



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

**PANEL VERSUS INDIVIDUAL INTERVIEWS: A META-
ANALYTIC INVESTIGATION OF EMPLOYMENT
INTERVIEW VALIDITY**

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Master of Social Science in Psychology
at the
University of Waikato
by
JUSTINE REBECCA GUILFORD

University of Waikato

1997

ABSTRACT

Further analysis using a similar data set to the McDaniel, Whetzel, Schmidt and Maurer (1994) meta-analysis of employment interviews was performed in the present study, in order to investigate four possible causes for the apparent superiority of individual employment interviews. These causes included (a) criterion contamination of individual interview studies, (b) greater prevalence of psychologists performing individual interviews, (c) greater number of trained/experienced individual interviewers, and (d) greater prevalence of high proximity to target positions in individual interviews. A research question was also proposed to investigate whether all interview panel sizes were inferior to individual interviews in terms of validity. Meta-analyses using 204 job and training performance validity coefficients indicated that individual interviews were superior in validity to panel interviews, but only when the criterion was training performance. Training/experience was the only explanation for the superiority of individual interview validity, such that individual interviewers were more likely to be trained/experienced, suggesting that individual interview validity was superior as a result. The explanation that the use of psychologists in interviews may account for superior individual interview validity was only partially supported. While psychologists were more prevalent in individual interviews, individual interview validity was lower than that of panel interviews, when the criterion was job performance. Only when training was the criterion was individual interview validity higher when psychologists were interviewers. Two hypotheses were not supported. Firstly, there was no evidence of criterion contamination inflating

individual interview validity. Secondly, no relationship between interviewer proximity and interview validity was found. Investigations of whether all panel sizes were inferior in terms of validity found that panel sizes of more than three members were superior in validity to individual interviews, when the criterion was job performance.

CONTENTS

	Page
Abstract	i
Contents	iii
List of Tables	v
Acknowledgments	vii
CHAPTER 1 INTRODUCTION	1
Panel Interviews: Are More Heads Be Better than One?	8
Aims of the Present Study	14
CHAPTER 2 METHOD	25
Data Set	25
Study Inclusion Rules	26
Coding Interview Studies	27
Reliability of Coding	31
Analyses	31
Artifact Information	35
CHAPTER 3 RESULTS	36
Overall Validity of Individual versus Panel Interviews	36
The “Criterion Contamination” Hypothesis	40
The “Psychologists vs Non-Psychologists” Hypothesis	43

The “Training/Experience” Hypothesis	47
Interviewer Proximity	50
Panel Size	52
CHAPTER 4 DISCUSSION	54
Limitations	59
Recommendations for Future Research	62
Conclusions	64
REFERENCES	65
APPENDICES	77
Appendix A Sources of Validity Data	77
Appendix B Coding Schedule	93

LIST OF TABLES

Table		Page
1	Summary of Previously Published Mean Validities for Employment Interviews	4
2	Summary of Previously Published Mean Validities of Panel and Individual Interviews (Criterion = Job Performance)	13
3	Comparison of The Validity Coefficients Used In The Present Study With McDaniel et al. (1994)	27
4	Judges' Percentage Agreement on Coding Schedule	32
5	Mean Validities for Panel and Individual Interviews	37
6	Mean Validities of "Multiple Assessor" and Individual Interviews	38
7	Panel and Individual Interview Validity Without Criterion Contamination	41
8	Mean Validities for Interviewers as Psychologists (Criterion = Job Performance)	44
9	Mean Validities for Interviewers as Psychologists(Criterion = Training Performance)	45
10	Mean Validities of Interviews for Training/Experience (Criterion = Job Performance)	48
11	Mean Validities of Interviews for Training/Experience (Criterion = Training Performance)	49

12 Summary of Mean Interview Validities for Different Panel Sizes 53

ACKNOWLEDGEMENTS

I would like to extend my thanks and appreciation to a number of people who were instrumental in the development of this thesis.

Firstly, my sincere thanks to Dr Paul Taylor who supervised this paper, for his invaluable advice, support, and patience throughout the year and numerous drafts! Secondly, I would like to sincerely thank Mike McDaniel for providing the validity data for further analysis, for his assistance in coding validity studies, and for his helpful comments. My thanks also to Joe Stauffer for the use of MetaQuik 16 and for his assistance in operating the programme. Finally, my thanks to Rob Bakker for his technical assistance.

I would next like to thank Karen Pratt and Nick Jacobs, with whom it was not only a privilege to share an office, but also for their support, understanding, and most of all, friendship. To my parents, Graeme and Kate, my sisters Gemma, Sarah, Cleo, and Felicity, and my niece Laura, thank you all for your love and understanding throughout the preparation of this thesis, particularly towards the end! Thank you also to Sharon, Derrek, Rodney, and Justin for being wonderfully tolerant flatmates. Finally, my thanks to Ross, during what has been a difficult couple of years, for his good humour and undying love and support.

CHAPTER ONE

INTRODUCTION

The employment interview is the most widely used selection technique in organisations for predicting the suitability of applicants for positions (Drake, Kaplan, & Stone, 1973; Guion, 1976; Harris, Toulson, & Livingston, 1996; Ryan & Sackett, 1987; Robertson & Makin, 1986; Taylor, Mills, & O'Driscoll, 1993; Ulrich & Trumbo, 1965;). Ryan and Sackett (1987) found in their survey of industrial and organisational psychologists that 93.8% used the interview as part of their assessments. Similarly, in New Zealand, the interview is used extensively across all organisations for selection purposes (Harris, Toulson, & Livingston, 1996; Taylor, Mills, & O'Driscoll, 1993). The widespread use of the employment interview as a selection technique has likewise been a popular focus of much research and analysis and several narrative and meta-analytic reviews have been performed on the employment interview literature since Wagner's (1949) first review.

Narrative reviews of the employment interview literature have often criticised the ability of interviews to reliably and validly predict applicant suitability for employment (Arvey, 1979; Arvey & Campion, 1982; Harris, 1989; Keenan, 1989; Mayfield, 1964; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949; Wright, 1969). For example, Wagner (1949) reported low validities for the traditional unstructured interview (median $r = .27$) and only moderate reliabilities

(median $r = .57$). Similarly, Mayfield (1964) noted that the employment interview literature still indicated low validities and reliabilities for the employment interview. Ulrich and Trumbo (1965) concurred with Mayfield's (1964) assertion of deficient validities and reliabilities, and directly questioned its continued use as a technique for selection, as did Wright (1969) and Schmitt (1976). The belief that employment interviews lacked predictive validity was upheld in a meta-analytic review conducted by Hunter and Hunter (1984) who found employment interviews to have mean validity of only .14. However, while this estimate of interview validity appears to be dismally low, Huffcutt and Arthur (1993) noted that the estimate was obtained from only ten correlation coefficients, and must therefore not be taken as a true indication of interview validity. Moreover, the coefficients used by Hunter and Hunter (1984) were only corrected for the statistical artifacts of sampling error and criterion unreliability, and not corrected for range restriction, which Hunter and Hunter (1984) acknowledge may have severely underestimated their estimate of employment interview validity.

While results of early reviews of the interview literature were discouraging, particularly in terms of validity, Dreher and Maurer (1989) noted that there were several complications with validity estimates of employment interviews. These estimates were complicated by the fact that validation study designs typically tended to aggregate multiple interviewers judgements of applicants, failing to take into account the differences in interviewer ratings (e.g., Albrecht, Glaser, & Marks, 1964; Miner, 1970; Tziner & Dolan, 1982). Dreher, Ash, and Hancock (1988) demonstrated that the use of aggregate analysis samples underestimates the validity of employment interviews and this underestimation is likely to be substantial. Secondly, there was a general failure to make important statistical adjustments in validity studies

of the employment interview, particularly for range restriction (which underestimates validity estimates, as demonstrated in the Hunter and Hunter (1984) estimate of interview validity). Finally, the existing domain of validity studies failed to take into account a number of methodological and contextual factors that were likely to affect interview validity estimates. Deficient criteria, lack of research on factors such as occupational and job differences, the purpose of the interview, and interview settings and formats all contributed to the underestimation of interview validity, according to Dreher and Maurer (1989).

The complications with inadequate validation study designs, lack of statistical adjustments, and a general failure to address methodological and contextual factors in older employment interview studies influenced the spurious conclusion that employment interviews were neither reliable or valid (Dreher & Maurer, 1989). However, the results of recent reviews which have utilised meta-analytic techniques to determine the true validity of employment interviews have provided researchers with evidence to consider employment interviews more positively (Huffcutt & Arthur, 1994; Huffcutt & Woehr, 1993; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Searcy, Woods, Gatewood, & Lance, 1993; Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989). Table 1 presents the results of previous meta-analytic findings.

The estimates obtained by Wiesner and Cronshaw (1988), McDaniel et al. (1994), and Huffcutt and colleagues, were all corrected for the statistical artifacts of sampling error, criterion unreliability and range restriction, while those estimates obtained by Searcy et al. (1993) and Wright et al. (1989) were corrected for predictor

Table 1

Summary of Previously Published Mean Validities for Employment Interviews

Authors	Mean Validities		
	Overall	Panel	Individual
All Interviews			
Wiesner & Cronshaw (1988)	.47	.44	.44
McDaniel et al. (1994)	.37	.32	.43
Huffcutt & Woehr (1993)	.43		
Huffcutt & Arthur (1994)	.37		
Unstructured Interviews			
Hunter & Hunter (1984)	.14 †		
Wiesner & Cronshaw (1988)	.31	.37	.20
McDaniel et al. (1994)	.33	.33	.34
Structured Interviews			
Wiesner & Cronshaw (1988)	.62	.60	.63
Wright et al. (1989)	.38 ‡		
McDaniel et al. (1994)	.44	.38	.46
Searcy et al. (1993)	.49 ‡		

Note. Blanks indicate that panel versus individual interviews were not investigated in the research, and therefore no validity estimates were available. "Structure" was operationalised in slightly different ways by each researcher. Unless otherwise noted, all estimates are corrected for sampling error, range restriction, and criterion unreliability. † - corrected for sampling error and criterion unreliability. ‡ - corrected for sampling error, predictor and criterion unreliability, and range restriction.

unreliability in addition to all of the aforementioned statistical artifacts. The validity estimates obtained for employment interviews in general, that is, irrespective of

moderators such as structure or panel versus individual interviews, were much higher than the estimate calculated by Hunter and Hunter (1984). For example, Huffcutt and Arthur (1994) and McDaniel et al. (1994) both estimated the corrected mean validity for employment interviews to be .37, while Huffcutt and Woehr (1993) and Wiesner and Cronshaw (1988) estimated the corrected mean validity of interviews to be .43 and .47, respectively. These results indicate that the overall validity of employment interviews was much higher than previously believed.

In addition to assessing the overall validity of employment interviews, authors of meta-analytic reviews have explored a number of variables hypothesised to moderate, or influence the predictive validity of employment interviews. Such moderators included interview structure (Huffcutt & Woehr, 1993; Huffcutt & Arthur, 1994; McDaniel et al, 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988; Wright et al., 1989), interview content (McDaniel et al., 1994; Searcy et al., 1993), criterion purpose (McDaniel et al., 1994; Searcy et al., 1993), test information and ancillary data (McDaniel et al., 1994; Searcy, et al., 1993), rating procedures (Searcy et al., 1993), job performance criteria (Searcy et al., 1993), and panel versus individual interviews (Huffcutt & Arthur, 1994; McDaniel et al., 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988). In addition to the moderators of interview validity, Conway, Jako, and Goodman (1995) analysed moderators of interview reliability, such as study design, that is panel versus individual interviewers, interviewer training, and interview structure. Of the moderators explored in meta-analyses, interview, structure appeared to be the largest moderator of interview validity and reliability (Conway et al., 1995; Huffcutt & Arthur, 1994; Huffcutt & Woehr, 1993; McDaniel et al., 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988;

Wright et al., 1989). For example, Table 1 shows that Wiesner and Cronshaw (1988) reported a corrected mean validity estimate of .62 for structured interviews, while other corrected estimates of structured interview validity were lower ranging from .38 (Wright et al., 1989) to .49 (Searcy et al., 1993). On the other hand, unstructured interviews fared far worse, yielding mean validities ranging between .14 (Hunter & Hunter, 1984, corrected for sampling error and criterion unreliability) to .33 (McDaniel et al., 1994, corrected for sampling error, criterion unreliability, and range restriction). The differences in predictive validity of structured and unstructured interviews was further highlighted by Huffcutt and colleagues, who argued that interview structure was more diverse than a simple dichotomisation of the variable, and proposed a taxonomy of dimensions of interview structure. While this approach may be more methodologically correct, the results told a similar story to other meta-analytic investigations of interview structure - structured interviews were more valid than unstructured interviews (Huffcutt & Arthur, 1994; Huffcutt & Woehr, 1993). In terms of reliability, Conway et al. (1995) found that interviews were more reliable when interview questions were more standardised, while Wiesner and Cronshaw (1988) reported mean reliability coefficients of .82 and .61 for structured and unstructured interviews, respectively. In sum, structure was found to have a profound influence on the reliability and predictive validity of employment interviews.

Although interview structure is perhaps the most notable moderator of interview validity and reliability, other moderators have also been found to influence interview validity and reliability. For instance, the content of the interview affects employment interview validity. Content refers to the type of questions used, such as situational or behavioural questions. McDaniel et al. (1994) found that interviews

which used situational interview questions were higher in predictive validity than interviews based on other types of questions, such as psychological or job-related questions. Similarly, Searcy et al. (1993) found in their analysis of structured interviews that situational interviews were also more valid than interviews based on behavioural or non-behavioural questions. Another interesting finding was that interview validity was lower when interviewers had access to tests scores or any other information prior to an interview, such as biographical data (McDaniel et al., 1994; Searcy et al., 1993). Interview validity was also shown to vary as a function of the performance criterion used. For example, Searcy et al. (1993) found that structured interviews were more valid when subjective job performance criteria, rather than objective criteria such as tenure, were used. Similarly, McDaniel et al. (1994) also found that subjective criteria such as job or training performance yielded higher mean corrected validities for interviews than an objective measure of performance, in particular, tenure.

The purpose of the criterion has an influence on the predictive validity of interviews, such that mean corrected validity was higher when job performance criteria were collected for research rather than administrative purposes. McDaniel et al. (1994) noted that estimates based on administrative criteria were more likely to be biased and contaminated by errors, such as halo effect, thereby decreasing validity. Another moderator of interviews, interviewer training, was completely overlooked in meta-analyses of interview validity, however, Conway et al. (1995) addressed this moderator in their investigations of interview reliability. These authors concluded that interview reliability was higher when interviewers were trained, and recommended some form of training for interviewers. Finally, job analysis was also shown to affect

interview reliability and validity. Wiesner and Cronshaw (1988) found that structured interviews were more valid when the interview questions were based on a formal rather than less formal job analyses, while Conway et al. (1995) noted that job analysis had an effect on interview reliability, albeit an indirect one, but nevertheless endorsed the need for job analyses to be performed when developing interviews for selection purposes. One other variable that was found to influence the validity and reliability of employment interviews (and was often analysed in conjunction with interview structure), was panel versus individual interviews (Conway et al., 1995; McDaniel et al., 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988). It was with this particular factor that the present study was concerned. The following section discusses panel and individual interviews in more detail.

Panel Interviews: Are More Heads Better Than One?

A number of meta-analytic reviews have compared the predictive validity and the reliability of panel and individual interviews (Conway et al., 1995; McDaniel et al., 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988). The notion of using a group to conduct an employment interview is appealing for a number of reasons. Firstly, the group literature suggests that groups may be better at tasks such as interviewing. For instance, groups have been found to be superior to individuals at recalling information (Martell & Borg, 1993; Stasser & Titus, 1987). Moreover, aggregating the individual opinions of interviewers is thought to improve the quality of decisions by balancing and cancelling out random error. Finally, the diversity of information and opinions

that are inherent in a group may enhance the accuracy of the decision made (Maier, 1967).

Secondly, industrial and organisational psychologists have advocated the use of multiple sources of information for the completion of such tasks as job analysis and performance appraisal. In reference to job analysis, many authors writing about job analysis methods have encouraged the use of multiple sources of information in the collection of job analysis data (e.g., Fisher Schoenfeldt, & Shaw, 1993; Gael, 1983; Gatewood & Feild, 1994; Harvey, 1990). Harvey(1990) noted that “...whenever possible, multiple raters...should be used to improve the quality of job analysis data” (Harvey, 1990, p. 112). By using multiple sources of information (as opposed to consulting only one source), it has been argued that the outcome of job analyses would present a more accurate description of a particular job (Gael, 1983).

Similarly, the use of multiple raters has been advocated in the field of performance appraisal. The notion of utilising multiple raters for the administration of performance appraisals is relatively new. Latham and Wexley (1981), along with Landy and Farr (1983), suggested that the ideal appraisal system is one based on multiple sources. This approach has a number of advantages, including an increased likelihood of obtaining a more complete understanding of an employee’s performance, thus leading to a more valid rating. In addition, an employee’s acceptance of the ratings of his/her performance was thought to increase. More recently, the multiple rater approach to performance appraisals has been more commonly referred to as 360-degree feedback, a technique involving multiple raters, including self-assessments, in the assessment of individuals. Tornow (1993) noted that this technique has been used for a variety of purposes, including performance appraisal,

facilitating personal development of managers (Hazucha, Hezlett, & Schneider, 1993), succession planning and assignment selection (Moses, Hollenbeck, & Sorcher, 1993), and organisational change (London & Beatty, 1993). The use of 360-degree feedback in performance appraisal for personal development has shown promise as a valid alternative to traditional performance appraisals in terms of acceptance and increased managerial performance (Hazucha et al., 1993), although it was not clear whether such a technique was effective for increased performance at other levels. While the use of multiple sources of information appears to have great appeal for job analysis or performance appraisal, there is no evidence to suggest that such an approach was superior to individuals.

The support for the use of groups for performing tasks is by no means only confined to the areas of job analysis and performance appraisal. The selection interview literature has also voiced its support for the use of groups for performing employment interviews, such that group or panel interviews have been purported to increase the reliability and validity of interviews (Anstey & Mercer, 1956; Arvey & Campion, 1982; Campion, Pursell, & Brown, 1988; Dipboye, 1991; Dipboye & Gaugler, 1993; Hollingworth, 1922; Lowry, 1994; Mayfield, 1964). Arvey and Campion (1982) noted in their narrative review of selection interviews that interviews conducted by panels appeared to show promise as a means of enhancing reliability and validity. This assertion was based on the favourable reliability and validity estimates obtained from authors of panel interview studies such as Landy (1976), Anstey (1977), and Reynolds (1979). Arvey and Campion (1982) argued that the sharing of perceptions and ideas with different interviewers might force panel members to be more attentive to the irrelevant inferences made on non job-related variables. In

addition, M. Campion et al. (1988) noted that using a panel for the interview process reduces the effect of idiosyncratic biases that were more likely to occur when individual interviewers were used.

In sum, a group approach to completing tasks, such as selection interviewing, job analysis or performance appraisal, appears to be popular. It is therefore surprising that meta-analytic reviews have found a group approach to interviewing to be no better, and even inferior to individual interviews in terms of predictive validity (McDaniel et al., 1994, Searcy et al., 1993, Wiesner & Cronshaw, 1988). The following section reviews the meta-analytic findings regarding panel versus individual interviews.

Recent Meta-analytic Evidence: More Heads Are Worse Than One

The results of meta-analyses performed by McDaniel et al. (1994), Searcy et al. (1993), and Wiesner and Cronshaw (1988) all indicated, contrary to popular belief, that groups were no better and possibly worse than individuals at predicting applicant suitability for employment. Wiesner and Cronshaw (1988) hypothesised that panel interviews would be more valid than individual interviews, because panel interviews would be more reliable, and reliability sets the upper limit of validity. McDaniel et al. (1994) made no such formal hypothesis, although they suggested, given the higher administration costs associated with performing panel interviews, that panel interviews ought to be more valid. Finally, based on suggestions of moderators of structured interviews by Dipboye and Gaugler (1993) who argued that panels would increase the reliability and validity of structured interviews, Searcy et al. (1993)

compared the amount of non-artifactual variance in the estimated population means (ρ) that was explained by panel versus individual interviews. Table 2 presents the estimates acquired by Wiesner and Cronshaw (1988) and McDaniel et al. (1994) for panel versus individual interview validity and the interaction with structured and unstructured interviews.

When panel and individual interview validity was considered without reference to structure, Wiesner and Cronshaw (1988) reported that the mean corrected¹ validities for panel and individual interviews were equal (.44), while McDaniel et al. (1994) found that individual interviews were superior in validity to panel interviews, yielding corrected mean validities of .43 and .32, respectively. Similarly, when panel and individual interview validity was considered with interview structure, McDaniel et al. (1994) reported substantially higher validity for structured individual interviews than for structured panel interviews (.46 versus .38, respectively), however, the difference reported by Wiesner and Cronshaw (1988) between structured panel and individual interviews was not as great (.60 versus .63, respectively). The only instance in which panel interviews were reported to have higher validity than individual interviews was when the panel interview was unstructured (Wiesner and Cronshaw, 1988), such that the corrected mean validities were .37 for panel interviews and .20 for individual interviews. Conversely, McDaniel et al. (1994) found that the difference in the corrected mean validity of unstructured panel and individual interviews was negligible (.33 versus .34, respectively).

¹ "Corrected" refers to the correction of mean validities for the statistical artifacts of sampling error, criterion unreliability, and range restriction unless stated otherwise.

Table 2

Summary of Previously Published Mean Validities of Panel and Individual Interviews
(Criterion = Job Performance)

Author	Mean Validities		
	Overall	Panel	Individual
All Interviews			
Wiesner & Cronshaw (1988)	.47	.44	.44
McDaniel et al. (1994)	.37	.32	.43
Unstructured Interviews			
Wiesner & Cronshaw (1988)	.31	.37	.20
McDaniel et al. (1994)	.33	.33	.34
Structured Interviews			
Wiesner & Cronshaw (1988)	.62	.60	.63
McDaniel et al. (1994)	.44	.38	.46

All estimates are corrected for sampling error, range restriction, and criterion unreliability.

While the estimates reported by McDaniel et al. (1994) and Wiesner and Cronshaw (1988) are slightly different in terms of the size of the validities, the conclusions of both meta-analyses remained the same - individual interviews were superior in predictive validity to panel interviews. These differences may be attributable to slightly different definitions of structure, and differences in the correlation coefficients used. However, neither of these authors statistically tested these differences in individual and panel interview validity to determine how likely they were to have occurred by chance alone. Searcy et al. (1993) came to a similar conclusion in their search for moderators of structured interviews, such that multiple

interviewers decreased interview validity. In terms of the reliability of panel and individual interviews, Wiesner and Cronshaw (1988) reported higher mean reliability coefficients for panel interviews compared to individual interviews (.85 versus .78, respectively). Conway et al. (1995) found that when interview questions were highly standardised, panel interviews were more reliable than individual interviews (.91 versus .59). Although panel interviews appeared to be more reliable than individual interviews, this was not reflected in the corrected mean validity estimates of panel interviews, despite suggestions that panel interviews should be more reliable *and* valid than individual interviews (e.g., Dipboye, 1991; Dipboye & Gaugler, 1993). Therefore the impetus for the present study stems from the lower validity of panel interviews observed in recent meta-analyses, given predictions which indicated that panel interviews would be more valid than individual interviews.

Aims of the Present Study

The aims of the present research were twofold: (1) to determine why panel interview validity was no better than individual interview validity, by investigating the possible causes for the inferior validity of panel interviews found in previous meta-analyses, and (2) to investigate the effect of panel size on Panel interview validity.

Aim 1

Three previous meta-analyses have found panel interviews to be no better and possibly worse than individual interviews in terms of predictive validity, yet the social and industrial and organisational psychology literature favoured a group approach to accomplishing tasks that require making decisions or judgements. Given that there

was also support for group approaches to selection interviewing in the selection interview literature (e.g., Arvey & Campion, 1982; Campion et al, 1988; Dipboye, 1991; Lowry, 1994), and given that other factors were not been considered in the panel and individual comparisons performed by McDaniel et al. (1994), Searcy et al., (1993), and Wiesner & Cronshaw (1988), it would be premature to conclude that a group approach to interviewing is an inefficient method of selection. Therefore, before discounting the predictive validity of panel interviews as such, a final look at possible causes for the inferiority of the predictive validity of panel interviews was warranted.

A number of possible causes for the inferiority of panel interview validity to individual interviews were postulated in the present study, which will be summarised here and elaborated on in the subsequent pages. These included:

- a) Criterion contamination artificially inflating individual interview validity
- b) Individual interview validity may be superior because interviewers may also be psychologists
- c) Trained/experienced interviewers make more valid decisions, and are more likely to be prevalent in individual interviews
- d) Interviewers more proximal to target positions may make more valid selection decisions, and are again more likely to be prevalent in individual interviews.

A detailed discussion of each of the possible causes of panel interview validity inferiority follows below.

Criterion Contamination. One of the prominent sources of criterion bias is knowledge of predictor information. Cascio (1987) provided an example of assessment center selection of executives relevant to the present study:

“If an individual’s immediate superior has access to the prediction to this individual’s future potential by assessment centre staff, and at a later date the supervisor is asked to rate the individual’s performance, the supervisor’s prior exposure to the assessment center prediction is likely to bias this rating. If the subordinate has been tagged as a “shooting star” by the assessment centre staff and the supervisor values that judgement, he or she may too, rate the subordinate as a “shooting star”..[and vice versa]. In either case - spuriously high or spuriously low ratings - bias is introduced and gives an unrealistic estimate of the validity of the predictor” (Cascio, 1987, pp. 65-66).

In the McDaniel et al. (1994) meta-analysis, interviews were found to be more valid when they were conducted by individuals rather than by a panel. However, the interview validity for individual interviews may, in fact, be overestimated due to criterion contamination. Since the interviewer in an individual interview is most likely to be the position manager, the interviewer is also just as likely to be the rater of performance. As Cascio (1987) illustrated, this situation would constitute criterion contamination, suggesting that interview validity for individual interviews could be artificially inflated, thus projecting the illusion that individual interviews are more valid than panel interviews. It is possible that criterion contamination may also be present in panel interview studies, however, I would argue that this problem would not present itself in the same degree as in individual interview studies. Because panel interviewers are more likely to be removed from the target position, it is feasible to

assume that panellists would be less likely to be involved in the administration of performance appraisals. The present study investigates the validity of interviews by removing those studies where interviewers were also raters of performance, thus providing a more accurate estimate of interview validity.

Hypothesis 1: Individual interview studies are more likely than panel interview studies to have criterion contamination, and when studies with criterion contamination are removed, mean individual interview validity might be expected to be no higher, and possibly less than mean panel interview validity.

THE UNIVERSITY OF WAIKATO
LIBRARY

Psychologists versus Non-Psychologists. Another explanation for the inferior validity of panel interviews found in the previous meta-analytic reviews may be that interviewers who are also psychologists increase the validity of selection decisions, and individual interviewers are more likely to be psychologists than panel interviewers, because psychologists should have greater expertise in selection interviewing. In a review of expertise and performance, Chi, Glaser, and Farr (1988) drew a number of conclusions. First, experts excel mainly in their own domains. Second, experts perceive large, meaningful patterns in their domains. Third, experts see and represent a problem in their domain at a deeper, more principled level than novices. Finally, experts have strong self-monitoring skills, meaning that experts were more aware when errors were made and when to check for errors. Therefore, experts might be expected to perform better than novices. Support for the expert hypothesis was also found by Gaugler, Rosenthal, Thornton, and Bentson (1987) in their review of assessment center validity such that psychologists' assessments were significantly

more valid than those made by managers. This finding contradicted a widely held belief that managers were more valid assessors, since managers were more familiar with job requirements than psychologists. The implications of these conclusions for the present study are that psychologists who are selection interviewers are more likely than non-psychologists to have expertise in the area of selection interviewing, and therefore make more valid selection decisions than non-psychologists. Psychologists are more likely to be involved in conducting individual interviews than panel interviews, since it would be costly for organisations to employ a panel of psychologists. While it is feasible that an interviewing panel may have a psychologist in its midst, it is more likely that psychologists would be prevalent in conducting individual interviews. The hypothesis concerning psychologists is as follows:

Hypothesis 2: Psychologists are more likely to make more valid interview judgements than non-psychologists, and psychologists are more likely to be prevalent in individual interviews than in panel interviews.

Interview Training/Experience. Interviewers have been found to make numerous errors in conducting the interview and evaluating applicants (Gatewood & Feild, 1994; Mayfield, 1964), such as using interview questions inconsistently with applicants, questions which are not job -related or which do not pertain to the applicant's ability to perform the job. Other errors include overestimating their ability to evaluate applicants, which can lead to hasty selection decisions, which may be influenced by all manner of personal biases. Employment interviewers are also subject to making rating errors, such as central tendency, leniency, and stringent errors in

their evaluations. Other errors such as halo, contrast, first impression, and similar-to-me effects have also been shown to influence interviewer judgements. Training interviewers in conducting interviews, the interview process, decision-making methods, and systematic scoring, can reduce the number of errors made by interviewers, and thus improve the reliability and validity of the employment interview (Dipboye, 1991; Dipboye & Gaugler, 1993; Gatewood & Field, 1994).

Interview experience and training is intuitively appealing as a moderating factor in interview validity (Keenan, 1989), since trained/experienced interviewers are more likely to know how to interview applicants effectively, process information presented in the interview, and be able to judge people better than those that are untrained/inexperienced. Arvey and Campion (1982) noted that previous findings suggested that intensive training courses that included practice with feedback and group discussions helped in reducing rating errors of contrast, halo, similarity, and first impression (Latham, Wexley, & Pursell, 1975; Wexley, Sanders, & Yukl, 1973). Keenan (1978) also found that trained interviewers were less biased in their evaluations of applicants, while other research found no significant effects for training in reducing rater errors (Vance, Kuhnert, & Farr, 1978; Maurer & Fay, 1988). There is some evidence to suggest that training may increase interview validity and reliability. For example, Dougherty, Ebert, and Callender (1986) obtained higher interview validities after interviewers were trained than prior to training, while Heneman (1975) found that low reliabilities resulted when untrained interviewers observed unstructured interviews. Conway et al. (1995) found that interview reliabilities were higher when interviewers were trained.

The case for experienced interviewers is not so clear, as there has been little research conducted on whether the effectiveness of interviewers' judgements varies with the amount of experience. One study, conducted by Keenan (1978), investigated interviewer experience and interviewer effectiveness and found that experienced interviewers were more biased than interviewers with less experience (when bias was operationalised as a correlation between personal liking for the candidate and a judgement of suitability). However, given that there has been little attempt to investigate interviewer effectiveness and experience, it may be too presumptuous to draw any major conclusions from this finding.

Despite the inconsistencies regarding interviewer training and lack research with respect to interviewer experience in the selection interview literature, support for training interviewers still prevails (Conway et al., 1995; Dipboye, 1991; Dipboye & Gaugler, 1993). In fact, Conway et al. (1995) recommended that interviewers should be trained in some capacity, based on their finding that training increased interview reliability. Moreover, there is evidence in the performance appraisal literature to suggest that rater training can increase accuracy (e.g., Sulsky & Day, 1992).

The concepts of training and experience were treated as a single variable in the present study based upon the following premise that interviewers who have either training or experience, suggests that they have some skills in interviewing, and therefore may contribute to superior validity. If training/experience does increase interview validity, then one explanation for the superiority in interview validity for individual interviews may be that individual interviewers are more likely to have training/experience than panel interviewers. If an organisation was to leave a selection decision to one person rather than a panel, then they would be more likely to

leave the interviewing task to a person with training/experience, rather than to someone without such skills. Therefore, the training/experience hypothesis states:

Hypothesis 3: More experienced/trained interviewers make more valid selection decisions than interviewers with no training/experience, and individual interviewers are more likely to be trained/experienced.

Proximity of Interviewer to Target Position. One of the possible explanations for the inferiority of panel interview validity may be found in the assessment of interviewers' proximity to the target position. Interviewers who are familiar and knowledgeable about the target position and its requirements, may increase the likelihood that a more valid selection decision will be made. Devine and Kozlowski (1995) investigated the effects of domain-specific knowledge and task characteristics on decision making. The results showed that individuals who had high degree of knowledge were more accurate on well-structured decisions, but are no better than low-knowledge individuals on ill-structured decisions. Secondly, high-knowledge individuals reduced information search when decisions were well-structured and alternatives were descriptively labelled. Finally, high knowledge individuals utilised more contextual information than low knowledge individuals. This supports the hypothesis that interviewers who are closely related to the target position are likely to have more relevant knowledge than individuals who are higher in the organisational hierarchy, from different departments, or from outside of the organisation. Therefore, interviewers who are more proximal to the target position might be expected to make more valid judgements than interviewers who are not.

Individual interviews are most likely to be conducted by persons proximal to the target position (such as the immediate manager of the target position or someone who is proximal to the vacant position), while panel interviews are more likely to comprise of members who are less related or proximal to the target position.

Therefore:

Hypothesis 4: Interviewer proximity to the target position is positively correlated with interview validity, and individual interviewers are more likely to have greater proximity to the target position.

Aim 2

The literature on groups and group performance indicates that groups can be more effective than individuals (in particular to average individual performance), in terms of the quality of decisions made (e.g., Bottger & Yetton, 1988; Martell & Borg, 1993; Miner, 1984;). In relation to group size, Littlepage (1991) noted that evidence of the relationship between group size and group performance was inconsistent, such that some studies indicated that increases in group size had no effect on performance, while other studies showed that performance increased with group size. Still other studies indicated that group performance increases with group size but with diminishing returns. Therefore, the nature of the relationship between group size and performance appears to vary as a function of the task. Of the tasks identified by researchers in the group literature, decision-making or judgemental tasks best characterise the task of employment interviewing, as interviewers are required to make a decision or judgement as to which applicant is best suited to the target

position. Most of the literature on group size and group performance has focused on tasks other than judgment or decision-making tasks, and there has been a general reluctance to put a number on an optimal group size for such tasks. However, Robbins (1993) and Worchel, Cooper, and Goethals (1990) noted that group performance would decrease when the size of a decision-making group exceeds five members.

Interestingly, several studies in the selection interview literature have suggested that an optimal group size for panel interviewing might range between two and five members (Campion et al., 1988; Daniel & Valencia, 1991; Lowry, 1994). However, such suggestions were not based on any evidence that panel interview validity increased as a result of the panel size ranging between two and five members. In fact, there appears to be no evidence that panel interview validity varies as a function of panel size. Rather, it would seem that the suggestion made by Campion et al. (1988), Daniel & Valencia (1991), and Lowry (1994) was derived from the number of interviewers typically employed in a panel interview. Therefore, based on these assumptions that the best size for an interviewing panel *might* range anywhere between two and five members, the following research question was proposed: Are all panel sizes inferior to individual interviews in terms of validity? By analysing the validity of various interview panel sizes, a clearer understanding of the effect group size has on decision-making or judgemental tasks may be achieved.

In sum, the first aim of the present study was to determine why earlier meta-analytic reviews found panel interview validity to be no better or worse than individual interview validity, by investigating the possible causes. Four possible causes were identified and included: (1) criterion contamination in individual interview

studies, (2) greater prevalence of psychologists in individual interview studies, (3) greater pervasiveness of training/experience in individual interview studies, and (4) greater prevalence of proximal interviewers in individual interview studies. The second aim was to investigate whether all panel sizes were inferior to individual interviews in terms of validity.

CHAPTER TWO

METHOD

Data Set

A total of 204 interview validity coefficients were analysed in the present study. The validity data were obtained from validation studies in which job performance and training performance were used as criteria for assessing the validity of selection interviews. The studies used in the present analysis were those employed in a previous meta-analysis of employment interviews conducted by McDaniel et al. (1994).

The studies were located by McDaniel and colleagues in an extensive literature search extending over a period of eight years. Validity data was obtained from the database of validity coefficients from the U. S. Office of Personnel Management and from the reference lists of five previous meta-analyses (Dunnette, Arvey, & Arnold, 1971; Hunter & Hunter, 1984; Reilly & Chao, 1982; Wiesner & Cronshaw, 1988; Wright et al., 1989). These validity coefficients were obtained from both unpublished and published research. Sources included journal articles (n = 60), dissertations (n = 20), technical reports (n = 32), master's theses (n = 1), unpublished or submitted manuscripts (n = 1), books (n = 4), and conference papers (n = 9). Research was conducted in a variety of settings, such as education (n = 16), business (n = 31), armed forces and law enforcement agencies (n = 46), manufacturing (n = 25), and other areas (n = 9), for example the Peace Corps. In some cases, studies included

more than one coefficient (e.g., Kennedy, 1986), all of which were coded. While this may appear to violate the principles of independence, it must be noted that some studies reported separate coefficients for training and job performance criteria, while others simply reported separate coefficients for separate studies.

Study Inclusion Rules

Several sources of validity data included in the McDaniel et al. (1994) analysis were excluded from the present study for a variety of reasons. First, studies excluded by McDaniel et al. (1994) were not included in the present analysis² Secondly, several validation studies could not be located by McDaniel in time for the present analyses. Of those studies not located were seven validity studies (Berkley, 1984; Delaney, 1954; English, 1983; Flynn & Peterson, 1972; McKinney, 1975; Wayne County Civil Service Commission, 1973) and six personal communications to McDaniel (Ard, 1985; Bosshardt, 1993; Brown, 1986; Davis, 1986; Maurer, 1986; Pearlman, 1978).

Table 3 shows the comparison of the number of studies used by McDaniel et al. (1994) and in the present study. As a result of these decision rules, over 200 validity coefficients remained for analysis in the present study. A list of all studies used for analysis in the present study are presented in Appendix A.

² The reader is referred to the "Decision Rules" section in the McDaniel et al. (1994) study for further clarification of the decision rules employed for the inclusion/exclusion of studies from analyses.

Table 3

Comparison of The Validity Coefficients Used In The Present Study with McDaniel et al. (1994)

Criterion	Number of Coefficients	
	McDaniel et al. (1994)	Present Study
Job Performance	160	136
Training Performance	75	68
Total	235	204

Coding Interview Studies

Each of the studies included in the present meta-analysis were coded by two independent judges, using a coding sheet developed from the hypotheses. Each coding category is described below.

Interview type. All studies were coded in terms of whether the interview was conducted by an individual or a panel of interviewers. In the present investigation, a “panel interview” was defined as one in which two or more interviewers were present at the same interview. McDaniel et al. (1994) treated such interviews and others in which individuals assessed applicants on separate occasions, yet made a collective decision, as “board” interviews. These were coded as individual interviews in the present analysis. While this is technically a form of panel interview (Dipboye, 1991), it was not considered to be a pure form of the panel interview as such a group of interviewers were not witness to the same applicant behaviours.

Criterion contamination. In order to investigate the impact of criterion contamination on interview validity, the judges were first required to ascertain for each study whether the performance criterion used for the computation of criterion-related validity was objective or subjective. Objective criteria referred to criteria such as tenure, test scores, and sales per hour (Weekley & Gier, 1987). Those studies identified as consisting of objective criteria were not coded further, since criterion contamination was not viewed to be an issue for such studies. Subjective criteria referred to criteria that consisted of appraiser judgements which included ratings or rankings on scales to assess performance such as BARS or graphic rating scales. Those studies identified to have used subjective criteria in the calculation of criterion-related validity were then coded as to whether the interviewers were also raters of performance, using a classification system comprised of four categories which were as follows: (a) it was certain that the selection interviewers were also raters of performance; (b) it was probable that the selection interviewers were also raters of performance; (c) it was certain that the selection interviewers were not raters of performance; and (d) it was probable that the selection interviewers were not raters of performance. These four categories were later collapsed into two categories which indicated as to whether the selection interviewers were or were not the raters of performance. Both individual and panel interview studies posed a few problems for coding criterion contamination, as some studies did not identify interviewers in relation to the target position. Therefore, it was difficult to determine as to whether the interviewers were associated with appraising performance. In such cases, these studies were not included for the analysis of criterion contamination. Also, there were some instances in which one panel member was the only rater of performance, however, this still constituted criterion contamination, and was coded accordingly.

Interviewers as psychologists. In order to investigate the impact a psychologist has on interview validity, and whether individual interviewers were more likely to be psychologists, the judges coded whether each interviewer was or was not a psychologist, irrespective of whether the interview was conducted by a panel or an individual. For those interviews with multiple raters, that is for both true panels and multiple rater interviews, there was often little information given regarding the interviewers, and thus made coding for psychologists difficult. However, in most cases, panels were often a mix of psychologists and non-psychologists, and thus separate codings for each interviewer were made. When the interview was conducted by a panel, studies in which there was at least one psychologist present in the interview panel were included in the analysis of psychologists. However, these were not included in the analysis of non-psychologists.

Training/experience. All studies were coded for interviewer training/experience. The coding sheet allowed for one of two possible codings for training/experience which included (1) yes and (2) no. A “yes” code included all those studies which stated or implied that interviewers had training/experience. For example, a number of validity studies used in the present met-analysis used the “Teacher Perceiver” interview (a structured interview used in the United States for the selection of teachers), and it was often never stated that the interviewer, who was also the researcher in most cases, had training/experience. However, in order to conduct a “Teacher Perceiver” interview, the interviewer is required to undergo an intensive training course (Zarnek, 1983), and therefore a “yes” was coded for training/experience in such cases. Similarly, those interviewers who were members of a personnel department within the organisation in which interviews were conducted

were also coded as trained/experienced. Panel interview studies in which some interviewers were trained/experienced and other panel members were not, a “yes” to training/experience was coded, as such panels were deemed to have some training/experience.

Interviewer proximity. The positions of interviewers in all of the studies were coded by the judges in order to analyse interviewer proximity to the target position. For panel interview studies, each interviewer was coded individually for his/her position in relation to the target position. In the original coding schedule, the following types of interviewer position categories relative to the target position were identified: (1) immediate manager; (2) higher-level manager; (3) incumbent; (4) personnel specialists or any personnel/human resources position; (5) interviewer from another department; (6) peer to target position; (7) subordinate to target position; (8) other member of the organisation; and (9) not a member of the organisation. For ease of analysis, these categories were later collapsed into a three-point scale ranging from low (1) to high (3) proximity. “Low” proximity included categories five, seven, eight and nine, while categories two, four and six were classified as “medium” proximity, and categories one and three classified as “high” proximity to the target position. In many cases, particularly for panel interviews and individual interviews in which there were multiple raters, interviewers varied in proximity to the target position, and therefore an average rating was calculated.

Each study was coded using the aforementioned rules. If a study provided insufficient information on a particular factor, it was coded as missing information and excluded from that particular analysis. For example, if it was not clear in a study whether the interviewer was or was not a rater on the performance criterion, the study

was omitted from the criterion contamination analysis. The coding sheet used in the present study is presented in Appendix B.

Reliability of Coding

Table 4 presents reliability information for the coding of studies. Two judges coded each study independently and then met periodically to compare codings. Discrepancies were discussed and an agreement was reached. In some cases, discussions of discrepant codings were not resolved, and a third judge was called upon to resolve a coding disagreement. The reliability prior to discussion for panel versus individual was 87%, 89% for objective versus subjective criteria, 82% for interviewer also criterion rater, and 80% for training/experience. The reliabilities for psychologist versus non-psychologist and interviewer position were averaged, since there was more than one interviewer in some studies. The average reliability for psychologist versus non-psychologist was 99%, while the reliability for interviewer position was somewhat lower with only 68% agreement. After discussion, agreement on the codings reached 100%.

Analyses

Meta-analysis was used for performing the analyses in the present study. It is a technique that evolved from a need to make sense of a vast number of accumulated study findings in many research domains. There are many other methods of integrating study results across studies, although these are generally inferior to meta-analysis. For example, the traditional narrative review is one method of comprehending study results. However, this method has been shown to be unreliable. Hunter & Schmidt (1990) noted that Cooper & Rosenthal (1980, cited in Hunter &

Table 4

Judges' Percentage Agreement on Coding Schedule.

Question #	Percentage Agreement
Panel or Individual Interview	87
Objective versus Subjective Criteria	89
Interviewer also Criterion Rater	82
Psychologist versus Non-Psychologist	99
Interviewer Training/Experience	80
Interviewer Position	68

Note: Agreements reached 100% after third party mediation. The percentage agreements for "Psychologist versus Non-Psychologist" and "Interviewer Position" categories (5 each) were averaged for studies with multiple interviewers.

Schmidt, 1990) found that even when the size of the sample of studies was as small as seven, reviewers who used narrative methods or quantitative methods reached different conclusions. Wolf (1986) noted many other potential disadvantages of the traditional narrative review method. Firstly, reviewers may selectively include studies. Secondly, there is a danger that studies may be weighted subjectively, thereby influencing the conclusions made. Thirdly, there is a likelihood of failure to examine characteristics of studies as possible explanations for divergent or consistent results across studies. Finally, narrative reviewers may fail to examine moderating variables in the relationship of interest. A number of narrative reviews of the employment interview literature have been conducted over the years (Arvey, 1979, Arvey & Campion, 1982; Harris, 1989; Keenan, 1989; Mayfield, 1964; Schmitt, 1976; Ulrich & Trumbo, 1964; Wagner, 1949; Wright, 1969) and the conclusions of these reviews, particularly with respect to interview validity, were not favourable. Meta-analysis is an advanced technique for statistically integrating results of independent studies to

obtain definitive conclusions in a research area. Unlike narrative reviews, this technique allows reviewers to estimate how much of the observed differences in results across studies is attributable to statistical artifacts rather than considerable differences in the underlying population relationships. Meta-analysis is more than a data-analysis technique, as it provides researchers with clearer directions regarding research needs within a particular domain. It can also be useful in the development of theory to identify what needs to be explained by a theory.

Essentially, meta-analysis provides estimates of the true average correlations and the true variation across studies, by calculating and removing variance attributable to artifacts. Hunter and Schmidt (1990) noted that there are several types of study and methodological artifacts that can alter the size of a study correlation to the extent that it actually underestimates the true correlation. Such statistical artifacts include sampling error, error of measurement in the independent and dependent variables, and range variation (ie., range restriction or enhancement).

The use of meta-analysis in estimating selection interview validity has contributed to a more lucid understanding of the factors that influence interview validity. For example, it is clear from previous meta-analytic reviews that structure is a powerful moderator of interview validity (Huffcutt & Arthur, 1994; Huffcutt & Woehr, 1993; McDaniel et al., 1994; Searcy et al. 1993; Wiesner & Cronshaw, 1988). While narrative reviewers have suspected that structure would influence interview validity (e.g; Keenan, 1989; Schmitt, 1976; Wagner, 1949; Wright, 1969), only the use of meta-analysis has confirmed such suspicions. Similarly, while many authors have concluded that panel interviews are, or at least should be, superior to interviews conducted by individuals (e.g., Campion, 1988; Dipboye, 1991; Lowry, 1994), the use of meta-analysis in the present review may contribute to an enhanced understanding of

interview validity. This can be achieved by determining the factors that not only affect interview validity, but are also correlated with whether the interview is conducted by an individual or a panel of interviewers.

A meta-analysis programme called MetaQuick 16, developed by Stauffer (1996) was used in the present study to analyse the proposed hypotheses. The programme is a psychometric meta-analysis programme based on the procedures developed by Hunter and Schmidt (1990) and Hedges and Olkin (1985).

Data sets were constructed according to the variable under consideration and the nature of the criterion. Hypotheses were tested by breaking the data set down by codings for a particular hypothesis, and comparing summary statistics. For example, the validities of panel and individual interviews with and without criterion contamination were analysed, by identifying the validity coefficients for those particular categories. For instance, all the coefficients which came from studies that were criterion contaminated individual interviews were analysed and then the result was compared to the result of criterion contaminated panel interview studies.

In all analyses, validity coefficients were corrected for sampling error and criterion unreliability. Following the reasoning of McDaniel et al. (1994), coefficients were not corrected for predictor unreliability, as the aim of this study was to obtain estimates of the operational validity of interviews for selection purposes. While range restriction data were available from McDaniel et al. (1994), coefficients were not corrected for range restriction due to MetaQuick 16's inability to use separate artifact distributions. As a result, the validities obtained without range restriction corrections must be regarded as lower bound, or downwardly biased, since range restriction attenuates or underestimates validity coefficients.

Artifact Information

As per the meta-analysis conducted by McDaniel et al. (1994), the studies used in the present analysis contained little information regarding the reliability of the job performance and training criteria. The present analysis thus uses the criterion reliability distributions used by Pearlman (1979) and later adopted by McDaniel et al. (1994), in which the average criterion reliabilities used for job performance and training performance were .60 and .80, respectively.

CHAPTER THREE

RESULTS

Overall Validity of Individual versus Panel Interviews

Analyses of the validities of individual and panel interviews were conducted for both job and training performance criteria to examine the extent to which individual and panel interviews differ in terms of validity, in the present sample. While McDaniel et al. (1994) conducted analyses using “tenure” as a performance criterion, this was not used in any of the present analyses, since the number of validity studies from which adequate conclusions could be drawn was too small (five studies in total). Comparisons with McDaniel et al.’s (1994) original findings were also made, but only with validities in which the criterion was job performance (as McDaniel et al did not compare the validities of individual and panel interviews when the criterion was training performance). The comparisons were performed in order to ensure that the results of the present study were not grossly conflicting with those obtained by McDaniel et al. Table 5 presents the results for individual and panel interviews for job and training performance criteria and show that panel interviews are equal in validity to individual interviews when the criterion was job performance, with both interviews yielding corrected mean validities of .28. However, when the criterion was training performance, panel interviews were inferior in terms of validity, achieving a mean corrected validity of .13 compared to

Table 5

Mean Validities for Panel and Individual Interviews

Interview Distribution	N	No. of r's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Criterion = Job Performance							
All interviews	21807 (25244)	136 (160)	0.22 (0.20)	0.15 (0.15)	0.28 (0.26)	0.20 (0.17)	-0.05
Individual Interviewer	11292 (11393)	65 (90)	0.22 (0.24)	0.18 (0.18)	0.28 (0.31)	0.23 (0.20)	-0.10
Panel Interview	7399 (11915)	47 (54)	0.22 (0.17)	0.11 (0.12)	0.28 (0.22)	0.14 (0.13)	0.05
Criterion = Training Performance							
All Interviews	58793	68	0.23	0.08	0.25	0.09	0.10
Individual Interviewer	51850	29	0.24	0.07	0.27	0.08	0.13
Panel Interview	5151	26	0.12	0.07	0.13	0.08	-0.01

Note: The first column of data in each table identifies the distribution of the validities analysed. The following four columns of data present the total sample size, the number of validity coefficients upon which each distribution was based, and the uncorrected mean and standard deviation of each distribution. Obs = observed; ρ = estimated population mean with corrections for sampling error and criterion unreliability; $\sigma\rho$ = estimated standard deviation. Values in parentheses refer to those obtained by McDaniel et al. (1994).

Table 6

Mean Validities for “Multiple Assessor” and Individual Interviews

Interview Distribution	N	No. of r's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Criterion = Job Performance							
Individual Interviewer	7749	70	0.28	0.21	0.36	0.26	-0.08
Multiple Assessor Interviews	12642	54	0.18	0.09	0.24	0.13	0.03
Criterion = Training Performance							
Individual Interviewer	10945	29	0.20	0.21	0.22	0.23	-0.16
Multiple Assessor Interviews	6055	18	0.22	0.16	0.24	0.18	-0.05

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation.

.26 for individual interviews. The results of the validities when the criterion was job performance conflict with those obtained by McDaniel et al., in which the corrected mean validities for individual and panel interviews were .31 and .22, respectively. These differences are attributable to two possible causes. Firstly, one possible cause for the differences could be that not all the validity coefficients used by McDaniel et al. (1994) could be obtained, and therefore it was not possible to include these in the present analyses. Secondly, some studies coded as individual interviews in the present study were coded as “board” interviews by McDaniel et al. In order to investigate which of these possible explanations accounted for the different results, analyses were performed using McDaniel et al.’s definition of panel and individual interviews (in which panel or “board” interviews encompassed all interviews in which there were multiple assessors, irrespective of whether all interviewers interviewed candidates simultaneously).

Table 6 presents the results of the validities for individual and multiple assessor interviews for job and training performance criteria. The results for interview validity when the criterion was job performance were similar to the original findings of McDaniel et al., such that the mean corrected validities for panel and individual interviews were .24 and .36, respectively, compared to McDaniel et al.’s findings of .22 for panel interviews and .31 for individual interviews. Therefore, the conflict between the results of the present study and those of McDaniel et al. can be attributed to the definition of “panel” interviews, rather than an incomplete data set.

The validity coefficients that were in the training criterion sample were also re-analysed using the “multiple assessor” definition. Interestingly, the mean corrected validities for individual and panel interviews were different, such that panel interview validity was *superior* to individual interviews (.24 versus .22, respectively). Although

these results can not be compared to McDaniel et al.'s (1994) (as no comparison between individual and panel interview validity was made for training performance criteria), these results illustrate that very different conclusions can be made when variables are operationalised in disparate ways.

Criterion Contamination

It was hypothesised in the present study that one explanation for the superior validity of individual interviews compared to panel interviews may be attributable to the presence of criterion contamination in individual interview studies. Criterion contamination is most likely to affect individual interview validity since individual interviewers are more likely to be the target position's manager, and thus are more likely to conduct performance appraisals for the target position. As a result of this potential bias, the corrected mean validity for individual interviews might be inflated, thereby leading to the false conclusion that individual interviews are superior to panel interviews in terms of validity. By removing criterion contaminated studies, I hypothesised that mean individual interview validity would be lower than panel interview validity.

Table 7 presents the results of analyses of individual and panel interviews with the omission of criterion contaminated studies when the criterion was job and training performance. Studies that could not be identified as either panel or individual interviews (twenty-four in total), were excluded from the criterion contamination analyses. A total of nine cases of criterion contamination were identified and subsequently removed. These contaminated cases were identified in

Table 7

Panel and Individual Interview Validity Without Criterion Contamination

Interview Distribution	N	No. of r's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Job Performance							
Individual Interviewer	11433	63	0.22	0.18	0.28	0.24	-0.11
Panel Interview	7256	45	0.22	0.11	0.28	0.14	0.05
Total	18689	108	0.22	0.16	0.28	0.20	-0.06
Training Performance							
Individual Interviewer	49890	27	0.24	0.07	0.27	0.08	0.13
Panel Interview	4602	23	0.11	0.06	0.12	0.08	-0.01
Total	54492	50	0.23	0.08	0.26	0.09	0.10

Note: Obs = observed; ρ = estimated population mean corrected for sampling error and criterion unreliability; $\sigma\rho$ = estimated standard deviation.

the studies performed by Arvey, Miller, Gould, and Burch (1987), Finesinger, Cobb Chapple, and Brazier (1948), Grove (1981), Johnson, Boots, Wherry, Hotaling, Martin, and Cassens (1994), Pasco (1979), Plag (1961), Roth, Campion, and Francis (1988), and Stohr-Gillmore, Stohr-Gillmore, and Kistler (1990).

When the criterion was job performance, there were two cases each of criterion contamination for individual and panel interviews, which when removed, made no difference to the mean corrected validities of individual and panel interviews from when criterion contamination was present. As such, the validities remained constant at .28. In addition, there were not more instances of criterion contamination in individual interview studies as hypothesised. Therefore, the hypothesis was not supported when the criterion was job performance.

A similar scenario occurred when the criterion was training performance. There were five cases of criterion contamination, two for individual interview studies and three for panel interview studies. When these cases of criterion contamination were removed, there was little difference in the mean corrected validities for individual and panel interviews, compared to when criterion contamination was present, such that the mean validities for individual and panel interviews were .12 and .27, respectively, compared to .26 for individual interviews and .13 for panel interviews when criterion contamination was present. However, the difference in the mean corrected validities for panel and individual interviews was large, suggesting that individual interviews are more valid than panel interviews when the criterion is training performance (irrespective of the existence of criterion contamination).

Overall, the criterion contamination hypothesis was inconclusive, to the extent that criterion contamination was not as prevalent in individual interview studies as

expected, and the removal of criterion contaminated studies did not alter the mean corrected validities of individual interviews.

Psychologists versus Non-Psychologists

The second hypothesis of the present study reasoned that individual interview studies were more likely to use psychologists as interviewers, and that psychologists would have higher validity. Tables 8 and 9 present the results for psychologists versus non-psychologists for individual and panel interview studies when the criterion was job and training performance, respectively. Studies that were not identified as individual or panel interviews were used in the calculation of the mean validities for “psychologists” and “non-psychologists” in total. As Tables 8 and 9 indicate, when the criterion was job performance, psychologists were more prevalent in individual interviews as hypothesised: twenty-two individual interview studies used at least one psychologist, while only three panel interview studies used psychologists as interviewers. While individual interviews used more psychologists than panel interviews, the difference between the number of panel and individual interview studies in which the interviewers were not psychologists was negligible (44 versus 43 studies respectively). When the criterion was training performance, a similar pattern emerged. Individual interview studies used psychologists as interviewers more than panel interview studies (11 versus 4), and non-psychologists were used as interviewers more often in panel interviews than in individual interview studies (22 versus 20 respectively). In sum, when interview studies used psychologists as interviewers, the interviews were more often individual interviews.

While the first half of the “psychologist versus non-psychologist” hypothesis was supported, the mean corrected validities obtained for both types of interviews

Table 8

Mean Validities for Interviewers as Psychologists (Criterion = Job Performance)

Interview Distribution	N	No. of r 's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Psychologist							
Individual Interviewer	6153	22	0.16	0.07	0.21	0.09	0.05
Panel Interview	167	3	0.22	0.13	0.28	0.17	0.01
Total	6580	29	0.16	0.07	0.21	0.10	0.05
Not Psychologist							
Individual Interviewer	5979	43	0.28	0.23	0.36	0.30	-0.14
Panel Interview	7232	44	0.22	0.11	0.28	0.14	0.05
Total	15227	107	0.24	0.17	0.31	0.22	-0.06

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation

Table 9

Mean Validities for Interviewers as Psychologists (Criterion = Training Performance)

Interview Distribution	N	No. of r 's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Psychologist							
Individual Interviewer	7688	10	0.29	0.13	0.32	0.15	0.08
Panel Interview	450	4	0.20	0.00	0.23	0.00	0.23
Total	8538	16	0.28	0.13	0.31	0.15	0.08
Not Psychologist							
Individual Interviewer	44162	19	0.23	0.05	0.26	0.06	0.16
Panel Interview	4701	22	0.11	0.07	0.12	0.08	-0.02
Total	50255	52	0.22	0.05	0.24	0.08	0.12

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation

under the respective performance criteria were somewhat surprising. Firstly, when interviewers were psychologists and the criterion was job performance, panel interview validity was superior to individual interview validity (.28 versus .21, respectively) as shown in Table 8. Secondly, individual interview validity was superior to panel interview validity when interviewers were identified as “non-psychologists”, achieving corrected mean validities of .36 compared to .28, respectively. Therefore, the second half of the “psychologist versus non-psychologist” hypothesis did not hold in this instance.

The hypothesis that individual interview validity would be higher than panel interview validity when psychologists were employed was supported when the criterion was training performance. Table 9 shows that the corrected mean validity for individual interviews was indeed superior to panel interview validity, such that the corrected mean validities were .32 and .23, respectively. Furthermore, individual interview validity was also higher than the validity for panel interviews when interviewers were “non-psychologists”. In this case, the corrected mean validities were .26 versus .12, respectively. In sum, the results suggest the following regarding the “psychologist versus non-psychologist” hypothesis. Firstly, psychologists are more prevalent in individual interviews for both performance criteria. Secondly, the results indicate that panel interview validity is superior to individual interview validity when psychologists are used and the criterion is training performance. Thirdly, panel interviews are inferior to individual interviews in terms of validity when non-psychologists are used as interviewers and the criterion is job or training performance. Finally, panel interviews are also inferior when interviewers are psychologists and the criterion is training performance.

Training/experience

The third hypothesis tested was that training/experience increases interview validity, and individual interviews are more likely than panel interviews to have training/experience. The results for studies using both job and training performance criteria, indicate that individual interviewers are more likely to have had training/experience than panel interviewers (as hypothesised), and individual interviews are more valid than panel interviews when interviewers were trained/experienced. However, individual interviews were less valid than panel interviews when the criterion was job performance and when interviewers were untrained/inexperienced. Tables 10 and 11 present the results of the analyses for both criterion types.

For both criterion types, there were more cases of trained/experienced interviewers in individual interviews than in panels, as hypothesised. When the criterion was job performance, there were 32 cases of trained/experienced interviewers in individual interview studies compared to only 14 cases in panel interview studies. Similarly, when the criterion was training performance there were more trained/experienced interviewers in individual than in panel interview studies (23 versus 15, respectively).

Individual interviewers were more valid than panel interviews when interviewers were trained/experienced for both performance criteria. When the criterion was job performance, the mean corrected validity for individual interviews was .44 compared to .31 for panels, as Table 10 shows. Similarly, the corrected mean validity was higher for individual interviews than panel interviews when the criterion was training performance (.35 versus .12, respectively). However, when interviewers were not trained/experienced, panel interviews yielded a higher

Table 10

Mean Validities of Interviews for Training/Experience (Criterion = Job Performance)

Interview Distribution	N	No. of r 's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Training/Experience							
Individual Interviewer	3938	32	0.34	.24	0.44	0.31	-0.07
Panel Interview	1570	14	0.24	0.01	0.31	0.02	0.28
Total	6411	56	0.30	0.20	0.39	0.26	-0.05
No Training/Experience							
Individual Interviewer	1208	3	0.14	0.14	0.18	0.18	-0.11
Panel Interview	1460	2	0.35	0.05	0.46	0.07	0.34
Total	2668	5	0.26	0.15	0.33	0.19	0.02

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation.

Table 11

Mean Validities of Interviews for Training/Experience (Criterion = Training Performance)

Interview Distribution	N	No. of r 's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Training/Experience							
Individual Interviewer	12835	23	0.31	.12	0.35	0.14	0.12
Panel Interview	3837	15	0.11	0.09	0.12	0.10	-0.04
Total	17730	44	0.26	0.14	0.29	0.16	0.02

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation. No studies were available for "no training/experience"

corrected mean validity than individual interviews (.46 versus .18, respectively). job performance. (There was no validity data available in reference to interviewer training/experience for the training performance criterion.) Overall, trained/experienced interviewers were more prevalent in individual interview studies, and individual interview validity was superior to panel interview validity for both performance criteria, when interviewers were trained/experienced. However, panel interviews were superior in validity to individual interviews when interviewers were not trained/experienced. Therefore, there was partial support for the "training/experience" hypothesis.

Interviewer Proximity

This hypothesis stated that the proximity of interviewer(s) to the target position would correlate positively with interview validity, and that individual interviewers were likely to be more proximal than panel interviewers. The first half of the proximity hypothesis was analysed using Spearman's rho since the proximity data was ordinal (low = 1, medium = 2, and high = 3). The second part of the "proximity" hypothesis was analysed by identifying the median proximity rating for interviewers in both individual and panel interviews. Consistent with other analyses in the present study, analyses are presented for both job and training performance criteria.

The results of the Spearman's rho correlations between interviewer proximity and interview validity were -0.16 (n = 80) and -0.03 (n = 49) for job and training performance criteria, respectively, both of which were insignificant. In spite of this finding, the interviewer proximity issue was pursued further by investigating the relationship between proximity and validity in conjunction with interview structure, based on the following logic. It would follow that interviewer proximity would not

affect interview validity when the interview is structured, as all interviewers, irrespective of how proximal they may be to the target position, would conduct the interview using a similar format. Accordingly, proximity might be expected to have a greater influence on interview validity when the interview is *unstructured*, as interviewers who are more proximal to the target position would have greater knowledge of the requirements for the target position, and would therefore make more valid selection decisions. This post hoc hypothesis was analysed using Spearman's rho. The results for the correlations between interview validity and interviewer proximity and structure were also inconclusive. When the criterion was job performance, the Spearman's correlation between interview validity and proximity when interviews were structured was .17 ($n = 80$) and $-.32$ ($n = 20$) when the criterion was training performance. Both correlations were not significant at the 0.05 level of significance. This supports the hypothesis that proximity is not important when interviews are structured. When the interviews were unstructured, the correlation between proximity and interview validity was $-.30$ ($n = 38$) and 0.13 ($n = 20$) when the criteria were job and training performance, respectively. Again, these correlations were not statistically significant. Therefore, there appears to be little relationship between interviewer proximity and interview validity when interviews are structured or unstructured.

The second part of the "proximity" hypothesis postulated that individual interviewers were likely to be more proximal than panel interviewers. This was analysed by calculating the median proximity for interviewers for both job and training performance criteria. The results indicated that this was the case. The median proximity for individual interviewers was 1.0 and 2.0 for job and training performance, respectively, compared to 0 and 1.0, respectively, for panel

interviewers. In sum, the results of the analyses of the “proximity” hypothesis, indicate that proximity is not related to interview validity, although individual interviewers are more proximal to target positions than panel interviewers.

Panel Size

The present study investigated the effect panel size would have on interview validity. The results, presented in Table 12, show that panel interview validity increases with panel size. When the criterion was job performance, the corrected mean validity for two member panels was .22 compared to the corrected mean validity of .44 when the number of panellists was five. The trend of increased mean corrected validity with increased panel size was also evident when the criterion was training performance. Table 12 shows that interview panels with four members were more valid than those with only two members. The results show that panel interview validity does increase as a function of panel size, and larger panels were higher than individual interviews in terms of predictive validity, particularly when the criterion was job performance.

Table 12

Summary of Mean Validities for Different Panel Sizes

Panel Size	N	No. of r 's	Mean r	Obs. σ	ρ	$\sigma\rho$	90% Credibility Value
Job Performance							
2 members	4360	25	0.17	0.08	0.22	0.10	0.05
3 members	1182	15	0.24	0.07	0.31	0.08	0.17
4 members	-†						
5 members	1548	4	0.34	0.11	0.44	0.11	0.21
Training Performance							
2 members	3537	10	0.09	0.09	0.11	0.10	-0.06
3 members	716	8	0.15	0.00	0.17	0.00	0.17
4 members	187	2	0.18	0.00	0.20	0.00	0.20
5 members	-†	-	-	-	-	-	-

Note. Obs = observed; ρ = estimated population mean; $\sigma\rho$ = estimated standard deviation. †No validity studies were available

CHAPTER 4

DISCUSSION

The aims of the present study were (a) to analyse the possible causes for the inferiority of panel interview validity compared to individual interview validity, and (b) to investigate whether all panel sizes were inferior to individual interviews in terms of predictive validity. In the present study, the overall findings indicated that individual interviews were equal to panel interviews in terms of validity when the criterion was job performance, although superior to panel interview validity when the criterion was training performance. Moreover, panel sizes of more than three members were more valid than individual interviews, when the criterion was job performance. The overall validity of individual and panel interviews and the outcomes the proposed hypotheses are discussed in turn below.

Overall Validity of Individual versus Panel Interviews

The validity of individual interviews was found to be superior to panel interviews for training performance criteria, and equal to panel interview validity when the criterion was job performance. When the McDaniel et al. (1994) definition of panel interviews was applied, and interviews were redefined as multiple assessor interviews, individual interviews were superior in predictive validity compared to panel interviews when the criterion was job performance, yet were inferior when the criterion was training performance. The results obtained after reclassifying the parameters of “panel” interviews, illustrate that varying results are likely to be

obtained when variables are not operationalised in a consistent fashion. In sum, the results yielded in the present study are only partially consistent with previous findings, such that individual interviews were superior to panel interviews in terms of validity, but only in the case of predicting training success (McDaniel et al., 1994; Searcy et al., 1993; Wiesner & Cronshaw, 1988).

Criterion Contamination

The present study hypothesised that individual interview studies were more likely than panel interview studies to suffer from criterion contamination. This was based upon the reasoning that individual interviewers were more likely to be the supervisor or manager of the target position, and might therefore be involved in assessments of performance, which would constitute criterion contamination. If this was the case, then individual interview validity would, in fact, be artificially inflated giving rise to the assertion that individual interviews were more valid. This hypothesis was not supported as individual interview studies had no more cases of criterion contamination than panel interviews for both performance criteria, and the number of criterion contaminated cases was minimal.

In addition, the criterion contamination hypothesis also stated that, once contaminated studies were omitted from comparisons of panel and individual interview validity, individual interview validity would be no higher than panel interview validity. However, in spite of the removal of criterion contaminated studies from analyses, validities for both panel and individual interviews remained unchanged from when criterion contamination was present.

It was interesting to discover that criterion contamination was not as prevalent as originally hypothesised. In fact, there were only nine cases of such bias in total. So

why might criterion contamination not be prevalent in the selection interview validation studies? Firstly, not all studies could be identified as to whether subjective appraisals of performance had occurred, and there were a number of cases in which it was not possible to identify how the interview was conducted (ie., individual versus panel formats). This was largely due to deficient reporting in many of the validity studies used in the present analysis. The results of the proximity hypothesis can also aid the interpretation of the criterion contamination result. Interviewers were generally not high in proximity to the target position, indicating that it was unlikely that interviewers were also raters of performance, thus eliminating the possibility of criterion contamination occurring. Furthermore, it is questionable as to how representative the validation studies used in the present study are of selection interviews in the field. It is possible that in many situations individual interviewers are indeed the position manager. Therefore, criterion contamination may be rife in practice. In conclusion, criterion contamination could not be attributed to explaining the superiority of individual interview validity in the present study.

Psychologists versus Non-Psychologists

In the present study, it was postulated that psychologists would increase interview validity, since psychologists would have more interviewing expertise than non-psychologists, and because evidence from the assessment center literature has indicated that psychologists made more valid ratings than non-psychologists. Furthermore, psychologists were predicted to be more prevalent in individual interviews than in panel interviews.

The findings indicated that there was a greater prevalence of psychologists in individual interviews. Secondly, individual interviews were more valid than panel

interviews when psychologists were used as interviewers, but only when the criterion was training performance, however when the criterion was job performance, panel interviews were more valid. So, in fact, individual interviewers who were psychologists decreased interview validity when the criterion was job performance. This finding contradicts the conclusions of Gaugler et al. (1987), who noted that psychologists made more valid assessments than managers in assessment centers. Also, individual interview validity was higher when interviewers were non-psychologists and when the criterion was job performance, which contradicts the conclusions made by Chi et al. (1992) that experts make more valid judgements. Interestingly, Camerer and Johnson (1991) noted that while experts were superior to novices in processing information, evidence showed that experts did not show superior performance. It is also possible that psychologists may not really be “experts” about the target position, nor may they be experts in interviewing. In conclusion, interviewers as psychologists only explained the superiority of individual interview validity when the criterion was training performance.

Training/Experience

Past research on interviewer training suggested that trained interviewers can increase the validity and reliability of ratings made (e.g., Conway et al., 1995; Dougherty et al., 1986). The present study hypothesised that trained/experienced interviewers were more likely to prevail in studies of individual interviews than in panel interview studies. Individual interview validity was subsequently hypothesised to be superior in validity to panel interviews. The findings of the present study found that there were more trained/experienced interviewers in individual studies than in panel interview studies, and individual interview validity was subsequently higher than

panel interview validity (for both job and training performance criteria), as hypothesised. In addition, panel interview validity was superior to individual interviews when interviewers were untrained/inexperienced when the criterion was job performance. This result was based on only three validity coefficients, and thus might suffer from second-order sampling error and should therefore be treated with caution. This issue is discussed in more detail in the limitations section of this chapter. Therefore, one factor which may explain why individual interviews were more valid than panel interviews, is that individual interviewers were more likely to be trained/experienced than panel interviewers. As such, training/experience or *skills* in interviewing contributes to increased interview validity.

Interviewer Proximity

The results of correlational analyses between interview validity and interviewer proximity were inconclusive, such that correlational analyses failed to demonstrate any relationship between the two variables. Additional analyses between interviewer proximity and interview structure and validity also yielded no significant relationship. One explanation for the insignificant findings regarding interviewer proximity may be that proximity does not connote effectiveness. An interviewer who possesses a great deal of knowledge about a particular position, may not be effective when required to extract relevant information from applicants, in terms of their ability to perform well on the job. Another explanation may be that the interviewers used in the studies analysed in the present analyses do not accurately reflect the positions of interviewers in practice. It could be that in practice, interviewers *are* more proximal to the target position, and perhaps the impact on interview validity is greater in the field. In conclusion, the present findings indicate that the proximity of the interviewer(s) to the

target position had no effect on interview validity, although individual interviewers were more proximal than panel interviewers.

Panel Size

In the present study the following question was posed: Were all panel sizes inferior to individual interviews in validity? The group literature has suggested that group performance may increase as a function of size, and was supported in the present study. The findings in the present study indicate that groups of four to five members are ideal, however only small samples of coefficients for panel interviews were available, which weakens validity generalisation somewhat. Moreover, there were no instances of panel interviews consisting of more than five members, and therefore comparisons with larger groups were not feasible. Nevertheless, while individual interviews were more valid than panel interviews, when interview validity was compared between individuals and varying interview panel sizes, larger interview panels yielded higher validity than individual interviews.

In sum, the possible causes for the inferiority of panel interview validity might be attributed to the greater prevalence of interviewer training/experience in individual interviews than in panel interviews. Small panel sizes also appear to contribute to lower interview validity compared to individual interviews. Criterion contamination, interviewers as psychologists and interviewer proximity could not explain the superior validity of individual interviews.

Limitations

A number of limitations existed in the present study and require discussion. Firstly, the small number of studies in many of the analyses performed may give rise to

second-order sampling error. According to Hunter and Schmidt (1990), second-order sampling error occurs when meta-analysis is based on a small number of studies and the outcome depends partly on study properties that vary randomly across studies. The effect of second-order sampling error is more problematic for estimates of standard deviations than for estimates of means, such that estimates can be distorted. Therefore, the results for "no training/experience" and panel size should be treated somewhat cautiously.

Secondly, the estimates of interview validity were not corrected for range restriction. The meta-analysis statistical package used to perform the analyses was unable to conduct analyses using separate artifact distributions, and so the results should be considered as lower-bound estimates of validity. While the program could perform analyses using artifact information, most of the validity studies used did not report separate artifact information for range variation, and therefore it was not possible to include such information. Nevertheless, the results indicated moderate validity for the employment interview, and it would be likely that these estimates of validity would be higher when range restriction was accounted for.

The reporting of methodological information in the validity studies used in the present study was generally poor. Often there was little information as to who performed interviews, how many people conducted the interviews, whether interviewers were trained/experienced, and it was often not made clear whether interviewers were also involved in performance appraisals. This limitation is by no means unique to the present study. However, it did serve to undermine the robustness of some validity estimates in the present study.

While every endeavour was made to obtain all the validity information used by McDaniel et al. (1994), it was not possible, and therefore it may not be entirely

appropriate to directly compare the present results with those of McDaniel et al. (1994). However, while only thirty-one coefficients were not located, the results obtained were similar, and therefore the concern of lack of comparability is minimal.

Another limitation of the present study resided in the treatment of interviewer training and experience as the same construct. The purpose of the merging of training and experience was purely to differentiate between interviewers with skills and those without interviewing skills to ascertain the differences in validity, if any. However, this may not be methodologically and conceptually correct. Training and experience are two quite different phenomena. Training refers to instructing interviewers on how to conduct an interview, how to minimise bias, and how to effectively obtain information from applicants so that an effective interview decision can be made. Experience, on the other hand, refers more to familiarity with conducting interviews, or that an interviewer has simply conducted a lot of interviews. Experience does not necessarily imply effectiveness.

Moreover, neither training nor experience have been considered as one and the same in interview research. Research has tended to focus on one or the other. Research evidence regarding interviewer experience and validity is nonexistent (unlike interviewer training and validity), with research merely focusing on experience and level of bias affecting the decisions. For example, Keenan (1978) noted that experienced interviewers were more biased than less experienced interviewers, while Marlowe, Schneider and Nelson (1996) found that less experienced managers were more biased in their hiring decisions than more experienced managers. In relation to training and validity, there is some research evidence to suggest that training can increase the validity of employment interviews. While the present results suggested that the superior validity for individual interviews was attributable to

trained/experienced interviewers, it was not clear as to whether training was more effective than experience or vice versa. Future research could explore the concepts of training and experience in order to improve our understanding of what constitutes effective interviewers.

Recommendations for Future Research

The present meta-analytic review of employment interviews focused on investigating possible causes of inferior panel interview validity compared to individual interviews. While the overall conclusion was that panel interviews were generally no better and even worse in terms of predictive validity, panel interviews were found to be superior in validity in some cases. Nevertheless, there remain a number of issues that may only be resolved as a result of further primary research. Firstly, studies are needed to ascertain what constitutes an effective interviewing panel, whether this be in terms of size, characteristics of interviewers, such as degree of training, level of experience, and processes which prevail within the interviewing group. Obviously, not all panels are equally effective, therefore, future research might investigate this. This would require indepth analysis of the group dynamics and processes that are in operation, and what size groups are more effective than others. Previous authors of panel interviews have randomly asserted that the ideal interview panel size ranges anywhere between two and five (e.g., Campion et al., 1988; Lowry, 1994), but without any empirical evidence to substantiate such claims. While the present research showed that larger panel sizes were not only more valid than smaller interview groups, but also more valid than individual interviews, there were no data for larger panel sizes to ascertain whether a diminishing effect on performance existed with larger interview

groups. Future research might investigate the ideal panel size by comparing varying sizes of interview panels.

Secondly, future research might address the issue of under what circumstances individual and panel interviews are more appropriate, by investigating occupation type and interview type. Individual interviews may be more valid in circumstances where a candidate is being assessed for an entry-level position or a position requiring few skills. Panel interviews on the other hand, may be more appropriate at executive levels where candidates might be expected to perform comfortably in such a situation. Indeed, a survey of selection practices of organisations in the United Kingdom indicate that individual interviews are commonly used for clerical/secretarial positions and manual/craftworker jobs in 19.9% and 23.7% of organisations surveyed, respectively. In addition, panel interviews are extremely common for graduate, managerial and professional/technical positions (Industrial Relations Services, 1997), although no validity information was available. Furthermore, applicant reactions to individual and panel interviews may also need to be addressed. It may be that individual interviews are more valid because applicants perform better at such interviews. The employment interview is not only a unique experience, but also one that is inherently stressful. Therefore candidates may feel more at ease in a one-on-one situation, and subsequently behave more naturally than candidates facing a group of five unfamiliar faces.

Finally, there has been virtually no research in the selection interview literature directly comparing individual and panel interviews. If the issue of individual versus group is one of importance, why has this not been done? This issue might be best addressed in conjunction with research investigating interview panel dynamics, processes, and size, by directly comparing individual interviewers with interview

panels of varying sizes. Furthermore, such research should not be limited to simply a comparison of methods of interviewing. The reactions and processes of applicants and interviewers should also be incorporated into such research.

Conclusions

The results of the present meta-analysis did not support previous findings that panel interviews are inferior to individual interviews in predictive validity in the case of predicting job performance. However, support was found for panel interview validity inferiority when the criterion was training performance. Panel interviews were superior in validity when (a) large panel groups were utilised, and (b) interviewers were not trained/experienced. However, a number of possible explanations were not addressed in the present study which might be the subject of future primary research. In sum, the panel interview should not be considered as an inferior selection technique until the processes and dynamics of interview panels are further explored.

REFERENCES

- Albrecht, P. A., Glaser, G. M., & Marks, J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology, 48*, 351-360.
- Anstey, E. (1977). A 30-year-follow-up of the CSSB procedure, with lessons for the future. *Journal of Occupational Psychology, 50*, 149-159.
- Anstey, E., & Mercer, E. O. (1956). *Interviewing for the selection of staff*. London: Allen & Unwin.
- Ard, R. F. (1985, January 30). Personal communication to M. A. McDaniel.
- Arvey, R. D. (1979). Unfair discrimination in the employment interview. *Psychological Bulletin, 86*, 736-765.
- Arvey R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35*, 281-322.
- Arvey, R.D., Miller, H. E., Gould, R., & Burch, P. (1987). Interview validity for selecting sales clerks. *Personnel Psychology, 40*, 1-12.
- Berkley, S. (1984). *VII Validation Report Corrections Officer Trainee*. Harrisburg: Commonwealth of Pennsylvania, State Civil Service Commission.
- Bottger, P. C., & Yetton, P. W. (1988). An integration of process and decision scheme explanations of group problem solving performance. *Organizational Behavior and Human Decision Processes, 42*, 234-249.
- Bosshardt, M. J. (1993, April 8). Personal communication to M. A. McDaniel.

- Brown, S. H. (1986). Personal communication to M. A. McDaniel.
- Camerer, C. F. & Johnson, E. J. (1991). The process-performance paradox in expert judgement: How can experts know so much and predict so badly? In K. A. Ericsson and J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 195-217). NY: Cambridge University Press.
- Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing, raising the psychometric properties of the employment interview. *Personnel Psychology, 41*, 25-42.
- Cascio, W. F. (1987). *Applied psychology in personnel management* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Daniel, C. & Valencia, S. (1991). Structured interviewing simplified. *Public Personnel Management, 20*(2), 127-133.
- Davis, R. (1986). Personal communication to M. A. McDaniel.
- Delaney, E. C. (1954). Teacher selection and evaluation with special attention to the validity of the personal interview and the National Teacher Examinations as used in one selected community (Elizabeth, New Jersey). (Doctoral dissertation, Columbia University, 1954). *Dissertation Abstracts International, 14*, 1334-1335.

- Devine, D. J., & Kozlowski, S. W. J. (1995). Domain-specific knowledge and task characteristics in decision making. *Organizational Behavior and Human Decision Processes*, 64 (3), 294-306.
- Dipboye, R. L. (1991). *Selection interviews: Process perspectives*. Cincinnati: Southwestern Publishing Co.
- Dipboye, R. L., & Gaugler, B. B. (1993). Cognitive and behavioral processes in the selection interview. In N. Schmitt, W. C. Borman, & Associates (Eds.), *Personnel Selection in Organisations*, pp. 135-170. New York: Jossey-Bass.
- Dougherty, T. W., Ebert, R. J., & Callender, J. G. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology*, 71, 9-15.
- Drake, L. R., Kaplan, H. R., & Stone, R. A. (1973). Organizational performance as a function of recruitment criteria and effectiveness. *Personnel Journal*, 52, 885-891.
- Dreher, G. F., Ash, R. A., & Hancock, P. (1988). The role of the traditional research design in understanding the validity of the employment interview. *Personnel Psychology*, 41, 315-327.
- Dreher, G. F. & Maurer, S. D. (1993). Assessing the employment interview: Deficiencies associated with the existing domain of validity coefficients. In R. W. Eder and G. R. Ferris (Eds.), *The Employment Interview: Theory, Research and Practice* (pp. 249-268). Newbury Park, CA: Sage.
- Dunnette, M. D., Arvey, R. D., & Arnold, J. A. (1971). *Validity study results for jobs relevant to the petroleum refining industry*. Minneapolis, MN: Personnel Decisions.

- English, J. J. (1983). *The relationship between structured selection criteria and the assessment of proficient teaching*. Unpublished doctoral dissertation, University of Virginia, Richmond.
- Finesinger, J. E., Cobb, S., Chapple, E. D., & Brazier, M. A. B. (1948). *An investigation of prediction success in naval flight training* (Rep. No. 81). Washington, DC: Civil Aeronautics Administration, Division of Research.
- Fisher, C. D., Schoenfeldt, L. F., & Shaw, J. B. (1993). *Human resource management* (2nd ed.). Boston, MA: Houghton Mifflin Company.
- Flynn, J. T. & Peterson, M. (1972). The use of regression analysis in police patrolman selection. *The Journal of Criminal Law, Criminology, and Police Science*, 63, 564-569.
- Gael, S. (1983). *Job analysis: A guide to assessing work activities*. San Francisco: Jossey-Bass.
- Gatewood, R. D., & Feild, H. S. (1994). *Human resource selection* (3rd ed.). London: Dryden Press.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III., & Bentson, C. (1987). Meta-analysis of assessment center validity: Monograph. *Journal of Applied Psychology*, 72 (3), 493-511.
- Grove, D. A. (1981). A behavioral consistency approach to decision making in employment selection. *Personnel Psychology*, 34, 55-64.
- Guion, R. M. (1965). *Personnel testing*. New York: McGraw-Hill.

- Guion, R. M. (1976). Recruiting, selection, and job placement. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*, p. 77-82. Chicago: Rand McNally.
- Harris, M. M. (1989). Reconsidering the employment interview: A review of the literature and suggestions for future research. *Personnel Psychology*, 42, 691-726.
- Harris, N. J., Toulson, P. K., & Livingston, E. M. (1996). New Zealand personnel consultants and the selection process. *Asia Pacific Journal of Human Resources*, 34(4), 71-87.
- Harvey, R. J. (1990). Job analysis. In M. D. Dunnette and L. M. Hough (Eds), *Handbook of industrial and organizational psychology* (Vol 2, 71-163). Palo Alto, CA: Consulting Psychologists Press.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management*, 32(2 & 3), 325-352.
- Hedges, L. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Heneman, H. G. (1975). The impact of interviewer training and interview structure on the reliability and validity of the selection interview. *Proceedings of Academy of Management*, 231-233.
- Hollingsworth, H. L. (1922). *Judging human character*. New York: Appleton.
- Huffcutt, A. & Woehr, D. J. (1993). *A meta-analytic review of the relationship between employment interview validity and degree of structure*. Paper presented at the eighth annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

- Huffcutt, A. & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry level jobs. *Journal of Applied Psychology, 79*(2), 184-190.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research across studies*. Beverley Hills, CA: Sage.
- Industrial Relations Services (1997). The state of selection: An IRS survey. *Employee Development Bulletin, 85*, 8-18.
- Johnson, H. M., Boots, M. L., Wherry, R. J., Hotaling, O. C., Martin, L. G., & Cassens, F. P., Jr. (1994). *On the actual and potential value of biographical information as a means of predicting success in aeronautical training* (Rep. No. 32). Washington DC: Civil Aeronautics Administration, Airman Development Division.
- Keenan, A. (1978). Selection interview outcomes in relation to interviewer training and experience. *Journal of Social Psychology, 106*, 249-260.
- Keenan, T. (1989). Selection interviewing. In C. L. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 1-23). John Wiley & Sons Ltd.
- Kennedy, R. (1986). *An investigation of criterion-related validity for the structured interview*. Unpublished master's thesis, East Carolina University, Greenville, NC.

Landy, F. J. (1976). The validity of the interview in police officer selection. *Journal of Applied Psychology*, 2, 193-198.

Landy, & Farr, (1983). *The measurement of work performance*. London: Academic Press.

Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimise rating errors in the observation of behavior. *Journal of Applied Psychology*, 60, 550-555.

Littlepage, G. E. (1991). Effects of group size and task characteristics on group performance: A test of Steiner's model. *Personality and Social Psychology Bulletin*, 17(4), 449-456.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, 32(2 & 3), 353 - 372.

Lowry, P. E. (1994). The structured interview: An alternative to the assessment center? *Public Personnel Management*, 23(2), 201-215.

Maier, N. R. (1967). Assets and liabilities in group problem solving: The need for an integrative function. *Psychological Review*, 74(4), 239-249.

Marlowe, C. N., 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(1), 11-21.

Martell, R. F., & Borg, M. R. (1993). A comparison of the behavioral rating accuracy of groups and individuals. *Journal of Applied Psychology*, 78(1), 43-50.

- Maurer, S. (1986). Personal communication to M. A. McDaniel.
- 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., Brown S. H., & Hamstra, B. W. (1980). Selective interviewing in the life insurance industry: An update of research and practice. *Personnel Psychology, 33*, 725-739.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*(4), 599-616.
- McKinney, T. S. (1975). *The criterion-related validity of entry level police officer selection procedures*. Phoenix, AZ: City of Phoenix, Personnel Department, Employee Selection Research.
- Miner, F. C. (1984). Group versus individual decision-making: An investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behavior and Human Performance, 33*, 112-124.
- Miner, J. B. (1970). Psychological evaluations as predictors of consulting success. *Personnel Psychology, 23*, 393-405.
- Moses, J., Hollenbeck, G. P., & Sorcher, M. (1993). Other people's expectations. *Human Resource Management, 32*(2 & 3), 283-297.
- Muchinsky, P. M. (1990). *Psychology applied to work: An introduction to Industrial and Organizational Psychology* (3rd ed.). Belmont, CA: Brooks/Cole.

- Pasco, D. C. (1979). The comparison of two interviewing techniques, the leaderless group discussion and the traditional interview, as a method for teacher selection (Doctoral dissertation, University of Maryland, 1979). *Dissertation Abstracts International*, 41, 486.
- Pearlman, K. (1978). Personal communication to M. A. McDaniel.
- Pearlman, K. (1979). *The validity of tests used to select clerical personnel: A comprehensive summary and evaluation* (TS-79-1). Washington DC: U. S. Office of Personnel Management, Personnel Research and Development Center. (NTIS No. PB 80.102650)
- Plag, J. A. (1961). Some considerations of the value of the psychiatric screening interview. *Journal of Clinical Psychology*, 17, 3-8.
- Pursell, E. D., Campion, M. A., & Gaylord, S. R. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal*, 59, 907-912.
- Reilly, R. A., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Reynolds, A. H. (1979). The reliability of a scored oral interview for police officers. *Public Personnel Management*, 8, 324-328.
- Robbins, S. P. (1993). *Organizational behavior: Concepts, theories, and applications* (6th ed.). Prentice Hall.
- Robertson, I. T., & Makin, P. J. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology*, 59, 45-57.
- Roth, P. L., & Campion, J. E. (1992). An analysis of the predictive power of the panel interview and pre-employment tests. *Journal of Occupational and Organizational Psychology*, 65, 51-60.

- Roth, P. L., Campion, J. E., & Francis, D. J. (1988, April). *A LISREL analysis of the predictive power of pre-employment tests and the panel interview*. Paper presented at the third annual meeting of the Society for Industrial and Organizational Psychology, Dallas, Texas. (Validity data not reported in the paper were obtained from P. L. Roth.).
- Ryan, A. M., & Sackett, P. R. (1987). A survey of individual assessment practices by I/O psychologists. *Personnel Psychology*, *40*(3), 455-488.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology*, *29*, 79-101.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 407-422.
- Searcy, C. A., Woods, P. N., Gatewood, R., & Lance, C. (1993). *The validity of structured interviews: A meta-analytical search for moderators*. Paper presented at the Society of Industrial Psychologists 1993 annual meeting. San Francisco, CA.
- Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, *113*(1), 67-81.
- Small, B. H. (1996). Variation in interview validity: Individuals, consensus panels, and statistical panels. Unpublished Masters Thesis, University of Waikato.
- Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology*, *53*, 81-93.

- Stohr-Gillmore, M. K., Stohr-Gillmore, M. W., & Kistler, N. (1990). Improving selection outcomes with the use of situational interviews: Empirical evidence from a study of correctional officers for new generation jails. *Review of Public Personnel Administration, 10*, 1-18.
- Sulsky, L. M. & Day, D. V (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology, 77*, 501-510.
- Taylor, P., Mills, A., & O'Driscoll, M. (1993). Personnel selection methods used by New Zealand organisations and personnel consulting firms. *New Zealand Journal of Psychology, 22*, 19-31.
- Taylor, D. W., & Faust, W L. (1952). Twenty questions: Efficiency in problem solving as a function of size of group. *Journal of Experimental Psychology, 44*, 360-368.
- Tornow, W W (1993). Editor's note: Introduction to special issue on 360-degree feedback. *Human Resource Management, 32(2 & 3)*, 211-220
- Tziner, A., & Dolan, S. (1982b). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology, 67*, 728-736.
- Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63(2)*, 100-116.
- Vance, R. J., Kuhnert, K. W., & Farr, J. L. (1978). Interview judgements: Using external criteria to complete behavioral and graphic ratings. *Organizational Behavior and Human Performance, 22*, 279-294.
- Wagner, R. (1949). The employment interview: A critical summary. *Personnel Psychology, 2*, 17-46.

- Wayne County Civil Service Commission. (1973). *Summary information regarding validity of the examination for patrolmen in the Wayne County Sheriff's Department*. Wayne County Civil Service Commission.
- 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.
- Wexley, K. M., Sanders, R. E., & Yukl, G. A. (1973). Training interviewers to eliminate contrast effects in employment interviews. *Journal of Applied Psychology*, 57, 233-236.
- Wiesner, W. H., & Cronshaw, S. F. (1988). The moderating impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275-290.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications, Inc.
- Worchel, S., Cooper, J., & Goethals, G. R. (1990). *Understanding social psychology* (5th ed.). Belmont, CA: Brooks/Cole.
- Wright, O. R. (1969). Summary of research on the selection interview since 1964. *Personnel Psychology*, 22, 391-413.
- Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology*, 62, 191-199.
- Zaranek, R. J. (1983). *A correlational analysis between the teacher perceiver interview and teacher success in the Chippewa Valley school system*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.

APPENDIX A

SOURCES OF VALIDITY DATA

SOURCES OF VALIDITY DATA

- Albrecht, P. A., Glaser, G. M., & Marks, J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology, 48*, 351-360.
- Aramburo, D. J. (1981). *The efficacy of the structured interview in the selection of special education teachers*. Unpublished doctoral dissertation, University of New Orleans, LA.
- Arvey, R.D., Miller, H. E., Gould, R., & Burch, P. (1987). Interview validity for selecting sales clerks. *Personnel Psychology, 40*, 1-12.
- Banta, G. W. (1967). A comparison of leaderless group discussion and the individual techniques for the selection of orientation assistants. *Dissertation Abstracts International, 28*, 937.
- Barrett, G. V., Svetlick, B., & Prien, E. P. (1966). Validity of the job-concept interview in an industrial setting. *Journal of Applied Psychology, 51*, 233-235.
- Bartleet, N. R. (1950). Review of research and development in examination for aptitude for submarine training, 1942-1945. *Medical Research Laboratory Report 153(9)*, 11-53.
- Bender, W. R. G., & Loveless, H. E. (1958). Validation studies involving successive classes of trainee stenographers. *Personnel Psychology, 11*, 491-508.
- Benz, M. P. (1974). *Validation of the examination for Staff Nurse II*. Urbana: University Civil Service Testing Program of Illinois, Testing Research Program.

- Bertram, F. D. (1975). *The prediction of police academy performance and on the job performance from police recruit screening measures*. Unpublished doctoral dissertation, Marquette University, Milwaukee, WI.
- Bloom, R. F., & Brundage, E. G. (1947). Prediction of success in elementary schools for enlisted personnel. In D. B. Stuit (Ed.), *Personnel research and test development in the Bureau of Naval Personnel* (pp. 251-254). Princeton, NJ: Princeton University Press.
- Bobbitt, J. M., & Newman, S. H. (1944). Psychological activities at the United States Coast Guard Academy. *Psychological Bulletin*, *41*, 568-579.
- Bolanovich, D. J. (1944). Selection of female engineering trainees. *Journal of Educational Psychology*, *35*, 545-553.
- Bonneau, L. R. (1957). An interview for selecting teachers (Doctoral dissertation, University of Nebraska, 1956). *Dissertation Abstracts International*, *17*, 537-538.
- Borman, W. C. (1982). Validity of behavioural assessment for predicting military recruiter performance. *Journal of Applied Psychology*, *67*, 3-9.
- Campbell, J. T., Prien, E. P., & Brailey, L. G. (1960). Predicting performance evaluations. *Personnel Psychology*, *13*, 435-440.
- Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology*, *41*, 25-42.
- Carlstrom, A. (1984). *Correlations between selection ratings and success as army aviators* (FOA Rep. C 53020-H2). Linkoping, Sweden: National Defence Research Institute.

- Cerf, A. Z. (1947). Conference for the interpretation of test scores and occupational background CE707A. In J. P. Guilford & J. I. Lacey (Eds.), *Printed classification tests: Report No. 5* (pp. 652-656). Washington, DC: Army Air Forces Aviation Psychology Program.
- Conrad, H. S., & Satter, G. A. (1946). *The use of test scores and quality classification ratings in predicting success in electrician's mates school* (Office of Scientific Research and Development Rep. No. 5667, 1945: Publications Board No. 132900). Washington, DC: U.S. Department of Commerce.
- Cook, R. E. (1981). *Chippewa Valley local validation study of the teacher perceiver interview process*. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.
- Darany, T. (1971). *Summary of state police trooper 07 validity study*: Ann Arbor: Michigan Department of Civil Service.
- Davey, B. (1984, May). *Are all oral panels created equal?: A study of differential validity across oral panels*. Paper presented at the International Personnel Management Association Assessment Council Annual Conference, Seattle, WA.
- DeHart, W. A. (n. d.). *Preliminary report on research findings of a demonstration project in recruitment and training of case workers sponsored by the Utah State Department of Public Welfare, 1966-1967*. Salt Lake City: Utah State Department of Public Welfare.
- Delery, J. E., Wright, P. M., & Tolzman, K. (1992, April). *Employment tests and the situational interview: A test of incremental validity*. Paper presented at the seventh annual meeting of the Society for Industrial and Organizational Psychology, Montreal, Quebec, Canada.

- Dipboye, R. L., Gaugler, B., & Hayes, T. (1990, April). *Differences among interviewers in the incremental validity of their judgements*. Paper presented at the fifth annual meeting of the Society for Industrial and Organizational Psychology, Miami, FL.
- Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology, 71*, 9-17.
- Dunlap, J. W., & Wantman, M. J. (1944). *An investigation of the interview as a technique for selecting aircraft pilots* (Rep. No. 33). Washington, DC: Civil Aeronautics Administration, Airman Development Division.
- Finesinger, J. E., Cobb, S., Chapple, E. D., & Brazier, M. A. B. (1948). *An investigation of prediction of success in naval flight training* (Rep. No. 81). Washington, DC: Civil Aeronautics Administration, Division of Research.
- Fisher, J., Epstein, L. J., & Harris, M. R. (1967). Validity of the psychiatric interview: Predicting the effectiveness of the first Peace Corps volunteers in Ghana. *Archives of General Psychiatry, 17*, 744-750.
- Fowlkes, R. D. (1984). *The relationship between the teacher perceiver interview and instructional behaviors of teachers of learning disabled students*. Unpublished doctoral dissertation. University of Virginia, Richmond.
- Freeman, G. L., Manson, G. E., Katzoff, E. T., & Pathman, J. H. (1942). The stress interview. *Journal of Abnormal and Social Psychology, 37*, 427-447.
- Friedland, D. (1973). *Selection of police officers*. Los Angeles: City of Los Angeles, Personnel Department.
- Friedland, D. (1980). A predictive study of selected tests for police officer selection. Los Angeles: City of Los Angeles, Personnel Department.

- Friedland, D. (n. d.). *Section II - City of Los Angeles Personnel Department (Junior Administrative Assistant Validation Study)*. Los Angeles: City of Los Angeles, Personnel Department.
- Ghiselli, E. E. (1966). The validity of a personnel interview. *Personnel Psychology*, *19*, 389-394.
- Gillies, T. K. (1988). *The relationship between selection variables and subsequent performance ratings for teachers in an Oregon school district*. Unpublished doctoral dissertation, University of Oregon, Eugene.
- Githens, W. H., & Rimland, B. (1964). *The validity of NRTOC selection interviews against career decisions and officer fitness reports: An eight year follow-up* (PRASD Rep. No. 234). San Diego, CA: U.S. Naval Personnel Research Activity.
- Glaser, R., Schwartz, P. A., & Flanagan, J. C. (1958). The contribution of interview and situational performance procedures to the selection of supervisory personnel. *Journal of Applied Psychology*, *42*, 69-73.
- Green, P. C., & Alter, P. (1990). *Validity of a multi-position, Behavioral Interviewing system when using an answer scoring strategy*. Memphis, TN: Behavioral Technology.
- Gregory, E. (1956). *Evaluation of selection procedures for women naval officers* (USN Bureau of Naval Personnel Technical Bulletin, No. 56-11). [Cited in Sawyer, J. (1966)]. Measurement and prediction, clinical and statistical. *Psychological Bulletin*, *55*, 178-200.
- Grove, D. A. (1981). A behavioral consistency approach to decision making in employment selection. *Personnel Psychology*, *34*, 55-64.

- Halliwell, S. T. (1984, November). *An evaluation of the Canadian Forces combat arms officer selection board*. Paper presented at the 26th annual conference of the Military Testing Association, Munich, Germany (Available from Canadian Forces Personnel Applied Research Unit, Willowdale, Ontario, Canada).
- Harris, J. G. (1972). Prediction of success on a distant pacific island: Peace Corps style. *Journal of Consulting and Clinical Psychology, 38*, 181-190.
- Hays, E. J. (1990). *Relationship of a situational interview to the job performance of convenience store clerks*. Unpublished master's thesis, University of North Texas, Denton.
- Hill, C. J., Jr., Russell, D. L., & Wendt, G. R. (1946). *A partial analysis of a three-man interview technique for predicting airline pilot success in the civilian pilot training program* (Rep. No. 65). Washington, DC: Civil Aeronautics Administration, Division of Research.
- Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology, 53*, 1-12.
- Hovland, C. I., & Wonderlic, E. F. (1939). Prediction of industrial success from a standardized interview. *Journal of Applied Psychology, 23*, 537-546.
- Huse, E. F. (1962). Assessments of higher-level personnel: IV The validity of assessment techniques based on systematically varied information. *Personnel Psychology, 15*, 195-205.
- Husen, T. (1954). La validité des interviews par rapport à l'âge, au sexe et à la formation des interview. *Le Travail Humain, 17*, 60-67
- Janz, T (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577-580.

Johnson, E.M., & McDaniel, M. A. (1981, June). *Training correlates of a police selection system*. Paper presented at the Annual Conference of the International Personnel Management Council, Denver, CO.

Johnson, G. (1990). *The structured interview: Manipulating structuring criteria and the effects on validity, reliability, and practicality*. Unpublished doctoral dissertation, Tulane University, New Orleans, LA.

Johnson, H. M., Boots, M. L., Wherry, R. J., Hotaling, O. C., Martin, L. G., & Cassens, F. P., Jr. (1944). *On the actual and potential value of biographical information as a means of predicting success in aeronautical training* (Rep. No. 32). Washington, DC: Civil Aeronautics Administration, Airman Development Division.

Jones, D. E. (1978). *Predicting teacher processes with the teacher perceiver interview*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Kelly, E. L., & Ewart, E. (1942). *A preliminary study of certain predictors of success in civilian pilot training* (rep. No. 7). Washington, DC: Civil Aeronautics Administration, Division of Research.

Kelly, E. L., & Fiske, D. W. (1950). The prediction of success in the VA training program in clinical psychology. *American Psychologist*, 5, 395-406.

Kennedy, R. (1986). *An investigation of criterion-related validity for the structured interview*. Unpublished master's thesis, East Carolina University, Greenville, NC.

Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., & Katzell, R. A. (1968). *Testing and fair employment: Fairness and validity of personnel tests for different ethnic groups*. New York: New York University Press.

- Knights, R. M. (1977). The relationship between the selection process and on-the-job performance of Albuquerque police officers (Doctoral dissertation, University of New Mexico, 1976). *Dissertation Abstracts International*, 38, 1229A-1230A.
- Komives, E., Weiss, S. T., & Rosa, R. M. (1984). The applicant interview as a predictor of resident performance. *Journal of Medical Education*, 59, 425-426.
- Landy, F. J. (1976). The validity of the interview in police officer selection. *Journal of Applied Psychology*, 61, 193-198.
- Latham, G. P., & Saari, 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., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-426.
- Lopez, F. M. (1966). Current problems in test performance of job applicants: I. *Personnel Psychology*, 19, 10-18.
- Martin, M. A. (1972). *Reliability and validity of Canadian Forces selection interview procedures* (Rep. 72-4). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.
- McKinney, T. S., & Meltzzer, D. M. (1979). *The validity and utility of entry level police recruit selection procedures*. Phoenix, AZ: City of Phoenix, Personnel Department, Employee Selection Research.
- McMurry, R. N. (1947). Validating the patterned interview. *Personnel*, 23, 263-272.
- Merenda, P. F., & Clarke, W. V. (1959). AVA validity for textile workers. *Journal of Applied Psychology*, 43, 162-165.

- Meyer, H. H. (1956). An evaluation of a supervisory selection program. *Personnel Psychology, 9*, 499-513.
- Miner, J. B. (1970). Psychological evaluations as predictors of consulting success. *Personnel Psychology, 23*, 393-405.
- Mischel, W. (1965). Predicting the success of Peace Corps volunteers in Nigeria. *Journal of Personality and Social Psychology, 1*, 510-517
- Modderaar, J. (1966). Over de validiteit van de selectie voor bedrijven [About the validity of selection for industry]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 21*, 573-589.
- Mosher, M. R. (1991, June). *Development of a behaviourally consistent structured interview*. Paper presented at the 27th International Applied Military Psychology Symposium, Stockholm, Sweden.
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughn, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77*, 571-587.
- Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology, 70*, 774-776.
- Parrish, J. A., Klieger, W. A., & Drucker, A. J. (1955). *Assessment ratings and experimental tests as predictors of performance in officer candidate school* (Personnel Research Branch Technical Research Note 49). Washington, DC: Department of the Army, Adjutant General's Office.
- Pasco, D. C. (1979). The comparison of two interviewing techniques, the leaderless group discussion and the traditional interview, as a method for teacher

selection (Doctoral dissertation, University of Maryland, 1979). *Dissertation Abstracts International*, 41, 486.

Pashalian, S., & Crissy, W. J. E. (1953). *The interview IV: The reliability and validity of the assessment interview as a screening and selection technique in the submarine service* (Medical Research Laboratory Rep. No. 216, XII, No. 1). Washington, DC: Department of the Navy, Bureau of Medicine and Surgery.

Plag, J. A. (1961). Some considerations of the value of the psychiatric screening interview. *Journal of Clinical Psychology*, 17, 3-8.

Prien, E. P. (1962). Assessments of higher-level personnel: V An analysis of interviewers' predictions of job performance. *Personnel Psychology*, 15, 319-334.

Pulos, L., Nichols, R. C., Lewinsohn, P. M., & Koldjeski, T. (1962). Selection of psychiatric aides and prediction of performance through psychological testing and interviews. *Psychological Reports*, 10, 59-520.

Rafferty, J. A., & Deemer, Jr., W. L. (1950). Factor analysis of psychiatric impressions. *Journal of Educational Psychology*, 41, 173-183.

Raines, G. N., & Rohrer, J. H. (1955). The operational matrix of psychiatric practice: Consistency and variability in interview impressions of different psychiatrists. *The American Journal of Psychiatry*, 3, 721-733.

Reeb, M. (1969). A structured interview for predicting military adjustment. *Occupational Psychology*, 43, 193-199.

Rhea, B. D. (1966). *Validation of OCS selection instruments: The relationship of OCS selection measures to OCS performance* (Tech. Bulletin STB 66-18). San Diego, CA: U. S. Naval Personnel Research.

- Rhea, B. D., Rimland, B., & Githens, W. H. (1965). *The development and evaluation of a forced-choice letter of reference form for selecting officer candidates* (Tech. Bulletin STB 66-10). San Diego, CA: U. S. Naval Personnel Research Activity.
- Ricchio, L. K. (1980). Evaluation of eligible list ranking as a predictor of teacher success. (Doctoral dissertation, Claremont Graduate School, 1980). *Dissertation Abstracts International*, 41, 50.
- Robertson, I. T., Gratton, L., & Rout, U. (1990). The validity of situational interviews for administrative jobs. *Journal of Organizational Behavior*, 11, 69-76.
- Roth, P. L., Campion, J. E., & Francis, D. J. (1988, April). *A LISREL analysis of the predictive power of pre-employment tests and the panel interview*. Paper presented at the third annual meeting of the Society for Industrial and Organizational Psychology, Dallas, Texas. (Validity data not reported in the paper were obtained from P. L. Roth.).
- Rundquist, E. A. (1947). Development of an interview for selection purpose. In G. A. Kelly (Ed.), *New methods in applied psychology* (pp. 85-95). College Park: University of Maryland.
- Shaw, J. (1952). The function of the interview for determining fitness for teacher-training. *Journal of Educational Research*, 45, 667-681.
- Simmons, J. E. (1976). *A study to test teacher perceiver interview as an instrument that would select vocational agriculture instructors that develop positive rapport with their students*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.

- Smedburg, C. A. (1984). *An examination of the validity of the empathy interview as an instrument for teacher selection*. Unpublished doctoral dissertation, University of South Florida, Tampa.
- Solomon, G. L. (1981). *The Omaha teacher perceiver interview as a predictor of job performance of selected new teachers in the Jefferson County school system*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.
- Sparks, C. P. (n. d.). [Title unknown]. Cited in Dunnette, M. D., Arvey, R. D., & Arnold, J. A. (1971). *Validity study results for jobs relevant to the petroleum industry*. Minneapolis, MN: Personnel Decisions.
- Stephen, M. J. (1980). The selection of computer salespersons: An evaluation of a structured interview and its underlying theory. (Doctoral dissertation, Northwestern University, 1980). *Dissertation Abstracts International*, 41, 2372.
- Stohr-Gillmore, M. K., Stohr-Gillmore, M. W., & Kistler, N. (1990). Improving selection outcomes with the use of situational interviews: Empirical evidence from a study of correctional officers for new generation jails. *Review of Public Personnel Administration*, 10, 1-18.
- Tarico, V. S., Altmaier, E. M., Smith, W. L., Franken, E. A., & Berbaum, K. S. (1986). Development and validation of an accomplishment interview for radiology residents. *Journal of Medical Education*, 61, 845-847
- Trankell, A. (1959). The psychologist as an instrument of prediction. *Journal of Applied Psychology*, 43, 170-175.
- Trites, D. K. (1960). Adaptability measures as predictors of performance ratings. *Journal of Applied Psychology*, 44, 349-353.

- Tubiana, J. H. , & Ben-Sitakhar, G. (1982). An objective group questionnaire as a substitute for a personal interview in prediction of success in military training in Israel. *Personnel Psychology*, *35*, 349-357.
- Tziner, A., & Dolan, S. (1982a). Evaluation of a traditional selection system in predicting success of females in officer training. *Journal of Occupational Psychology*, *55*, 269-275.
- Tziner, A., & Dolan, S. (1982b). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology*, *67*, 728-736.
- U. S. Office of Personnel Management (1987). *The structured interview*. Washington, DC: Office of Examination Development, Alternative Examining Procedures Division.
- Waldron, L. A. (1974). The validity of an employment interview independent of psychometric variables. *Australian Psychologist*, *9*, 68-77.
- Walters, L. C., Miller, M., & Ree, M. J. (1993). Structured interviews for pilot selection: No incremental validity. *Internal Journal of Aviation Psychology*, *3*, 25-38.
- 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.
- Wilkins, W L., Anderhalter, O. F., Rigby, M. K., & Stinson, P (1955). *Statistical description of criterion measures for USMC junior officers* (Office of Naval Research Contract N7ONR-40802; NR 151-092). St. Louis, MO: St. Louis University, Department of Psychology.
- Woodworth, D. G., Barron, F., & MacKinnon, D. W (1957). *Analysis of life history interview's rating for 100 air force captains* (Research Rep. AFPTRC-TN,

ASTIA Document AD 146-401). Lackland Air Force Base, TX: Air Research and Development Command.

Wright, P. M., Pursell, E. D., Lichtenfels, P. A., & Kennedy, R. (1986, August). *The structured interview: Additional studies and a meta-analysis*. Paper presented at the Academy of Management Convention, Chicago.

Wunder, S. (1973a). *Validity of a selection procedure for process trainees*. Unpublished manuscript.

Wunder, S. (1973b). *Validity of a selection program for refinery technician trainees*. Unpublished manuscript.

Wunder, S. (1974). *Validity of a selection program for commercial drivers*. Unpublished manuscript.

Wunder, S. (1977a). *The refinery and chemical plant test battery: Validity for laboratory selection*. Unpublished manuscript.

Wunder, S. (1977b). *Selecting process operator and laboratory trainees*. Unpublished manuscript.

Wunder, S. (1978a). *Preemployment tests for pipefitter trainees*. Unpublished manuscript.

Wunder, S. (1978b). *Selecting laboratory technicians*. Unpublished manuscript.

Wunder, S. (1979). *Preemployment tests for electrician trainees*. Unpublished manuscript.

Wunder, S. (1980). *Validity of preemployment measures for instrument technician trainees*. Unpublished manuscript.

- Wunder, S. (1981). *Selection of manufacturing technicians*. Unpublished manuscript.
- Yonge, K. A. (1956). The value of the interview: An orientation and a pilot study. *Journal of Applied Psychology, 40*, 25-31.
- Zarnek, R. J. (1983). *A correlational analysis between the teacher perceiver interview and teacher success in the Chippewa Valley school system*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- Zedeck, S., Tziner, A., & Middlestadt, S. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology, 36*, 355-370.

APPENDIX B

CODING SCHEDULE

Individual/Panel Interviewer Meta-Analysis

Study #: _____

Authors: _____

Coder: _____

__ 1a. Who conducted the interviews?

- (1) individual interviewers (2) panels of 2 or more interviewers
 (3) unable to tell whether individual or panel

__ 1b. If panels, what was the average number of interviewers on each panel? (if number not mentioned, code as “?”).

2. (first digit) What was each interviewer’s position relative to the target position (ie., the position being interviewed for)?

- (1) Immediate manager of the position
 (2) Higher-level manager of the position
 (3) Incumbent of the position
 (4) Personnel specialist/consultant/psychologist/psychiatrist (any HR position)
 (5) From another department
 (6) Peer to target position
 (7) Subordinate to target position
 (8) Other member of org. or not a work org. (role: _____)
 (9) Not member of organisation (role: _____)
 (10) Study did not mention interviewer’s role in org.

(2nd digit) Was the interviewer(s) a psychologist?

- (1) said was psychologist (2) was not psych. (or didn’t say)

- ___ 2a (interviewer 1)
- ___ 2b (interviewer 2)
- ___ 2c (interviewer 3)
- ___ 2d (interviewer 4)
- ___ 2e (interviewer 5)

___ 3. Did interviewers have prior interview experience or training?

- (1) yes (2) no (3) no mention

___ 4. Does the study mention that any job analysis (e.g., interviews, surveys, observation, review of literature) was conducted?

- (1) yes (2) no (If “no”, go to item 9)

___ 5a. Were SME’s *interviewed* to gather job analysis information? (include here mention of “critical incidents”).

- (1) yes (2) no (3) no mention

___ 5b. If SME’s were interviewed, *how many* SME’s were interviewed?

___ 6a. Were SME’s *surveyed (questionnaire)* to gather job analysis information?

- (1) yes (2) no (3) no mention

___ 6b. If surveys were administered, *how many* SME’s were surveyed?

___ 7a. Were job incumbents *observed* to gather job analysis information?

- (1) yes (2) no (3) no mention

___ 7b. If incumbents were observed, *how many total hours* of observation?

- ___ 8a. Was any *job analysis literature* (e.g., job descriptions, training manuals) reviewed to gain job-related information?
(1) yes (2) no (3) no mention
- ___ 8b. If so, *how many sources* of job analysis literature were consulted?
- ___ 9. Did the performance criterion used for the computation of criterion-related validity consist, at least partially, of appraiser judgements (e.g., graphic scale/BARS/BOS/ratings or rankings?)
(1) yes (2) no (3) insufficient information to judge
(Unless “yes”, skip question 10)
- ___ 10. Were interviewers also raters of performance?
(1) *It is certain* that the selection interviewers *were* also the individuals who judged performance?
(2) *It is probable* that the selection interviewers *were* also the individuals who judged performance?
(3) *It is certain* that the selection interviewers *were not* those individuals who judged performance?
(4) *It is probable* that the selection interviewers *were not* those individuals who judged performance?
(5) The report provides *insufficient information* to judge whether interviewers were judges of performance.