A Comparative Validation Study

of the

Behavioral Description Interview and Situational Interview

© Copyright

Copyright by

Maxon R. Mosher

September, 1991

Submitted in partial fulfillment of

the requirements for the degree in

Master of Applied Science

(I/O Psychology)

Saint Mary's University

Halifax, Nova Scotia
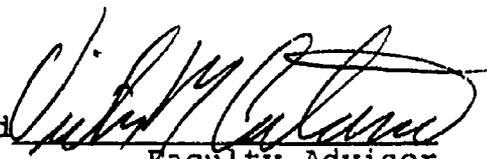
Approved _____
Faculty Advisor
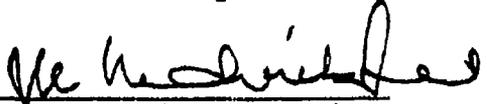
Approved _____
Committee Member

Approved _____
Committee Member

Canadä

Acknowledgements

Table of Contents

List of Tables

Abstract

Behavioral Description and Situational questions were developed based on a critical incident job analysis of initial level military training. Interviewers were 12 Military Career Counsellors (MCCs) and the interviewees were 53 Canadian Forces applicants. All were interviewed using the traditional interview and the Behavioral Description Interview (BDI), 20 were also interviewed using the Situational Interview (SI), and 15 independent ratings of the BDI responses were also obtained. The criterion measure for all of the interviews was a specifically designed rating form based on the same dimensions identified in the critical incident analysis. The BDI was found to be valid and reliable (r=.44, p<.01; interrater reliability .69, p<.01). Neither the SI nor the traditional interview produced significant validities. Small sample size and methodological problems may have affected the results of the SI, but other factors such as the nature of the success criteria and experience level of the applicant population may also have had an impact.

Introduction

Background

The interview is an integral and important part of standard personnel selection procedures, and recent surveys indicate that it is currently used by over 99% of organizations (Wiesner, 1991). The Canadian Forces (CF) is no exception, as the interview is considered to be the main tool for assessing the future performance of both officer and non-officer applicants. Martin (1972) examined the traditional CF interview procedure, and found its predictive validity to be relatively low (r=.14) using success/release in recruit training as the criterion measure. Recent examinations of interview procedures used to evaluate the success of CF naval officers (Bradley, 1990; Okros, Johnston, and Rodgers, 1988) suggested that the board interview was a poor predictor of future success in basic officer and initial level naval officer training, and did not add to the validity of the overall selection process.

Two important issues are currently impacting on the CF and other organizations, and they have direct relevance to the use of the employment interview, particularly considering its questionable predictive validity:

1. Efficiency. The current political and economic

climate in Canada has placed certain constraints on the military as well as on other businesses and organizations. This "lean and mean" approach translates into a requirement for the CF recruiting system to process applicants efficiently and in a manner that produces maximum utility using a top-down selection approach. This requires that concerns about interview validity and utility be addressed.

2. Human Rights. The CF, like other organizations, is subject to certain regulations intended to ensure the fair and equitable treatment of all citizens in hiring and other situations. Human Rights legislation in both the U.S. and Canada has focused attention on the interview, and has resulted in a number of legal challenges which may make the issues of validity and job-relevance of the interview even more important in the future (Cronshaw, 1989).

Purpose

The purpose of this research is to examine two recent and promising approaches to structured interviewing, namely the Situational Interview (SI) and the Behavior Description

will be examined, and they will be compared and contrasted with the current traditional interview approach. As well, the study will examine their potential value in addressing Utility and Human Rights concerns.

The Reviews

A considerable body of information and research has been amassed, which casts doubt on the reliability and validity of the employment interview. Numerous reviews have documented these problems, and they have been the subject of countless published and unpublished research papers, particularly since World War Two. Arvey and Campion (1982), Harris (1989), Mayfield (1982), Ulrich and Trumbo (1965), Wagner (1949), Webster (1964 & 1982), and Wright (1969) represent the most influential reviews (and research reports) and, for the most part, they are all in general agreement that numerous and serious problems exist. They also point out that research efforts have shed some light on the processes encountered in the interview, but only recently has there been any indication of possible significant improvement in the areas of reliability and validity.

Wagner (1949) presented one of the first comprehensive reviews of research on the employment interview. He examined some 106 articles, the earliest of which was a study by Scott in 1915 in which low reliability was reported

for evaluations of 26 sales applicants interviewed by six personnel managers. Overall, this review found validity coefficients which ranged from .09 to .94 with a median of .27 and reliabilities which ranged from .23 to .97 with a median of .57. Even in this early article, Wagner indicated encouraging results with the use of standardized or patterned interview approaches.

Mayfield (1964) produced the next major review on the employment interview. In it, he advocated the use of a micro-analytic approach to better understand which factors were influencing the interviewers' judgements. He reported a number of research findings, the most important of which are:

1. General ratings on unstructured interviews have low reliability.

2. Material is not covered consistently in unstructured interviews.

3. Different interviewers rate the same information differently.

4. Structured interviews result in higher interrater reliability.

5. Interview validity is low.

6. If the interviewer has valid test information, predictions based on the interview plus test information are no better than those based on test results alone.

information are no better than those based on test results alcne.

7. Interviewers can assess intelligence reliably and validly, but have not been able to do so with other traits.

8. The form of the question affects the answers given.

9. The attitude of the interviewer affects the interpretation of the interviewee's responses.

10. In unstructured interviews, interviewers tend to talk most.

11. Interviewers are influenced more by unfavourable than favourable information.

12. Interviewers make their decisions quite early in unstructured interviews.

A review by Ulrich and Trumbo (1965) published only six months later, reached conclusions which were not greatly different from those of Mayfield (1964). They suggested more limited evaluation goals for the interview, in that they considered the interview had greater possibilities for predicting certain performance areas than others. They also suggested a structured approach to interviewing which they felt would improve reliability and focus on specific job relevant areas and, further, they recommended the examination of interview results separately from other sources of information such as employment tests, which could

the interview would be most useful in assessing
interpersonal relations and career motivations.

Reviews by Wright (1969) and Schmitt (1976) both dealt
with decision making in the interview and relied heavily on
the significant research conducted by Webster (1964) and his
graduate students at McGill University, who used Canadian
Army Personnel Selection Officers as subjects. Several
specific variables and related findings from these studies
are worthy of note:

1. <u>Information Favorability</u>. Interviewers place more
   weight on unfavourable rather than favourable
   information. Based on the work done by Webster,
   it was suggested that this latter tendency was
   attributable to the preponderance of negative
   rather than positive feedback provided to
   interviewers by their superiors.

2. <u>Temporal Placement of Information</u>. Interviewers
   reach a final decision quite early in the
   interview, typically within the first four
   minutes. Impressions formed early in the
   interview are more important than other factual
   information in determining the judgement of the
   interviewer.

3. <u>Interviewer Stereotypes</u>. Interviewers possess
   stereotypes of idealized successful candidates
   against which interviewees are judged.

4. <u>Job Information</u>. Interrater reliability increases
    as interviewers have more information about the
    job to be filled. There is a concomitant
    reduction in the impact of irrelevant attributes.

5. <u>Individual Differences in the Decision Process</u>.
    Different interviewers use different processes and
    weigh information differently in reaching
    decisions. This in turn results in some of the
    observed interrater differences.

6. <u>Visual Clues</u>. Nonverbal sources of information
    are more important than verbal clues.

7. <u>Attitudinal, Sexual and Racial Similarity</u>.
    Certain similarities between the interviewer and
    interviewee will impact upon the evaluation.
    Females were given lower ratings although both
    males and females were more likely to be
    recommended for role-congruent jobs.

8. <u>Contrast Effects</u>. The rating of a candidate is
    affected by the quality of preceding interviewee.

9. <u>Structured Interview Guides</u>. Interrater
    reliability is enhanced through the use of
    structured interview guides.

10. <u>Miscellaneous</u>. A number of other relevant finding
    were presented showing that experience did not
    increase reliability, pressure to meet quotas
    influenced decisions of experienced interviewers

more than inexperienced ones, and that such factors as appearance and personal history had some effect on decisions.

Arvey and Campion (1982) reviewed and summarized more recent research. They noted that Landy (1976) produced reasonably favourable results through the use of board interviews for the selection of police officers, although a statistically derived predictor, using averaged interview factor scores identified through principle components analysis, rather than the overall recommendations of the interviewers, was the only predictor that produced valid results. Further, Arvey and Campion reported on several studies which resulted in improvements in validity as a consequence of the use of directly related job analyses and other job information as a basis for highly structured interview questions.

The most recent comprehensive review by Harris (1989) followed an organization similar to Arvey and Campion (1982) in examining validity, methodological issues, decision making, applicant characteristics, and interview training, as well as making recommendations for future research. Harris noted shortcomings in methodology such as small sample sizes, the use of paper and pencil interviewees, and the preponderance of studies using students as subjects. Harris also remarked on the dearth of studies focusing on decision making theories. With respect to applicant

characteristics, Harris confirmed the potentially negative influence of demographic variables, nonverbal behavior, and other personal characteristics. He also reviewed the effects of both interviewer and interviewee training, but results were largely inconclusive. In discussing validity, the rather positive results achieved by several highly structured interview approaches were highlighted. These included the situational interview (SI; Latham, Saari, Pursell, and Campion, 1980; Latham and Saari, 1984), the behavior description interview (BDI; Janz, 1982; Janz, Hellervik, and Gilmore, 1986) and the comprehensive structured interview (CSI; Pursell, Campion, and Gaylord. 1980). This latter approach used questions similar to the SI and items that might best be viewed as job knowledge test items. The first two types of structured interview approaches will be discussed in greater detail later, in that they are the approaches to interviewing which are the subject of this study.

Meta-Analytic Studies

Both Arvey and Campion (1982) and Harris (1989) noted recent meta-analytic studies undertaken with the idea that possible problems with research on the interview may be associated with the approaches taken to examining and cumulating research findings. Hunter, Schmidt, and Jackson (1982) in fact, suggest that standard research review

practices can lead to conflicting results and false conclusions which may be entirely artifactual. Hunter et al. point out that subjective, narrative reviews fail to consider the different sample sizes and do not adjust for potential sources of error variance such as sampling error, error of measurement (unreliability of predictor and criterion measures), and range restriction. Utilizing the meta-analytic approach, which corrects for these potential sources of error, Hunter and Hunter (1984) reanalysed several studies which focused on the validity of the employment interview. In one instance they obtained an average validity of .16 and in the other .23. They also reported on a new meta-analysis they had conducted based on three more recent and comprehensive reviews. With supervisor ratings as the criterion, and a sample size of 2,694, they estimated an average validity of .14 for the employment interview. With promotion, training success, and tenure as criterion measures, average validities ranged from .03 to .10. These results are reasonably consistent with previous findings, although sample sizes were still relatively small (by meta-analytic standards).

The most recent and comprehensive study utilizing meta-analytic techniques was conducted by Wiesner and Cronshaw (1988). Utilizing 150 validity coefficients (which represented a sample of over 50,000 interviews), and using the same techniques as those described by Hunter et al.

(1982), several validity estimates were obtained. The population estimate of the validity of the interview, considering all types of interviews and all categories of criterion measures, was found to be .47, (corrected for direct restriction of range and criterion unreliability; .26 uncorrected). These values for the validity of the employment interview were considerably larger than those reported by Hunter and Hunter (1984). However, since less than 75% of the variance in correlations could be explained by pooling all interview validity coefficients and correcting for sampling error, attenuation, and restriction of range, Wiesner and Cronshaw undertook a search for possible moderator variables. Interview type (individual, board, structured, unstructured and the various combinations thereof) was examined and found to be a significant moderator. In the case of the individual structured interview (operationally defined as questions with predetermined answers, rating scales for the answers, and statistical combination of the ratings), a mean population validity estimate of .63 (corrected; .35 uncorrected) was determined. The meta-analysis revealed the presence of still other moderating variables however, suggesting that the category of interview type could be broken down further and that considerable variance existed across the sub-categories. Subsequent analysis showed that within the category of individual structured interviews, those based on

formal job analyses were the strongest, with a mean validity coefficient of .87 (corrected; .48 uncorrected).

The meta-analytic approach generally supported the findings of other researchers and qualitative reviews, that structure in the interview is of primary importance to increased validity. Considering the findings of the previous reviews, one might also reasonably conclude that within this broad category, the best results will occur when the interview questions are based on thorough job analyses and are directly related to behaviors found to be important to good performance on the job. Although similar approaches have been used in the development and validation of employment tests and situational tests, relatively few efforts have been made to improve the interview by utilizing such an approach.

Behavior Consistency

It is interesting to note that Wernimount and Campbell (1968), in a classic article entitled Signs, Samples, and Criteria, challenged the traditional notion of validity in which predictors and criteria are different (eg. predicting performance in a skilled trade area from personality traits, attitudes, verbal fluency, etc.). They argued that psychology should return to its original mandate and restrict itself to the measurement of behaviors. As well, they reaffirmed the familiar wisdom that the best predictor

of future performance is past performance (behavioral consistency approach). They went on to point out, that in a selection situation, this sampling of behavior at two points in time makes the comparison of measures more akin to reliability than the popular idea of validity. Schmitt and Ostroff (1986) attempted to operationalize this behavioral consistency approach by delineating a standardized approach to the development of tests, job samples, and interviews. With respect to the interview, the proposed process was very similar to the SI developed and reported on by Latham et al. (1980) and was based on concepts not unlike those of the BDI developed by Janz et al. (1986).

Situational Interview Research

Latham (together with various colleagues), conducted research related to development of what he called the Situational Interview (SI). It was, in part, an outgrowth of his earlier work in the development of Behavior Observation Scales (BOS) for performance appraisal purposes (Latham et al., 1980). Underlying this attempt at development of a valid interview procedure is Locke's (1980) goal setting theory. The premise of this cognitive theory is that conscious ideas affect what people do. Locke reported on a number of experimental studies, which he concluded demonstrate, among other things, that behavioral intentions regulate choice behavior. That is, people tend

to do what they say. This mediating effect of goals or intentions, as enunciated in this finding, would seem to have direct applicability to the interview. Latham and Yukl (1975) reviewed Locke's work along with twenty-seven other studies and concluded, on the basis of both experimental and applied organizational research, that there was strong evidence to support goal setting theory. Intentions (goals) appeared to be directly related to future behavior and further, Latham and Yukl believed these intentions could be determined from the verbal behavior of an interviewee. This led directly to research on a situationally based interview which focused on the stated intentions of the interviewee.

Latham (Latham et al., 1980; Latham, 1989) asserted that an interview approach which focused exclusively on past job-relevant behaviors could discriminate unfairly against those who had not had the opportunity to engage in such behaviors in the past, but who might have the ability to perform well in a given situation. This would be particularly problematic if adverse impact against a minority group was demonstrated. Latham suggested that situational interview questions which focused on the intentions of the candidate, based on that individual's knowledge of their capabilities or past behavior in similar situations, be used instead of behavior based questions.

Latham et al. (1980), examined this proposed situational interview procedure using concurrent and

predictive validity designs. The concurrent validity studies utilized 49 workers and 63 foremen of a sawmill firm as interviewees. Following a procedure developed for performance appraisal purposes, rating scales were developed for performance evaluation, with the same measures serving as the basis for interview questions. This involved a critical incident technique as described by Flanagan (1954), in which employees (it is not clear from the literature who these employees were) generated examples of behaviors on the job which represented either effective or ineffective performance. The incidents were then grouped into common areas and a Behavioral Observation Scale (BOS) developed. The same scale was then used to generate questions for the interview through a rewording process. Possible responses to the questions were generated and used to prepare an objective rating scale. The development of the questions and the behavioral anchors for the rating scales was done by three to five company supervisory personnel. The subjects were interviewed by two or more personnel and asked what they felt they would do if placed in the various hypothetical situations. The responses were then assigned a score by independent judges (twenty company superintendents) based on the previously developed criteria. The same subjects were then rated on their current performance by supervisors in the case of the workers, and superintendents in the case of the foremen, using the BOS previous

developed. Concurrent validities for the workers and foremen were .46 and .30 respectively (both significant, p<.05) and inter-observer reliabilities were .76 and .79 respectively (again both significant, p<.05).

The predictive validity study involved 56 applicants who were applying for work in a pulp mill. Of this number, 30 were female and all were black. All were subsequently hired. The same procedures as previously described were used to develop performance rating scales, interview questions, and response rating scales. Interview responses were examined together with actual performance measures taken 12 months later. Figures for blacks and females respectively were: predictive validity .39 and .33 (both significant, p<.05) and inter-observer reliability .87 and .82 (again significant, p<.05). Overall the results were encouraging, in view of the results of much past interview research. Latham attributed success to the systematic job analysis, the job relatedness of the questions, the careful development of the criterion ratings, and the fact that both measures were based on behaviors.

A second study (Latham and Saari, 1984) was intended to confirm the previous findings as well as to contrast a somewhat different behavioral consistency approach used by Ghiselli (1966), with the situational technique. It involved two separate sub-studies, one a concurrent validity approach, the second involving predictive validity. The

first, using essentially the same procedures as those described in the 1980 report, compared concurrent validity coefficients for questions based directly on those used by Ghiselli (self-reports of past job-relevant behaviors) and situational questions developed specifically for the jobs in question (clerical positions with a major wood products company). The results for the situational questions were consistent with Latham's previous study (concurrent validity approximately .40, significant at $p < .05$), while validities for the past experience questions were only about .15 and not significant. These latter, rather surprising results could not be explained; they suggested however, that the scoring methods could have been considerably different in that specific, related information was not provided by Ghiselli. The second sub-study involved 349 new employees hired to work in a newsprint mill. Only the predictive validity of the situational interview was examined, using the same techniques as previously described for this and the 1980 study. The resulting coefficient (.14), while significant and consistent with the traditional estimated validity of the interview procedure, was nevertheless quite disappointing, particularly considering the results of the 1980 study. After some closer examination, it was discovered that the individuals conducting the interviews had not correctly followed the carefully developed procedures. Although they had asked the agreed-upon

questions, they had not rated each according to the scale, but rather had reacted with an overall impression based on the whole interview.

Only three related studies on the SI are reported in the literature since these original studies. The first (Weekley and Gier, 1987) was a concurrent validity study for a sales position. The criterion measure was not a behaviorally anchored scale, but rather a hard criterion, namely sales productivity. The resulting validity coefficient was .45 ($p < .05$, $N = 24$). Campion, Pursell and Brown (1988) report on a study which, among other measures, examined a structured interview which included simulation questions (similar to SI questions), job knowledge questions, as well as worker characteristic and willingness questions. Predictive validity coefficients using performance scales based on the determined behaviors were reported as .34 ($p < .05$). Neither of these studies constitutes a complete replication of the original SI studies.

The most recent partial replication of an SI study, Robertson, Gratton and Rout (1990), examined the ability of the SI to predict future performance ratings of administrators selected for a company's "fast track". With a sample of 63, the validity of the SI was found to be .28 ($p < .05$).

Overall, the results obtained for the situational interview provide encouraging evidence of an improvement in predictive validity over traditional interview methods. They suggest that if situationally based questions, with behaviorally anchored rating guides, are developed on a thorough job analysis, and compared with job performance rating scales corresponding to the same dimensions derived from the job analysis, then significant improvements in prediction can be realized over more commonly used interview approaches.

Behavioral Interview Research

Janz (1982) also started with the work of Ghiselli (1966) to develop what he termed a Patterned Behavior Description Interview (BDI). This is essentially a structured interview utilizing questions about past behaviors which are developed in a manner very similar to the SI questions. The difference is that they require a self-report of past behaviors (behavioral approach) instead of a verbal report of intentions regarding future behaviors (cognitive approach). The study compared the BDI with an unstructured interview in terms of reliability and predictive validity. Interviewers were undergraduate students. Those conducting the BDI's were enroled in a directed studies course in personnel research; they were given special training and spent additional time developing

the questions and rating scales. The interviewees were teaching assistants and the criteria were student ratings. Unlike Latham's approach, Janz used a comparative rating scale for the interview questions, and as the criterion measure, the traditional departmental five-point rating scale, rather than a scale prepared specifically in agreement with the question areas (however, one might assume there would be considerable overlap). While the interrater reliability for the standard interview was greater than the BDI (.71 vs. .46) the predictive validity of the BDI was higher (.54, p<.01 vs .07,ns). Janz argued that the high agreement in the standard interview was the result of a common but inaccurate stereotype while the lower interrater reliability for the BDI was the result of a more complex process which agreed less often but more often bracketed the criterion. The weakness of this study was in the non-random assignment of interviewers to the two conditions, and the more extensive training and related backgrounds of the BDI Interviewers.

A partial replication of Janz's research was undertaken with applicants for life insurance sales positions (Orphen, 1985). The interviewers were 16 male employees of a large life insurance company. None had any interviewing experience. They were randomly assigned to the control and experimental groups, and both received comparable training in either standard interviewing techniques or BDI (two 3-

hour sessions). The interviewees were 26 male applicants
for sales positions and each was interviewed four times,
twice each by standard interviews and twice by BDI
interviews. Criteria for the predictive validity study
included the total value of life insurance sold in the year
following their selection, and ratings received from the
immediate supervisor (global ratings on a 7-point scale).
BDI questions were developed using a critical incident
technique similar to that used by Janz (1982). Reliability
was found to be .72 for the BDI and .68 for the standard
interview, which were not significantly different. The
overall validity for the BDI was .48 using supervisors
ratings as the criteria and .61 using value of sales as the
criteria. The overall validity of the standard interview
was .08 and .05 respectively for the two criteria. In both
cases the BDI validities were significantly higher (p<.05)
than the standard interview. There have not been any other
replications of research on the BDI reported in the
literature.

Like the SI, the BDI questions are based on performance
dimensions derived from a detailed job analysis. Unlike the
SI, behavioral anchors for the interview rating scales are
not provided, but rather responses are assessed on a
comparative rating scale (Janz, Hellervik, and Gilmore,
1986). Also, in the BDI studies reviewed, the performance
measures were either the standard performance evaluation

ratings, global supervisor ratings, or the dollar value of sales. Nevertheless, the demonstrated improvement in predictive validity was significant and indicates promise for this type of approach, focusing on the past behavior of the applicant.

Hypotheses

Both the SI and the BDI are structured and adhere to the behavior consistency principle. Both are developed utilizing approaches found by previous researchers and reviewers to be promising. These include carefully developed predictor measures which are in agreement with performance measures (complete agreement in the case of the SI), training of raters and interviewers, and a highly structured interview. The initial results of the SI have not been confirmed owing to methodological differences, and only two studies of the BDI are reported. As well, the two approaches have not been compared to one another (and to a standard interview) in the same setting. The purpose of this study was to examine the relative validity of the two approaches and contrast them to the current and somewhat more traditional approach used by the Canadian Forces to predict the likely success of non-officer applicants.

There were four research hypotheses:

1. The SI would demonstrate significant predictive validity.

2. The BDI would demonstrate significant predictive validity.

3. Both structured interview approaches would show a significant improvement in prediction over the traditional rating method.

4. Of the two structured approaches, the SI would show significantly greater predictive validity than the BDI.

The first three hypotheses follow logically from most of the literature which has previously been cited. The fourth follows from a consideration of the nature of the applicants for the CF. In large part they are young (17-24 is the primary target group) and they have relatively little work experience. The BDI studies focus on an examination of past behavior in directly related job abilities and the interviewees all had related work experience. This factor would seem to pose some potential difficulty for the unskilled applicant. The SI on the other hand, in that it focuses on behavioral intentions, does not appear to have this drawback. In fact, Janz himself, in comparing the BDI to the SI (Janz, 1989) felt the SI had the advantage of not requiring past experience. Wiesner (1991) also suggests the BDI is better suited to candidates with prior work experience.

Method

Participants

The interviewers for the study were 12 Military Career
Counsellors (MCCs) employed in five different Canadian
Forces Recruiting Centres (CFRCs) in the Atlantic region.
All MCCs had at least one year experience in the job, and
had completed a training program which focused on
interviewing, decision making and report writing. The
interviewees were 53 Canadian Forces (CF) non-officer
applicants who were about to be enroled and then proceed to
the Canadian Forces Recruit School (CFRS) to undergo initial
level recruit training.

The applicants had all been interviewed previously as a
normal part to their processing, had been rated on their
potential to succeed on initial level military training and
had been selected as suitable for available quota. The
prior assessments, based on the traditional interview, were
completed from several weeks to a year previous to the
conduct of this study. The applicants had been placed on
merit lists and were assigned positions as quota became
available.

The sample of applicants was a sample of opportunity,
and represented slightly less than 10% of the applicants
enroled at the designated units over the approximate nine
month period of the study. As interviewers, enrolees, and

time were available on enrolment days, one or more of the experimental interview procedures was administered (thus random assignement of applicants to interviewers was not possible). All 53 of the applicants were interviewed using the BDI; of those, 20 were also interviewed using the SI. A second independent BDI rating on 15 of the interviewees was also obtained when a second interviewer was available. The second rating was made independently by the second MCC who listened to the applicant's responses. Workload made it impossible to obtain a larger sample of independent ratings.

The smaller sample of applicants administered the SI was in part the result of data collection problems early in the study. Some of the criterion data sample for applicants who had been administered one or both of the experimental interviews (approximately 15) was lost during the summer period when staff of the Recruit School was changing, and instructions for data collection were not effectively communicated to incoming staff. At this early stage it was noted that the SI questions seemed somewhat easy and transparent to most interviewees, and this raised questions as to their potential value. As both time constraints and limited sample possibilities were present, every effort was made to maximize the sample for the BDI, while at the same time obtaining an adequate sample for the SI, keeping in mind the expected effect size, approximately $r=.40$ (Wiesner, 1991).

Procedure

In keeping with the procedures initially described by
Latham et al. (1980) and Janz (1982), a job analysis
utilizing a critical incident technique was undertaken to
determine the primary behavioral dimensions contributing to
success in initial level military training. These were used
as a basis upon which to develop both structured questions
and criterion rating scales. The recruit training course
was chosen, in that it has as its aim to train and test new
recruits in those areas of performance deemed essential to
overall adaptation to the military environment.

Six staff members of the recruit school (ranging in
rank from Major to Master Corporal), provided assistance in
this activity. The standard assessment procedure was
reviewed and discussed and then five of the staff members
were requested to generate approximately six critical
incidents each, which illustrated effective behaviors on the
part of one or more of their best recruits. They were also
requested to generate a similar number of representative
ineffective behaviors on one or more of their worst
recruits. It should be noted that the number of critical
incidents is traditionally considerably larger than used in
this study, but the instructors were very familiar with the
full range of behaviors of the recruits, and were also
familiar with the traditional rating dimensions used by the

recruit school. As a result the group and the researcher were satisfied that the domain was adequately sampled. The incidents were grouped for similarity and discussed in order to label the basic dimensions. Six possible behavior dimensions were identified by the staff:

1. Adheres to formal and informal rules and norms.

2. Accepts direction, criticism, and censure.

3. Interacts positively with peer groups.

4. Actively participates in co-operative tasks and activities.

5. Reacts appropriately and effectively under mental and/or physical stress.

6. Adapts quickly and effectively to new or novel situations.

Further discussion with the Chief Instructor and an in-depth examination of the identified dimensions, led to the conclusion that there was considerable overlap of these dimensions. Three more generally stated dimensions were considered to adequately describe the range of identified behaviors:

1. Conduct (follows rules, takes direction, and accepts criticism)

2. Teamwork (gets along with peers, and works well as a team member)

3. Coping (reacts well under stress, and adapts to new situtions)

Following the procedure described by Latham et al. (1980), the instructors were requested to develop situationally based questions corresponding to the six dimensions. The six scales were maintained for the criterion ratings as the instructors were comfortable with them and the wording was not unlike several of scales used on the traditional Recruit School student rating form. The six rating scales were later statistically combined (averaged) into the three more general scales. A copy of the criterion rating form is contained at Appendix A.

Further, as a group the staff provided suggested behavioral anchors for the SI questions (questions and behavioral anchors are contained at Appendix B). In the case of the BDI, a number of questions were developed by the author following the procedures and examples provided by Janz (1986) (questions are contained at Appendix C). Rating forms for the questions were based on the three general behavioral dimensions, with two or more questions designated for each scale. A copy of the form is at Appendix D.

Seven point rating scales were used for both the questions and performance measures. This differs from the five point scales used by both Latham and Janz, but was chosen to more closely correspond to the Military Potential (MP) scale, which is the traditional predictor used by MCCs to rate CF applicants.

Prior to beginning any data collection, visits were

made by the author to all of the CFRCs participating in the
study, in order to familiarize and train interviewers prior
to administration of the experimental protocols. The
training consisted of having the MCCs identify potentially
important behavioral dimensions they felt would contribute
to success in recruit training. In a sense, the MCCs
themselves were subject matter experts in that they had all
been through the same or similar basic training course, and,
naturally, are current CF members. There was a strong, and
not unexpected overlap in the dimensions identifi..d, with
those identified by CFRS staff. There was a tendency for
the MCCs to identify a slightly larger number of dimensions,
some of which had been discussed but rejected by the CFRS
instructors. This activity produced a better understanding
of the dimensions to be assessed, and appeared to elicit
support on the part of the MCCs for the rating scales.

In the case of the SI, specific targeted training was
limited to reviewing the interview procedure and ensuring an
understanding of the rating process. In the case of the
BDI, additional discussion covered use of supplementary
probing questions to better understand responses, and
discussion and role play of potential difficulties as
described by Janz (1986). These included:

1. In response to a BDI question, and particularly an
   initial one, some interviewees experience
   difficulties and discomfort in thinking of a

response. Techniques including putting the applicant at ease, use of silence, and patience in eliciting a response were covered to assist in this area.

2.   Some interviewees will initially claim that a past situation as requested by the BDI has never happened to them. Again, techniques to deal with this problem were covered including reassuring the applicant that similar things happen to all of us, and again the development of an encouraging and patient approach.

3.   The third typical response is for the interviewee to respond with a generality rather than a specific example of an incident from the past. Here the emphasis was on having the interviewer continue to insist on a specific response until the expectation is created.

MCCs were instructed not to interview an applicant using an experimental measure, whom they had previously assessed and rated during the normal processing phase (as was noted the original traditional assessment was made at least several weeks earlier). Further they were instructed not to look at the applicant's previous assessment report and rating, at least not until after they had made their ratings on the experimental measures.

On completion of the structured interviews, interview

rating forms were completed by the MCCs and the forms, together with the MP ratings from the traditional interview were gathered and held until the applicants were enroled in the CF, and had begun recruit training at CFRS. CFRS administrative staff were then notified and requested to have training staff, who had observed the candidates under training (for at least six weeks in the case of any early terminations for training or other reasons, and on completion of the ten week course for all others), to complete the performance assessment form.

Some discussion of the traditional Military Potential rating, given on completion of the current interview procedure, is also in order. While it is directly connected to the interview, it should be noted, that unlike the experimental methods discussed here, it is in fact a summary rating not made on the basis of information gained uniquely in the interview. Prior to the interview, considerable additional information is available to the interviewer, specifically general ability test results, results of an aptitude test battery, language aptitude test scores, educational transcripts, work history, and other biodata from various application forms and questionnaires. This information will, and is intended to, influence the rating. It should also be noted that the MP ratings is intended as an overall predictor of early military training success, which consists not only of recruit school and early overall

adaptation, but also initial level occupational training.
In most cases this initial level occupational training
requires certain specific aptitudes and educational
achievements (quite high for technical occupations), which
may be irrelevant to (or even negatively correlated with)
performance at recruit training.

Results

As is common practice for traditional interview methods, and for the original research carried out by Latham and Janz (Latham, 1980; Janz, 1982), individual interview and performance rating scales were combined into an overall predictor and criterion score (mean of the individual scales). In addition, where there were two independent ratings for the BDI, they were averaged to produce one single score for each applicant (a difference of less than .01 in the validity coefficient resulted if all 68 ratings were used). As was noted earlier, the MP rating is a global rating based in part on the interview, and also to some degree, on other information available to the interviewer at the time of the interview.

An examination of the intercorrelations of the individual BDI predictors and individual performance scales, and the same intercorrelations for the SI (Tables 1 and 2 respectively) suggested that there were indeed underlying dimensions related to the total predictor and total performance scores. Therefore, principal components analyses of the intercorrelations of the individual scales of the SI and BDI, together with their respective individual performance scales, were conducted. In the first analysis, the resulting factor loadings for the BDI ratings clearly showed two factors, the first corresponding to the combined

Table 1

Intercorrelations of Individual BDI Scales and Performance

Scales

Variable	1	2	3	4	5	6
1. Teamwork Performance	-					
2. Conduct Performance	.72**	-				
3. Coping Performance	.70**	.73**	-			
4. Teamwork Predictor	.35*	.18	.27	-		
5. Conduct Predictor	.38**	.29*	.28*	.54**	-	
6. Coping Predictor	.47**	.37**	.45**	.60**	.72**	-

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

predictor scores, the second corresponding to the combined performance scores (Table 3). Two factors also emerged in the second analysis of the SI ratings. Again the individual performance scores loaded heavily on one factor. The second factor was related to the overall performance rating, although the loading for the coping scale was in a direction opposite to the teamwork and conduct ratings (Table 4). Certainly, the relatively smaller sample for the SI may have contributed to these results, and therefore this finding was considered an anomaly. As a result, the overall performance measure was retained for the SI.

Table 2

Intercorrelations of Individual SI Scales and Performance

Scales

Variable	1	2	3	4	5	6
1. Teamwork Performance	-					
2. Conduct Performance	.76**	-				
3. Coping Performance	.74**	.88**	-			
4. Teamwork Predictor	.45*	.24	.33	-		
5. Conduct Predictor	.11	.14	-.04	.20	-	
6. Coping Predictor	-.16	-.15	- 31	-.65**	-.23	-

* p<.05
**p<.01

Neither the overall traditional interview rating (MP), nor the overall SI score, correlated significantly with the overall performance score. However, the overall BDI rating was significantly related to the overall performance score (Table 5). The resulting validity coefficient was in keeping with the results of the earlier reviewed research on the BDI (r=.44, p<.01). The 95% confidence interval for the validity coefficient of the BDI did not include zero, while those for the MP and SI both included zero. It should be noted that a sample size of approximately 125 would have

been required for the observed validity coefficient of the SI to reach significance at p<.05 (r=.17).

Table 3

Principal Components Analysis of Individual BDI and Performance Scales. Varimax Rotation (N=53)

| | Factors | |
Scale	1	2
BDI Teamwork	.10	.83
BDI Conduct	.17	.85
BDI Coping	.31	.84
Performance Teamwork	.84	.30
Performance Conduct	.91	.11
Performance Coping	.88	.19

To further examine the nature of the potential improvement in prediction offered by the BDI and/or the SI over the traditional interview, as well as the difference between the BDI and SI, z-tests were performed using Fisher's r-to-z transformation (Ferguson, 1981). The BDI prediction was substantially and significantly higher than the prediction based on the traditional MP rating (z=1.65, P<.05, one-tailed). There was no significant difference between the SI and the MP validities (z=.11, ns), nor was there a significant difference between the validity of the SI and BDI (z=1.07, ns). The means and standard deviations of the MP, SI, and BDI ratings are contained at Table 6.

Table 4

Principal Components Analysis of Individual SI and

Performance Scales, Varimax Rotation (N=20)

Scale	Factors 1	2
SI Teamwork	.30	.81
SI Conduct	-.04	.54
SI Coping	-.13	-.87
Performance Teamwork	.88	.17
Performance Conduct	.94	.05
Performance Coping	.94	.12

In addition, a hierarchical multiple regression analysis was conducted to examine the improvement in prediction which could be realized by using the BDI in combination with current assessment practices. The analysis showed a substantial and significant improvement in the amount of variance in the criterion accounted for, when the contribution of the BDI was considered (Table 7).

Only with the BDI was it possible to obtain a sample of independent ratings in order to provide an estimate of interrater reliability. Interrater reliability for the BDI was found to be .69 (p<.01, N=30).

Table 5

<u>Validity Coefficients and 95% Confidence Intervals of r for</u>

<u>MP, SI, and BDI Scores with Performance Scores</u>

	Interview Type		
	MP	SI	BDI
Validity Coefficient	.14(53)	.17(20)	.44*(53)
95% CI of r	-.13 to .39	-30 to .57	.26 to .62

<u>Note 1</u>. Sample size shown in parentheses

<u>Note 2</u>. Validity Coefficients are uncorrected

*p<.01

Statistics from a data base held by the Recruiting Zone
Headquarters (Atlantic) were available on a large sample of
files (N=1807), collected over the same approximate perioa
of time as the experimental sample, thus providing
information (SD of the MP rating) on an unrestricted sample
of CF applicants. In that there was a substantial and
significant relationship between the MP rating and the BDI
rating (.28, p<.05), an examination for possible indirect
restriction of range (due to selection on the correlated
traditional measure) was undertaken (Angus, 1985; Gilford,

1965). No substantial increase in r due to range restriction was discovered for either the SI or BDI (increase in r<.01).

Table 6

Comparison of Means, and Standard Deviations for MP, SI, and BDI

Measure	N	Mean	SD
MP	53	3.8	1.10
SI	20	5.8	.45
BDI	53	4.3	.80

Table 7

Hierarchical Multiple Regression Analysis Results of MP and BDI with Performance Score

Predictor(s) in Equation	R	R^2	R^2 Change
MP	.14	.02	.02
MP&BDI	.44*	.19	.17*

*p<.01

Discussion

Of the four research hypotheses, the study provided
support for only one, namely that the BDI would show
significant predictive validity. It also provided partial
support for another, namely that the BDI would demonstrate a
significant improvement in prediction over the traditional
rating method. Contrary to what had been predicted, the BDI
produced greater predictive validity than the SI, although
the difference was not significant.

Behavior Description Interview

The results of the study with respect to the BDI were
consistent with past research on this approach (Janz, 1982;
Orphen, 1985). The mean r (criterion-related validity
coefficient) across these former studies was .55 (Wiesner,
1991) as compared to .44 for the current study. In
addition, substantial and significant improvement over the
traditional interview was shown. The results therefore,
clearly support the value of a highly structured interview
process, developed on a detailed job analysis and using as a
criterion-measure, the same behavioral dimensions as the
predictor. It also demonstrates the behavior consistency
principle which is expressed in the familiar wisdom that the
best predictor of future performance is past performance.

As was noted in the discussion of the research hypotheses, it was anticipated that the BDI might be expected to produce weaker results than the SI because of the relative lack of work experience of the interviewees for this study. However, this did not appear to adversely effect the predictive ability of the BDI questions. Two possible factors which may explain this finding should be considered:

1. The Canadian Forces, unlike most other employers, does not hire skilled personnel as non-officer recruits. In large part the abilities required to succeed, particularly in initial level general military training, are not related to highly developed specific job knowledge or skills, but are broad and general in nature, and are not necessarily developed only through previous job training or exposure.

2. The BDI questions for this study, in recognition of the nature of the anticipated applicant background, made specific provision for responses which could be based on experiences in school, sports, clubs or other social activities, and family settings, as well as in work experience.

Overall, this study provides support for the BDI as a behaviorally consistent, structured interview approach which can provide significant predictive validity for initial

level training success. It has shown in the past to be suitable for experienced applicants (Janz, 1982; Orpen, 1985) and this study confirms that it is also effective for inexperienced individuals.

Situational Interview

The results of the study with respect to the SI were somewhat disappointing. Several possible factors may have had an effect on the results, and suggest it would be inappropriate to conclude the SI has no value as a potential predictor of success for non-officer CF applicants. Already mentioned was the small sample size, which resulted from initial data collection problems and the decision to maximize the sample for the BDI, when the tendency for the SI questions to be somewhat transparent and easy was noted. Latham (1989) noted this potential problem of overly easy questions, and suggested avoiding lower rated anchors which represented responses so ridiculous that no one would respond in such a fashion. He further noted that success lies in having the questions so abstruse that applicants cannot determine the desired response and therefore must report their true intentions. While this would be possible if behaviors known only to skilled (or well prepared) applicants were used (job knowledge and job skill questions), this did not seem to be possible for the general behaviors under consideration. They represented fairly well

known dimensions connected with life in the military, and therefore "faking good" was a likely (and apparently observed) outcome.

This finding is contrary to the original expectation that the SI would be better for an inexperienced target population, but was noted as a possible outcome by Janz (1989) who pointed out that applicants with the ability to reason and express themselves well, could possibly respond with quite satisfactory answers to SI questions, yet their stated intentions might depart from their future behavior on the job.

Military Potential Rating

Notwithstanding the validity coefficient for the MP rating was approximately that which might be anticipated from past research on the traditional interview, the nature of its composition, and its intended use, as previously discussed, do not make its mediocre performance surprising. In fact, the interview itself is based on a semi-structured guide, in which most of the factors have been rationalized as bearing a relationship to dimensions known to be important to success at recruit training. If the contribution of this information could be factored out, it might well bear a more substantial and significant relationship to the general factors identified in this study.

Restriction of Range

It was noted earlier, that there was no improvement in the validity coefficient for either the SI or BDI when they were corrected for indirect restriction of range. Although the SD of the MP rating (the correlated measure upon which selection was made) was slightly smaller in the study group, the difference was not sufficiently large to produce a substantial correction for either of the experimental measures. This was due to a rather unusual situation which had occurred in the available occupational quota over the period of the study. Although the number of applicants had seen a rather dramatic increase during the early stages of the study (perhaps due to publicity surrounding the Gulf War), the number and type of occupations open were drastically reduced, and only rather less attractive choices were available (Combat Arms and Navy). The number of applicants for these particular occupations was insufficient to fill all the available quota, and therefore, even the least suitable applicants were selected.

Utility

One factor which is naturally of concern to any employer is the possible dollar benefit or utility to be realized by the introduction of any selection system. This is certainly true of the CF, and indeed it was a primary

consideration in the study. For this reason, an estimate of utility of the BDI was derived using procedures and tables contained in Janz (1986) which in turn were developed on a system described in Schmidt, Hunter, McKenzie and Muldrow (1979). Rather than being an actuarial method for determining utility, this method assumes a linear relationship between the job performance of an individual and the dollar value of that individual to the organization, as well as a linear relationship between performance on a selection predictor and later performance on the job. Taking into account the improvement in validity realized by implementing a new selection system, the overall selection ratio (number of openings/number of applicants), and estimating the standard deviation of performance in dollars, a dollar benefit can be derived. In the case of the BDI, implementation of such a measure would result in an improvement in validity of approximately .30. Using a conservative estimate of the SD of performance in dollars (40% of annual salary; Schmidt et al., 1979), a selection ratio of .50 (an accurate reflection of the selection ratio for the Canadian Forces at present), and a projected hire of approximately 2,000 non-officer personnel over the next 12 month period, the potential dollar benefit would be over $8 million for a three year period (the standard initial engagement period for the Canadian Forces). Considering that the costs of fully developing and implementing such a

system would be relatively inexpensive for the CF, the overall net benefit by improving the prediction of the interview process could be substantial. Smaller organizations with fewer hires, and lack of support from a personnel research organization, could well encounter relatively high startup costs, and a much smaller dollar benefit. It is less clear in such cases, whether or not overall benefits would result.

Human Rights

Human Rights legislation in the U.S., and more recently in Canada, has and will focus greater attention on issues of fairness and job relevance in relation to selection practices, including the employment interview (Arvey,1979; Campion and Arvey, 1989; Cronshaw 1989; Gatewood and Field,1987). Fortunately, achievement of validity and utility are significantly related to fairness and job relevance.

In the U.S., Title VII of the 1964 Civil Rights Act set out specific guidelines regulating discrimination in the workplace, and established an enforcement agency, the Equal Employment Opportunity Commission (EEOC). Original legislation and subsequent revisions established a prohibition on discrimination on such grounds as sex, race, color, religion, national origin, age (between 40 & 70), physical and mental handicaps, and disabled veterans.

Discrimination is defined as intentional prejudice (negative treatment of a group because of personal characteristics), unequal treatment (situations where different standards are set for different groups), and adverse impact (different proportions of certain groups are selected even where the same standard is applied; the most common measure is known as the four-fifths rule, ie the ratio of any group must be at least 80% of the ratio of the most favourably treated group). Three possible defences may be employed in the event adverse impact is proven:

1. <u>Business Necessity</u>. This defence requires the employer to demonstrate that a strong relationship between the selection device and performance exists and that training and failure costs would be prohibitive if the method were discontinued.

2. <u>Bona Fide Occupational Qualification (BFOQ)</u>. This is difficult to use, is primarily seen in sex discrimination cases, and means that no person of a particular group (ie sex, age, religion, etc.) can perform the job effectively (eg. restroom attendant).

3. <u>Validation</u>. Normally, this will include evidence of statistical (predictive) or content validity. It is interesting to note that in 1987 the American Psychological Association (APA) filed an amicus brief with the U.S. Supreme Court arguing

that validation requirements applying to selection
tests be applied as well to subjective measures
such as the interview (Field and Gatewood, 1989).
This would likely place a higher value on
scientific evidence (statistical validity) to
defend the employment interview.

Canadian legislation, while having similar aims, is
somewhat different than the U.S. legislation, and the number
of precedent setting cases are fewer. The Canadian Human
Rights Act (CHRA) provides for 10 specific grounds on which
discrimination is forbidden, namely race, national or ethnic
origin, color, religion, age, sex, marital status, family
status, pardoned conviction or disability. In addition, the
Charter of Rights and Freedom provides more generally for
all Canadians to enjoy equality and fair treatment in a
variety of areas including hiring. It is not necessary to
show adverse impact (although this may be an issue) as it is
in the U.S., only that as an individual one has been
unfairly treated (particularly on the proscribed grounds).
In such cases, the burden of proof shifts to the employer,
who must defend a Bona Fide Occupational Requirement (BFOR).
Cronshaw (1989), in examining cases of challenges against
the employment interview in Canada, points out that evidence
of job relatedness may be used to satisfy tribunals that a
BFOR exists. While professional validation may be a
sufficient condition, at the present time it has not proven

to be a necessary one. He also notes that the cases he has examined have been the result of Human Rights Commission Tribunals, and that none has yet been heard by an appeal or the Supreme Court.

Cronshaw (1989) suggests that a well designed structured interview, based on a job analysis, will likely have an advantage over unstructured interviews for two reasons:

1. The structure of the interview will decrease the chances that the interviewer will stray into prohibited areas of questioning.

2. The structured interview should, if properly designed, be able to demonstrate job relatedness in that it is based on questions developed from a job analysis.

Certainly, if the employment interview is to continue as a viable and widely utilized selection instrument, and is to prove capable of withstanding Human Rights challenges it should be carefully examined to ensure that it:

1. Demonstrates validity (particulary if adverse impact is observed).

2. Is directly and visibly related to identified job requirements.

3. Is recorded in such a manner as to permit independent review.

4. Adheres to common standards and criteria.

With these conditions in mind, it would seem that the BDI offers considerable potential to provide an ideal selection instrument which can withstand the scrutiny of human rights challenges. Initial results with non-officer CF applicants show significant predictive validity, the questions asked are directly related to dimensions of future success in training, interview responses can be recorded to permit independent review, and the fact that the procedure is structured means that a set of common questions and standards will be applied.

Future research

Additional research on the SI and BDI needs to be done to examine the specific applicant and criterion characteristics which may have a moderating effect on the results. Wiesner (1991) has in fact proposed, and has begun a study in conjunction with Latham, to examine both the BDI and SI as potential predictors for selecting Naval Officers for the Canadian Forces. In addition to replicating the work of this study, Wiesner intends to incorporate the use of an anchored scoring guide for the BDI, although initial work on the BDI did not do so. As well, it can be anticipated that the nature of the behavioral dimensions important to officer training might more likely be found in work experience, and may be more appropriate for SI questions (i.e. leadership behaviors). The outcome of this

research should provide valuable information on the performance of the two structured interview measures.

An additional area which requires attention, is a careful examination of the criterion measures for success in early military training. Traditionally, as in the case of most organization, the Canadian Forces uses a single composite predictor for early success, yet, as was noted earlier, many possible dimensions contribute to success, and much information is gathered and incorporated into the composite predictor. There is no evidence, however, that the criterion measures (which again are composites) represent the same construct. In fact, it is highly likely that they do not, and that little or no relationship (or even a negative relationship) might exist among them. Thus, although structured interview procedures seem to show promise, particularly if matching criterion measures are tailored for their use, the nature of currently used criterion measures for both initial general military training and initial level occupational training need to be critically examined and decomposed into their elements, before truly effective and specifically designed predictors can achieve their maximum potential.

References

Angus, R.J. (1985). <u>Restriction of range in predictive validity studies: General correction procedures.</u> (Technical Note 1/85). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Arvey, R.D. (1979). Unfair discrimination in the employment interview: Legal and Psychclogical aspects. <u>Psychological Bulletin</u>, <u>86</u>, 736-765.

Arvey, R.D., & Campion, J.E. (1982). The employment interview: A summary and review of recent research. <u>Personnel Psychology</u>, <u>35</u>, 281-322.

Bradley, J.P. (1990). <u>A validation study on the Naval Officer Assessment Board's ability to predict MARS officer training success.</u> (Working Paper 90-7). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Campion, M.A., Pursell, E.D., & Brown, B.K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. <u>Personnel Psychology</u>, <u>41</u>, 25-42.

Cronshaw, S.F. (1989, June). <u>Legal implications for the employment interview in Canada</u>. Paper presented at the 50th Annual Convention of the Canadian Psychological Association, Halifax, N.S.

Ferguson, C.A. (1981). Statistical analysis in psychology and education. New York: McGraw-Hill Book Company.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Phillippines: Addison-Wesley.

Gatewood, R.D., & Field, H.S. (1990). Human Resource Selection. Orlando, FL: The Dryden Press.

Ghiselli, E.E. (1966). The validity of the personnel interview. Personnel psychology, 19, 389-394.

Ghiselli, E.E., Campbell, J.P., & Zedek, S. (1981). Measurement theory for the behavioral sciences. San Francisco: W.H. Freeman and Company.

Gilford, J.P. (1965). Fundamental statistics in psychology and education. New York: McGraw-Hill Book Company.

Harris, M.M. (9189). Reconsidering the employment interview: A review of recent literature and suggestions for future research. Personnel Psychology. 42, 691-726.

Hunter, J.E., & Hunter, R.F, (1984). Validity and utility of alternate predictors of job performance. Psychological Bulletin, 96(1), 72-98.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982).
Meta-analysis: Cumulating research findings
across studies. Beverly Hills, CA.: Sage
Publications.

Janz, T. (1982). Initial comparisons of patterened
behavior description interviews vs. unstructured
interviews. Journal of Applied Psychology, 67,
577-580.

Janz, T., Hellervik, L., & Gilmore, D.C. (1986).
Behavior description interviewing. Newton, Mass:
Allyn and Bacon Inc.

Latham, G.P. (1989). The reliability, validity, and
practicality of the situational interview. In
Eder, R.W. and Ferris, G.R. (Eds.), The employment
inter ew: Theory, research, and practice.
Newbury Park, California: Sage Publications Inc.

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

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

Latham, G.P. & Yukl, G.A. (1975). A review of research on the application of goal setting in organizations. Academy of Management Journal, 18, 824-845.

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

Martin, M.A. (1972). Reliability and Validity of Canadian Forces Selection Interview Procedures (Report 72-4). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Mayfield, E.C. (1980). The selection interview: A reevalation of published research. Personnel Psychology, 17, 239-260.

Nunnally, J.C. (1978). Psychometric theory. New York: McGraw-Hill Book Co.

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

Okros, A.C., Johnston, W.V., & Rodgers, M.N. (1988). An evaluation of the Naval Officer Selection Board as a predictor of success on the Basic Officer Training Course. (Working Paper 88-1). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

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

Robertson, I.T., Gratton, L. & Rout, U. (1990). The validity of situational interviews for administrative jobs. Journal of Organizational Behavior, 11, 69-76.

Schmidt, F.L. (1976). Social and situational determinants of intervew decisions: Implications for the employment interview. Personnel Psychology, 29, 79-101.

Schmidt, F.L., & Hunter, J.E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. Journal of Applied Psychology, 68, 407-414.

Schmidt, F.L., Hunter, J.E., McKenzie, R.C., & Muldrow, T.W. (1979). Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Schmidt, F.L., Hunter, J.E., & Pearlman, K. (1982). Assessing economic impact of personnel programmes on work-force productivity. Personnel Psychology, 35, 333-347.

Schmidt, N., & Ostroff, C. (1986). Operationalizing
the behavioral consistency approach: Selection
test development based on a content oriented
strategy. Personnel Psychology, 39(1), 91-108.

Ulrich, L., and Trumbo, D. (1965). The selection
interview since 1949. Personnel Bulletin, 63,
100-116.

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

Webster, E.C. (1964). Decision making in the
employment interview. Montreal: Industrial
Relations Center, McGill University.

Webster, E.C. (1982). The employment interview: A
social judgement process. Schomberg Ontario: SIP
Publications.

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

Wernimont, P.F., & Campbell, J.P. (1969). Signs,
samples, and criteria. Journal of Applied
Psychology, 52, 372-376.

Wiesner, W.H., (1991). Developing and evaluating
improved interview procedures for the naval
officer assessment board. (Technical Note 3/91).
Willowdale, Ontario: Canadian Forces Personnel
Applied Research Unit.

Wiesner, W.H., and 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.

Wright, O. R. (1969). A summary of research on the
    selection interview since 1964. Personnel
    Psychology, 22, 391-413.

Appendix A

PROTECTED B

RECRUIT ASSESSMENT FORM

1. Recruit Information: _____ _____ _____ _____
   SIN   NAME & INIT   MOC   CFRS
                                              SERIAL

2. CFRC/CFRC Det: _____

3. Critical Behavior Areas (circle appropriate assessment):

   a. Follows formal/informal rules and norms.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

   b. Accepts direction and criticism.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

   c. Gets along well with peers.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

   d. Works well as a team member.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

   e. Reacts effectively and appropriately to stress.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

   f. Adapts well to new and different situations.

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

Appendix B

Situational Interview Questions

Conduct

Follows Directions/Rules

1. Imagine a situation at work where your boss assigns you a job that you do not at all enjoy doing, and which requires you to work beyond normal working hours. You think it could wait until the next day. You have made other plans for that time to go out with friends, something you have been looking forward to doing for sometime. In fact the plans you have made will be difficult if not impossible to change. What would you do? What would you say to your boss?

Behavioral Anchors:

High	Accept the assignment and do your best to change plans and explain to friends.
Average	Explain to your boss the importance of the plans and ask if you can do it later or if someone else can fill in for you. Accept task if that is outcome.
Low	Tell boss you won't do it, and that you have other plans which can't be changed.

Accepts Criticism

2. Imagine that you have just completed a very difficult and long task that you have been assigned. You have put a lot of effort in the task and feel you have done your best. Your boss is not all that impressed. In fact, after looking at the results he becomes very angry and criticizes you quite severely. What would you do? What would you say to your boss?

Behavioral Anchors:

High	Accept the criticism and ask for direction as to how you could have completed the task better.
Average	Accept the criticism, but tell your boss you did your best and feel that his reaction might not be altogether deserved.
Low	Get angry and tell your boss you think you did well and that the criticism is wrong and not deserved.

Teamwork

Interacts with Group

3. Your group has decided to hold a car wash to raise money. You show up to help, but feel the way things are organized and how things are being done are wrong. You try to change some things but no one will listen. What would

you do?

Behavioral Anchors:

> High Join in and do your part. Don't complain,
> but when you can, make suggestions you feel
> would be helpful.
>
> Average Join in, but continue to try to have things
> done right, even if nobody still listens.
>
> Low Tell them that if they won't listen, then you
> won't help.

Cooperation

4. Imagine a situation in which you are working with a group of other people. You have been assigned a specific individual task to do, which you are having some difficulty completing because you do not have all the required skills and knowledge. As well, the group as a whole has been given a general task to perform. Both tasks must be completed by the day. Most everyone else seems to have their own work done and are working on the group task. You will be held accountable for completion of both tasks. What would you do?

Behavioral Anchors:

> High Pitch in and do your best to see that the
> group task gets done. Then see if anyone
> with the necessary skills or knowledge could
> help you complete your task.

Average   Tell the group you will be there to help
soon, that you are having difficulty with
your task and that you need a little more
time to try to get it done. Help as soon as
you can get your job done.

Low   Tell the group they will have to get along
without you because you won't be able to get
your task done if they don't leave you alone.

Coping

Performs Under Stress

5. You have had a hard day at work. You are physically
exhausted and pretty frustrated because it seems nothing was
going right. Your boss arrives on the scene and becomes
angry over something you don't think is very important.
What would you do?

High   Relax, accept the criticism and ask for
advice about what your boss feels is wrong.

Average   Accept the situation but tell your boss you
have had a hard day and that his outburst is
not helpful.

Low   Get mad and tell the boss he has no right to
criticize you. You have had a bad day and
don't need this.

Adapts to New Situations

6. Imagine that you have been working at the same job for several years. You like the work and the people you work with. The place you live in is nice and you have a lot of good friends. You are transferred by your company to a new location quite far away. When you get there you find the work very hard and not to your liking. The new work group does not accept you and you have not made any new friends. The place you live is not nearly as nice as your former place. What would you do?

Behavioral Anchors:

High	Accept the situation knowing its always hard at first to fit into a new group and that it takes time to make new friends and adjust to a new location.
Average	Give it a try, but if things don't change and become like they used to be, you might consider returning to your former location and looking for another job, if you have to.
Low	Quit and go back, even if it means losing your job.

Appendix C

Behavior Description Interview Questions

Conduct

1. In any organization (work, school, sports, and clubs,
for example), there are many rules, regulations and policies
which are supposed to be understood and followed. Tell me
about a time when, even though you may not have agreed with
those rules, you carried them out.

Possible Probes:

What were the circumstances of the case?

What was the outcome of the situation?

What feedback did you receive, if any?

How did you feel about the outcome?

(Option) Tell me about another one.

2. Now I would like you to tell me about a time when you
didn't adhere to the rules, regulations etc., and the
results were a bit sticky.

Possible Probes:

What led to the situation?

What reason did you have for handling the situation the
way you did?

What was the outcome?

How would you handle this situation if it came up
again?

How often has this type of situation arisen in the past
year?

(Option) Tell me about another one.

3. Sometimes we all are given direction to do something we
do not want to do or with which we do not agree. Tell me
about a time when a parent, teacher, boss , etc., gave you
such direction.
Possible Probes:

What were the circumstances?

How did you handle the situation and why?

How would you handle this type of situation if it came
up again?

(Option) Tell me about another one.

4. Sometimes, even when we think we have done our best, we
are criticized for something we have done. Tell me about a
time when you have been criticized or got into trouble for
something you did at home, school, or work.
Possible Probes:

What were the circumstances?

How did you feel?

How did you handle the situation?

How would you handle it another time?

(Option) Tell me about another one.

Teamwork

5. It is often necessary to work together in a group to accomplish a task. Can you tell me about a recent experience you had working as part of a group?

Possible Probes:

What was the task?

How many people were in the group?

What difficulties arose as a result of working as a group?

What role did you play in resolving these difficulties?

How successful was the group?

How often do you work as part of a group?

(Option) Tell me about another one.

6. Even people who are reasonable can have disagreements. Tell me about the most heated disagreement you experienced with a friend, family member, peer or fellow worker.

Possible Probes:

When did this take place?

Who was involved?

How did the disagreement surface?

Why was the disagreement heated?

What was the outcome?

How often has this situation come up in the past year?

(Option) Tell me about another one.

7. Sometimes we see friends, peers or fellow workers doing something incorrectly and we offer advice or criticism. Tell me about a time when you have offered such advice or criticism.

Possible Probes:

What were the circumstances that led to your offering the advice?

How did you approach the situation?

How was the advice received?

How did the situation end up?

(Option) Tell me about another one.

8. Can you think of a specific incident when you did something for a friend, peer or fellow worker without being asked?

Possible Probes:

When did this happen?

What were your reasons for taking this action?

How did they react?

What sort of feedback did you receive?

How often has this happened in the past year?

(Option) Tell me about another one.

Coping

9. Sometimes we encounter new or unusual situations with
which we are not familiar, and as a result we feel ill at
ease and/or don't know what to do. Tell me about a time
when you ran into such a situation.

Possible Probes:

>What were the circumstances?
>
>How did you deal with the situation?
>
>What was the outcome?
>
>How often have you had to deal with such a situation
over the past year?
>
>(Option) Tell me about another one.

10. Sometimes we are faced with very stressful situations
with which we do not feel we can cope. Tell me about a time
when you were faced with such a stressful situation.

Possible Probes:

>What were the circumstances?
>
>How did you react?
>
>How did things turn out?
>
>How often have you had to deal with such situations
over the past year?
>
>(Option) Tell me about another one.

Appendix D

PROTECTED
(When Completed)

INTERVIEW

ASSESSMENT FORM

Applicant

SIN: _____ Name: _____

CFRS Serial: _____ Date: _____ MOC: _____

Rater

Name: _____ Date: _____

Interviewer: _____ or Observer: _____ (check one) CFRC/Det: _____

Applicant Assessment

Teamwork (circle one)

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

Conduct (circle one)

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

Coping (circle one)

1	2	3	4	5	6	7
Substantially Below Average	Below Average	Slightly Below Average	Average	Slightly Above Average	Above Average	Substantially Above Average

Remarks: _____

_____

_____

PROTECTED
(When Completed)