. Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

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

Cronshaw, S. F., & Wiesner, W. H. 1989. The validity of the employment interview: Models for research and practice. In: The Employment Interview: Theory, Research and Practice, eds. R. W. Eder & G. R. Ferris, 269-279. Newbury Park, CA: Sage Publications.

Dalessio, A. T., & Silverhart, T. A. 1994. Combining biodata test and interview information: Predicting decisions and performance criteria. Personnel Psychology, 47 : 303-315.

Dipboye, R. L. 1982. Self-fulfilling prophecies in the selection-recruitment interview. Academy of Management Review, 7(4) : 579-586.

Dipboye, R. L., Arvey, R. D., & Terpstra, D. E. 1977. Self and physical attractiveness of raters and applicants as determinants of resume evaluations. Journal of Applied Psychology, 62 : 288-294.

Dipboye, R. L., & Flanagan, M. F. 1979. Research settings in industrial/organizational psychology: Are findings in the field more generalizable than in the laboratory? American Psychologist, 34 : 141-150.

Dipboye, R. L., Fontenelle, G. A., & Garner, K. 1984. Effects the previewing the application on interview process and outcomes. Journal of Applied Psychology, 69(1) : 118-128.

Dipboye, R. L., Fromkin, H. L., & Wiback, K. 1975. Relative importance of sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 60 : 39-43.

Dipboye, R. L., & Gaugler, B. B. 1993. Cognitive and behavioural processes in the selection interview. In: Personnel Selection in Organizations, eds. N. Schmitt & W. C. Borman, 135-170. San Francisco, CA: Jossey-Bass Publishers.

Dipboye, R. L., Stramler, C. S. & Fontenelle, G. A. 1984. The effects of the application on recall of information from the interview. Academy of Management Journal, 27(3) : 561-575.

Dipboye, R. L., & Wiley, J. W. 1977. Reactions of college recruiters to interviewee sex and self-presentation style. Journal of Vocational Behaviour, 10 : 1-12.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. 1986. Policy capturing in the employment interview. Journal of Applied Psychology, 71 : 9-15.

Dougherty, T. W., Turban, D. B., & Callender, J. C. 1994. Confirming first impressions in the employment interview: A field study of interviewer behaviour. Journal of Applied Psychology, 79(5) : 659-665.

Doverspike, D., Carlisi, A. M., Barrett, G. V., & Alexander, R. A. 1983. Generalizability analysis of a point-method job evaluation instrument. Journal of Applied Psychology, 68(3) : 476-483.

Eder, R. W. & Buckley, M. R. 1988. The employment interview: An interactionist perspective. In: Research in Personnel and Human Resources Management: Volume 6, eds. G. R. Ferris & K. M. Rowland, 75-107. Greenwich, CT: JAI Press.

Eder, R. W., & Ferris, G. R. 1989. The Employment Interview: Theory, Research and Practice. Newbury Park, CA: Sage Publications.

Einhorn, H. J. 1972. Expert measurement and mechanical combination. Organizational Behaviour and Human Performance, 7 : 86-106.

Elliott, A. G. P. 1981. Sex and decision making in the selection interview: A real-life study. Journal of Occupational Psychology, 54 : 265-273.

Farr, J. L. 1973. Response requirements and primacy-recency effects in a simulated selection interview. Journal of Applied Psychology, 57(3) : 228-232.

Farr, J. L., & York, C. M. 1975. Amount of information and primacy-recency effects in recruitment decisions. Personnel Psychology, 28 : 233-238.

Fay, C. H., & Latham, G. P. 1982. Effects of training and rating scales on rating errors. Personnel Psychology, 35, 105-116.

Ferguson, J., & Fletcher, C. 1989. An investigation of some cognitive factors involved in person-perception during selection interviews. Psychological Reports, 64 : 735-745.

Flanagan, J. C. 1954. The critical incident technique. Psychological Bulletin, 51 : 327-358.

Fletcher, C. 1989. Impression management in the selection interview. In: Impression Management in the Organization, eds. R. A. Giacalone & P. Rosenfield , 269-281. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Forbes, R. J., & Jackson, P. R. 1980. Nonverbal behaviour and the outcome of selection interviews. Journal of Occupational Psychology, 53 : 65-72.

Forsythe, S., Drake, M. F., & Cox, C. E. 1985. Influence of applicant's dress on interviewer's selection decisions. Journal of Applied Psychology, 70(2) : 374-378.

Furnham, A., & Burbeck, E. 1989. Employment interview outcomes as a function of interviewers' experience. Perceptual and Motor Skills, 69 : 395-402.

Ghee, K. L., Rozelle, R. M., & Baxter, J. C. 1986. The effects of race similarity on interpersonal judgement in the Graduate Selection Interview. British Journal of Social Psychology, 25 : 277-285.

Giacalone, R. A., & Rosenfeld, P. 1986. Self-presentation and self-promotion in an organizational setting. Journal of Social Psychology, 126 : 321-326.

Gifford, R., Ng, C. F., & Wilkinson, M. 1985. Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgements. Journal of Applied Psychology, 70(4) : 729-736.

Gilmore, D. C., & Ferris, G. R. 1989. The effects of applicant impression management tactics on interviewer judgements. Journal of Management, 15(4) : 557-564.

Gordon, M. E. 1970. The effect of the correctness of behaviour observed on the accuracy of ratings. Organizational Behaviour and Human Performance, 5 : 366-377.

Gordon, M. E., Slade, L. A., & Schmitt, N. 1986. The "science of the sophomore" revisited: From conjecture to impression. Academy of Management Review, 11 : 191-207.

Gordon, R. A., Rozelle, R. M., & Baxter, J. C. 1988. The effect of applicant age, job level, and accountability on the evaluation of job applicants. Organizational Behaviour and Human Decision Processes, 41 : 20-33.

Gorman, C. D., Clover, W. H., & Doherty, M. E. 1978. Can we learn anything about interviewing real people from "interviews" of paper people? Two studies of the external validity of a paradigm. Organizational Behaviour and Human Performance, 22 : 165-192.

Green, P. C., Alter, P. & Carr, A. F. 1993. Development of standard anchors for scoring generic past-behaviour questions in structured interviews. International Journal of Selection and Assessment, 1(4).

Greenberg, J. 1987. The college sophomore as guinea pig: Setting the record straight. Academy of Management Review, 12(1) : 157-159.

Guilford, J. P. 1956. Fundamental Statistics in Psychology and Education (3rd edition). New York: McGraw Hill.

Haefner, J. E. 1977. Race, age, sex, and competence as factors in employer selection of the disadvantaged. Journal of Applied Psychology, 62 : 199-202.

Hakel, M. D. 1982. Employment interviewing. In: Personnel Management, eds. K. M. Rowland & G. R. Ferris. Boston, MA: Allyn and Bacon.

_____. 1989. The state of employment interview theory and research. In: The Employment Interview: Theory, Research and Practice, eds. R. W. Eder & G. R. Ferris, 285-293. Newbury Park, CA: Sage Publications.

Hakel, M. D., Dobmeyer, T. W., & Dunnette, M. D. 1970. Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. Journal of Applied Psychology, 54(1) : 65-71.

Hakel, M. D., Hollmann, T. D., & Dunnette, M. D. 1970. Accuracy of interviewers, certified public accountants, and students in identifying the interests of accountants. Journal of Applied Psychology, 54(2) : 115-119.

Hakel, M. D., Ohnesorge, J. P., & Dunnette, M. D. 1970. Interviewer evaluations of job applicants' resumes as a function of the qualifications of the immediately preceding applicants: An examination of contrast effects. Journal of Applied Psychology, 54(1) : 27-30.

Hamilton, D. L., Katz, L. B., & Leirer, V. O. 1980. Cognitive representation of personality impressions: Organizational processes in first impression formation. Journal of Personality and Social Psychology, 39(6) : 1050-1063.

Harris, M. M. 1989. Reconsidering the employment interview: A review of recent literature and suggestions for future research. Personnel Psychology, 42 : 691-726.

Heilman, M. E., & Saruwatari, L. R. 1979. When beauty is beastly: The effects of appearance and sex on evaluation of job applicants for managerial and nonmanagerial jobs. Organizational Behaviour and Human Performance, 23 : 360-372.

Heneman III, H. G., Schwab, D. P., Huett, D. L., & Ford, J. J. 1975. Interviewer validity as a function of interviewee structure, biographical data and interviewee order. Journal of Applied Psychology, 60(6) : 748-753.

Higgins, E. T., Herman, C. P., & Zanna, M. P. 1981. Social Cognition: The Ontario Symposium on Personality and Social Psychology Volume 1. Hillsdale, NJ: Erlbaum and Associates.

Hoff-Macan, T., & Dipboye, R. L. 1990. The relationship of interviewers' preinterview impressions to selection and recruitment outcomes. Personnel Psychology, 43 : 745-768.

Hollman, T. D. 1972. Employment interviewers' errors in processing positive and negative information. Journal of Applied Psychology, 56 : 130-134.

Huffcutt, A. I., & Arthur, W. A. 1994. Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. Journal of Applied Psychology, 79(2) : 184-190.

Huguenard, T., Sager, E. B., & Ferguson, L. W. 1970. Interview time, interview set, and interview outcome. Perceptual and Motor Skills, 31 : 831-836.

Hunter, J. E., & Schmidt, F. L. 1990. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Newbury Park, CA: Sage.

Ilgen, D. R. 1986. Laboratory research: A question of when, not if. In: <u>Generalizing from Laboratory to Field Settings: Research Findings from Industrial-Organizational Psychology, Organizational Behaviour, and Human Resource Management</u>, ed. E. A. Locke, 257-267. Toronto, ON: Lexington Books.

Imada, A. S., & Hakel, M. D. 1977. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. <u>Journal of Applied Psychology</u>, 62 : 295-300.

Janz, J. T. 1982. Initial comparisons of patterned behaviour description interviews versus unstructured interviews. <u>Journal of Applied Psychology</u>, 67 : 577-580.

Janz, J. T., Hellervik, L., & Gilmore, D. C. 1986. <u>Behaviour Description Interviewing: New, Accurate, Cost Effective</u>. Newton, MA: Allyn and Bacon.

Jensen, A. R. 1980. <u>Bias in Mental Testing</u>. New York, NY: The Free Press.

Johns, G. 1975. Effects of informational order and frequency of applicant evaluation upon linear information-processing competence of interviewers. <u>Journal of Applied Psychology</u>, 60 : 427-433.

Kataoka, H., Latham, G. P., & Whyte, G. 1995. <u>The relative resistance of the situational, patterned behaviour, and conventional structured interviews to anchoring effects</u>. (Manuscript submitted for publication).

Keenan, A., & Wedderburn, A. A. I. 1975. Effects of non-verbal behaviour of interviewers on candidates' impressions. <u>Journal of Occupational Psychology</u>, 48 : 129-132.

Kinicki, A. J., Lockwood, C. A., Hom, P. W., & Griffeth, R. W. 1990. Interviewer predictions of applicant qualifications and interviewer validity: Aggregate and individual analyses. <u>Journal of Applied Psychology</u>, 75(5) : 477-486.

Landy, F. J., & Bates, F. 1973. Another look at contrast effects in the employment interview. <u>Journal of Applied Psychology</u>, 58(1) : 141-144.

Langdale, J. A., & Weitz, J. 1973. Estimating the influence of job information on interviewer agreement. <u>Journal of Applied Psychology</u>, 57(1) : 23-27.

Latham, G. P., & Saari, L. M. 1984. Do people do what they say? Further studies on the situational interview. Journal of Applied Psychology, 69(4) : 569-573.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. 1980. The situational interview. Journal of Applied Psychology, 65 : 422-427.

Latham, G. P., & Skarlicki, D. P. 1995. The relative effectiveness of the situational and patterned behaviour description interviews in minimizing the similar-to-me bias of Canadian Francophone managers. (Manuscript submitted for publication).

Latham, G. P., Wexley, K. N., & Pursell, E. D. 1975. Training managers to minimize rating errors in the observation of behaviour. Journal of Applied Psychology, 60 : 550-555.

Leonard Jr, R. L. 1976. Cognitive complexity and the similarity-attraction paradigm. Journal of Research in Personality, 10 : 83-88.

Liden, R. C., Martin, C. L., & Parsons, C. K. 1993. Interviewer and applicant behaviours in employment interviews. Academy of Management Journal, 36(2) : 372-386.

Liden, R. C., & Parsons, C. K. 1989. Understanding interpersonal behaviour in the employment interview: A reciprocal interaction analysis. In: The Employment Interview: Theory, Research and Practice, eds. R. W. Eder & G. R. Ferris, 219-231. Newbury Park, CA: Sage Publications.

Lin, T. R., Dobbins, H. G., & Farr, J. L. 1992. A field study of race and age similarity effects on interview ratings in conventional and situational interviews. Journal of Applied Psychology, 77 : 363-371.

Locke, E. A. 1968. Toward a theory of task motivation and incentives. Organizational Behaviour and Human Performance, 3 : 157-189.

London, M., & Hakel, M. D. 1974. Effects of applicant stereotypes, order, and information on interview impressions. Journal of Applied Psychology, 59(2) : 157-162.

Macan, T. M., & Dipboye, R. L. 1988. The effects of interviewers' initial impressions on information gathering. Organizational Behaviour and Human Decision Processes, 42 : 364-387.

Maas, J. B. 1965. Patterned scaled expectation interview: Reliability studies on a new technique. Journal of Applied Psychology, 49(6) : 431-433.

Marchese, M. C., & Muchinsky, P. M. 1993. The validity of the employment interview: A meta-analysis. International Journal of Selection and Assessment, 1 : 18-26.

Marlowe, C. M, Schneider, S. L., Nelson, C. E. 1996. Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? Journal of Applied Psychology, 81 : 11-21.

Maurer, S. D., & Fay, C. 1988. Effect of situational interviews, conventional structured interviews, and training on interview rating agreement: An experimental analysis. Personnel Psychology, 41 : 329-344.

Mayfield, E. C. 1964. The selection interview - A re-evaluation of published research. Personnel Psychology, 17 : 239-260.

Mayfield, E. C., & Carson, R. E. 1966. Selection interview decisions: First results from a long-term research project. Personnel Psychology, 19 : 41-55.

McCormick, E. J., Jeanneret, P. R., & Meecham, R. C. 1972. A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology, 56 : 347-368.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., Hunter, J. E., Maurer, S., & Russell, J. 1986. The validity of employment interviews: A review and meta-analysis. Manuscript submitted for publication.

McDonald, T., & Hakel, M. D. 1985. Effects of applicant race, sex, suitability, and answers on interviewer's questioning strategy and ratings. Personnel Psychology, 38 : 321-334.

McGovern, T. V., & Tinsley, H. E. 1978. Interviewer evaluations of interviewee nonverbal behaviour. Journal of Vocational Behaviour, 13 : 163-171.

McMurry, R. N. 1947. Validating the personnel interview. Personnel, 23 : 263-272.

Meehl, P. E. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46(4) : 806-834.

Moore, L. F., & Lee, A. J. 1974. Comparability of interviewer, group, and individual interview ratings. Journal of Applied Psychology, 59(2) : 163-167.

Motowidlo, S. J. 1986. Information processing in personnel decisions. In: Research in Personnel and Human Resources Management: Volume 4, eds. K. M. Rowland & G. R. Ferris, 1-44. Greenwich, CT: JAI Press.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. 1992. Studies of the structured behavioural interview. Journal of Applied Psychology, 77(5) : 571-587.

Nunnally, J. C. 1967. Psychometric Theory. New York: McGraw Hill.

Olian, J. D. 1986. Staffing. In: Generalizing from Laboratory to Field Settings: Research Findings from Industrial-Organizational Psychology, Organizational Behaviour, and Human Resource Management, ed. E. A. Locke, 13-42. Toronto, ON: Lexington Books.

Orpen, C. 1985. Patterned behaviour description interviews versus unstructured interviews: A comparative validity study. Journal of Applied Psychology, 70 : 774-776.

Osburn, H. G., Timmrick, C. & Bigby, D. 1981. Effect of dimensional relevance on accuracy of simulated hiring decisions by employment interviewers. Journal of Applied Psychology, 66 : 159-165.

Parsons, C. K., & Liden, R. C. 1984. Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. Journal of Applied Psychology, 69(4) : 557-568.

Paunonen, S. V., & Jackson, D. J. 1987. Accuracy of interviewers and students in identifying the personality characteristics of personnel managers and computer programmers. Journal of Vocational Behaviour, 31 : 26-36.

Peters, L., O'Connor, E. J., & Eulberg, J. 1985. Situational constraints: Sources, consequences, and future considerations. In: Research in Personnel and Human Resources Management Volume 3, eds. K. M. Rowland & G. R. Ferris, 79-114. Greenwich, CT: JAI Press.

Peters, L. H., & Terborg, J. R. 1975. The effects of temporal placement of unfavorable information and of attitude similarity on personnel selection decisions. Organizational Behaviour and Human Performance, 13 : 279-293.

Peterson, D. K., & Pitz, G. F. 1986. Effect of input from a mechanical model on clinical judgement. Journal of Applied Psychology, 71 : 163-167.

Phillips, A. P., & Dipboye, R. L. 1989. Correlational tests of predictions from a process model of the interview. Journal of Applied Psychology, 74(1) : 41-52.

Popper, K. 1981. Science, Pseudo-science and falsifiability. In: On Scientific Thinking, eds. R. D Tweney, M. E. Doherty & C. R. Mynatt. New York, NY: Columbia University Press.

Prewett-Livingston, A. J., Feild, H. S., Veres, J. G., & Lewis, P. M. 1996. Effects of race on interview ratings in a situational panel interview. Journal of Applied Psychology, 81(2) : 178-186.

Pulakos, E. D., Schmitt, N., Whitney, D., & Smith, M. 1996. Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on the validity of a structured interview. Personnel Psychology, 49 : 85-102.

Pursell, E. D., Campion, M. A., & Gaylord, S. R. 1980. Structured interviewing: Avoiding selection problems. Personnel Journal, 59 : 907-912.

Rand, T. M., & Wexley, K. N. 1975. Demonstration of the effect "similar to me" in simulated employment interviews. Psychological Reports, 36 : 535-544.

Rasmussen, Jr., K. G. 1984. Nonverbal behaviour, verbal behaviour, resume credentials and selection interview outcomes. Journal of Applied Psychology, 69(4) 551-556.

Raza, S. M., & Carpenter, B. N. 1987. A model of hiring decisions in real employment interviews. Journal of Applied Psychology, 72(4) : 596-603.

Riggio, R. E., & Throckmorton, B. 1988. The relative effects of verbal and nonverbal behaviour, appearance, and social skills on evaluations made in hiring interviews. Journal of Applied Social Psychology, 18 : 331-348.

Rosen, B., & Jerdee, T. H. 1976. The influence of age stereotypes on managerial decisions. Journal of Applied Psychology, 61 : 428-432.

Rowe, P. M. 1967. Order effects in assessment decisions. Journal of Applied Psychology, 51 : 170-173.

_____. 1984. Decision processes in personnel selection. Canadian Journal of Behavioural Science, 16 : 326-337.

Rowe, P. M., Williams, M. C., & Day, A. L. 1994. Selection procedures in North America. International Journal of Selection and Assessment, 2(2).

Rozelle, R. M., & Baxter, J. C. 1981. Influence of role pressures on the perceiver: Judgements of videotaped interviews varying judge accountability and responsibility. Journal of Applied Psychology, 66 : 437-441.

Sackett, P. R. 1982. The interviewer as hypothesis tester: The effects of impressions of an applicant on interviewer questioning strategy. Personnel Psychology, 35 : 789-804.

Sawyer, J. 1966. Measurement and prediction, clinical and statistical. Psychological Bulletin, 66 : 178-200.

Schmitt, N. 1976. Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 29 : 79-101.

Schuh, A. J. 1978. Effects of early interruption and note taking on listening accuracy and decision making in the interview. Bulletin of the Psychonomic Society, 12 : 242-244.

Schwab, D. P., & Heneman III, H. G. 1969. Relationship between interview structure and interviewer reliability in an employment situation. Journal of Applied Psychology, 53(3) : 214-217.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. 1989. Generalizability Theory. American Psychologist, 44(6) : 922-932.

Shaw, E. A. 1972. Differential impact of negative stereotyping in employee selection. Personnel Selection, 25 : 333-338.

Siegel, J. M., & Steele, C. M. 1980. Environmental distraction and interpersonal judgments. British Journal of Social and Clinical Psychology, 19 : 23-32.

Singer, M., & Eder, G. S. 1989. Effects of ethnicity, accent, and job status on selection decisions. International Journal of Psychology, 24 : 13-34.

Singer, M. S., & Sewell, C. 1989. Applicant age and selection interview decisions: Effect of information exposure on age discrimination in personnel selection. Personnel Psychology, 42 : 135-154.

Smith, P. C., & Kendall, L. M. 1963. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47 : 149-155.

Snyder, M., & Cantor, N. 1979. Testing hypotheses about other people: The use of historical knowledge. Journal of Experimental Social Psychology, 15 : 330-342.

Snyder, M., & Swann, Jr., W. B. 1978. Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology, 36(11) : 1202-1212.

Springbett, B. M., 1958. Factors affecting the final decision in the employment interview. Canadian Journal of Psychology, 12(1) : 13-22.

Srull, T. K., & Wyer, Jr., R. S. 1980. Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgements. Journal of Personality and Social Psychology, 6 : 841-856.

Stevens, C. K., & Kristof, A. L. 1995. Making the right impression: A field study of applicant impression management during job interviews. Journal of Applied Psychology, 80(5) : 587-606.

Stone, T. H., & Meltz, N. M. 1988. Human Resource Management in Canada (2nd Edition). Toronto, ON: Holt, Rinehart and Winston of Canada.

Sulsky, L. M., & Balzer, W. K. 1988. Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73(3), 497-506.

Sydiaha, D. 1961. Bales' interaction process analysis of personnel selection interviews. Journal of Applied Psychology, 45, 393-401.

Tessler, R., & Sushelsky, L. 1978. Effects of eye contact and social status on the perception of a job applicant in an employment interviewing situation. Journal of Vocational Behaviour, 13 : 338-347.

Tucker, D. H., & Rowe, P. M. 1979. Relationship between expectancy, causal attributions, and final hiring decisions in the employment interview. Journal of Applied Psychology, 64(1) : 27-34.

Tullar, W. L., Mullins, T. W., & Caldwell, S. A. 1979. Effects of interview length and applicant quality on interview decision time. Journal of Applied Psychology, 64 : 669-674.

Ulrich, L & Trumbo, D. 1965. The selection interview since 1949. Psychological Bulletin, 63 : 100-116.

Vance, R. J., Kuhnert, K. W., & Farr, J. L. 1978. Interview judgments: Using external criteria to compare behavioural and graphic scale ratings. Organizational Behaviour and Human Performance, 22 : 279-294.

Wagner, R. 1949. The employment interview: A critical summary. Personnel Psychology, 2 : 17-46.

Weekley, J. A., & Gier, J. A. 1987. Reliability and validity of the situational interview for a sales position. Journal of Applied Psychology, 72(3) : 484-487.

Wernimont, P. F., & Campbell, J. P. 1968. Signs, sample and criteria. Journal of Applied Psychology, 52 : 372-376.

Wexley, K. N., & Nemeroff, W. F. 1974. The effects of racial prejudice, race of applicant, and biographical similarity on interviewer evaluations of job applicants. Journal of Social and Behaviour Sciences, 20 : 66-78.

Wexley, K. N., Sanders, R. E., & Yukl, G. A. 1973. Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 57 : 233-236.

Wexley, K. N., Yukl, G. A., Kovacs, S. Z., & Sanders, R. E. 1972. Importance of contrast effects in employment interviews. Journal of Applied Psychology, 56(1), 45-48.

Wiesner, W. H. 1988. Improving the criterion-related validity and reliability of the structured interview: The contributions of job relevance, timing and rating scale. Unpublished doctoral dissertation, University of Waterloo.

_____. 1989. The contributions of job relevance, timing, and rating scale to the validity of the employment interview. In S. F. Cronshaw (Chair), Improving interview validity and legal defensibility through structuring. Symposium conducted at the 50th Annual Convention of the Canadian Psychological Association, Halifax, Nova Scotia.

Wiesner, W. H., & Cronshaw, S. F. 1988. A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. Journal of Occupational Psychology, 61 : 275-290.

Wiesner, W, H., & Oppenheimer, R. J. 1990. An investigation of the effectiveness of note-taking versus the use of rating scales in employment interview decisions. In: Interview Judgement Processes: Edward Webster's Vision and Current Research Efforts, Chairs K. M. Kacmar & R. W. Eder. Symposium presented at the Fifth Annual Conference of the Society for Industrial and Organizational Psychology, Miami Beach, Florida.

_____. 1993. Effects of interview structure and member dominance and accuracy on the validity of employment interview board decisions. Unpublished manuscript.

Wright, O. R. 1969. Summary of research on the selection interview since 1964. <u>Personnel Psychology</u>, <u>36</u>, 355-370.

Wright, P. 1974. The harassed decision maker: Time pressures, distractions, and the use of evidence. <u>Journal of Applied Psychology</u>, <u>59(5)</u>, 555-561.

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. 1989. The structured interview: Additional studies and a meta-analysis. <u>Journal of Occupational Psychology</u>, <u>62</u>, 191-199.

Wyer, R. S., Jr., & Srull, T. K. 1981. Category accessibility: Some theoretical and empirical issues concerning the processing of social information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), <u>Social Cognition: The Ontario Symposium on Personality and Social Psychology</u>. Hillsdale, NJ: Erlbaum.

Young, D. M., & Beier, E. G. 1977. The role of applicant nonverbal communication in the employment interview. <u>Journal of Employment Counseling</u>, <u>14</u>, 154-165.

Zikmund, W. G., Hitt, M. A., & Pickens, B. A. 1978. Influence of sex and scholastic performance on reactions to job applicant resumes. <u>Journal of Applied Psychology</u>, <u>63</u>, 252-255.

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"
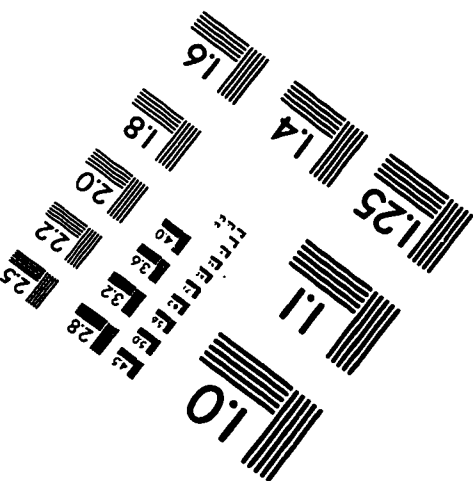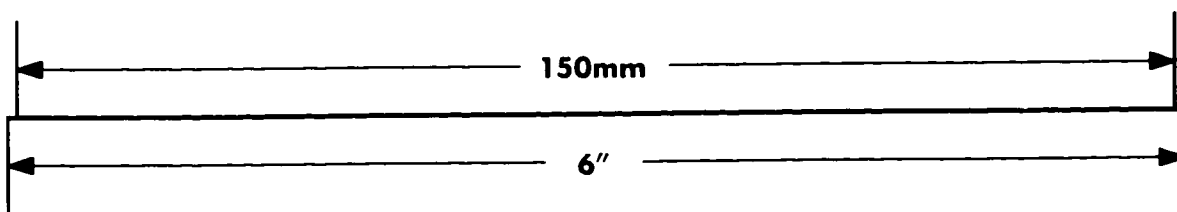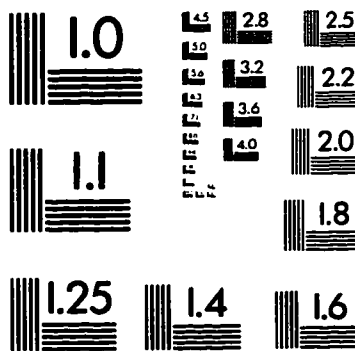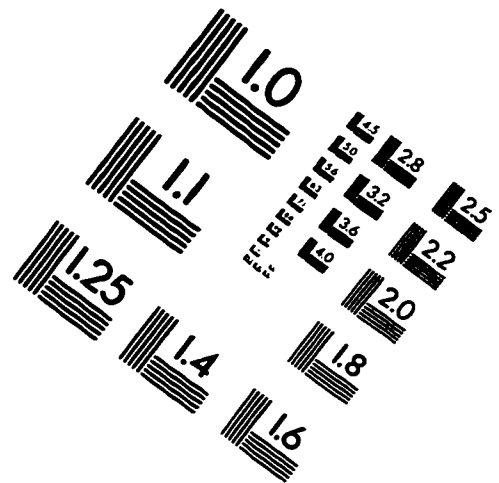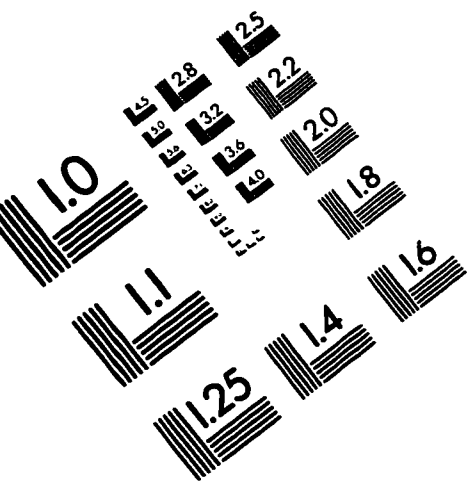
APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989