

**A REVIEW OF VALIDATION RESEARCH ON  
STRUCTURED EMPLOYMENT INTERVIEWS:  
EXPLORING THE THREATS TO VALIDITY**

by

Ameetha Garbharran

Supervisor: Ms Sonia Hill

A dissertation submitted in partial fulfillment of the  
requirements for the degree of

Master of Social Science (Industrial Psychology)

Centre For Industrial, Organisational And Labour Studies

University of Natal - Durban

2000

## **Abstract**

The main purpose of this study was to explore the threats to internal and external validity that international research on structured employment interviews may have been prone to. On the basis of this exploration, suggestions were rendered on how the perceived threats to internal validity could have been managed. In addition, commentary was offered on whether or not the threats to external validity appear to have been actualised in the international studies that were compared. These constituted the secondary aims of the study. In order to accomplish the primary aim of the study a qualitative approach was employed. Using the literature on the threats to internal and external validity as a fundamental point of departure, studies on the two main variants of structured employment interviews (viz. behaviour description interviews and situational interviews) were analysed. This analysis yielded an insight into the potential threats that are likely to have impacted on the validity findings that were obtained in these studies. On the basis of these insights, suggestions, pertaining to how the threats to internal validity could have been managed, were proposed. In addition, a meta-analytic technique, for comparing the findings across multiple studies, was employed to comment on whether or not the threats to external validity appear to have manifested in the studies in question. These combined insights served as the foundation for offering a South African perspective on the threats to internal and external validity, which included recommendations on how they could be effectively managed in validation research in the South African context.

## **Acknowledgments**

This dissertation constitutes one of the most challenging academic tasks that I have undertaken. The following individuals have played an instrumental role in providing me with invaluable support and insights during the completion of this study:

My husband and soul-mate, Roopesh Ramklass, for his unwavering support, patience and love.

My supervisor, Sonia Hill, for her guidance, perspectives and encouragement.

My nieces, Kiara Ramklass and Amira Ramklass, for their invaluable perspectives and constant lessons in the University of Life.

My parents, Dr Hari Lall Garbharran and Mrs Indra Garbharran, for their encouragement and support.

I would like to take this opportunity to offer my heart-felt thanks to each of these special people.

## Declaration Of Originality

I, Ameetha Garbharran, hereby declare that this dissertation is my own original work, unless it is specified to the contrary in the text. This dissertation has not been submitted for a degree at any other university.

*Garbharran*

Ameetha Garbharran

University of Natal - Durban

December 2000

## Table of Contents

Abstract .....	ii
Acknowledgments.....	iii
Declaration of Originality.....	iv
<b>1. Introduction .....</b>	<b>1</b>
1.0 Introduction .....	1
1.1 Background to the Study.....	1
1.1.1 The Complex Nature of Selection: An International Perspective .....	1
1.1.1.1 The Human Element .....	2
1.1.1.2 Choosing Effective and Appropriate Selection Instruments .....	3
1.1.2 In Search of a Universal Selection Instrument.....	5
1.1.2.1 The Popularity of the Employment Interview.....	5
1.1.2.2 Understanding the Dominance of Employment Interviews.....	6
1.1.2.3 Difficulties with the Popularity of Employment Interviews .....	10
1.1.2.4 The Employment Interview: A Universal Selection Instrument.....	11
1.1.3 Selection in South Africa .....	11
1.1.3.1 The Employment Interview: The Most Frequently used Selection Tool.....	11
1.1.3.2 The Implications of the Pervasiveness of Employment Interviews .....	11
1.2 The Rationale for the Study .....	14
1.3 Statement of the Research Aims .....	15
1.4 The Value of the Research.....	16
1.5 Outline of the Chapters .....	16
<b>2. Literature Review .....</b>	<b>19</b>
2.0 Introduction .....	19
2.1 An Overview of the Dominant Trends in Employment Interview Research .....	19
2.1.1 A Theoretical Conceptualisation of the Major Trends .....	20
2.1.2 A Curvilinear Perspective of the Major Trends .....	20
2.2 An In-Depth Exploration of the Dominant Trends.....	23
2.2.1 The First Trend: Research on Unstructured Interviews.....	24

- 2.2.1.1 In Search of a Definition for Unstructured Interviews .....24
- 2.2.1.2 The Main Highlights of Research in the First Trend ..... 25
- 2.2.1.3 An Appraisal of Research on Unstructured Interviews.....31
- 2.2.2 The Second Trend: Research on Decision-Making in Interviews..... 35
- 2.2.3 The Third Trend: Research on Structured Employment Interviews..... 37
- 2.2.3.1 Early Perspectives on Structured Interviews.....37
- 2.2.3.2 Early Investigations of Structured Interviews ..... 38
- 2.2.3.3 In Search of a Definition for Structured Interviews .....41
- 2.2.3.4 Major Contributions of Research in the Third Trend.....48
- 2.2.3.5 Major Criticisms of Research on Structured Interviews .....53
- 2.3 Conclusion..... 55
  
- 3. Methodology..... 58**
- 3.0 Introduction ..... 58
- 3.1 A Statement of the Research Problem ..... 58
- 3.2 Exploring the Concepts of Internal and External Validity..... 58
- 3.2.1 Internal Validity..... 58
- 3.2.2 External Validity..... 59
- 3.2.3 The Link Between Internal Validity and External Validity ..... 60
- 3.3 The Sample..... 61
- 3.4 The Method ..... 63
- 3.5 Conclusion..... 69
  
- 4. Results ..... 70**
- 4.0 Introduction ..... 70
- 4.1 The Exploration of Threats to Internal Validity..... 70
- 4.1.1 Research Design Strategies ..... 71
- 4.1.2 Threats to Internal Validity and Research Design Strategies ..... 73
- 4.2 The Exploration of Threats to External Validity..... 78
- 4.3 Conclusion..... 89

<b>5.</b>	<b>Discussion.....</b>	<b>90</b>
5.0	Introduction .....	90
5.1	Research on Patterned Behaviour Description Interviews.....	91
5.1.1	An Exploration of the Threats to Internal Validity .....	91
5.1.1.1	Research Design Strategy: Within-Subjects Experiments.....	91
5.1.1.2	Threats to Internal Validity in Within-Subjects Experiments .....	92
5.1.2	An Exploration of the Threats to External Validity .....	104
5.2	Research on Structured Behavioural Interviews .....	106
5.2.1	An Exploration of the Threats To Internal Validity .....	107
5.2.1.1	Research Design Strategy: Correlational Studies.....	107
5.2.1.2	Threats to Internal Validity in Correlational Studies .....	107
5.2.2	An Exploration of the Threats to External Validity .....	114
5.3	Research on Situational Interviews .....	117
5.3.1	An Exploration of the Threats to Internal Validity .....	117
5.3.1.1	Research Design Strategies .....	117
5.3.1.2	Threats to Internal Validity.....	117
5.3.2	An Exploration of the Threats to External Validity .....	129
5.4	Conclusion.....	132
<b>6.</b>	<b>Conclusions and Recommendations.....</b>	<b>135</b>
6.0	Introduction .....	135
6.1	Recommendations.....	136
6.1.1	Research Design Strategies .....	136
6.1.2	Threats to Internal Validity .....	136
6.1.2.1	Organismic Variables.....	137
6.1.2.2	Time-Tied Extraneous Variables.....	142
6.1.2.3	Invalid Operational Definitions .....	143
6.1.2.4	Sampling Biases .....	145
6.1.2.5	Uncontrolled Variation in Information .....	147
6.1.3	Threats to External Validity .....	147
6.2	Conclusion.....	149
	<b>Bibliography.....</b>	<b>152</b>

## List Of Figures, Graphs and Tables

### Figures

Figure 3.1 A Decision-Tree for Identifying Research Design Strategies.....	.65
--	-----

### Graphs

Graph 2.1 The Sigmoid Curve.....	.21
Graph 2.2 A Curvilinear Perspective of the Dominant Trends .....	.22

### Tables

Table 2.1 The Contributors to Structure in Employment Interviews.....	.47
Table 2.2 The Impact of the Fifteen Components of Structure on Validity.....	.50
Table 3.1 The Studies Considered for Inclusion in this Investigation.....	.61
Table 3.2 Frequency of Studies Per Structured Interview Variant .....	.64
Table 4.1 Classification of Studies in terms of Research Design Strategies.....	.71
Table 4.2 Threats to Internal Validity on the Basis of Research Design Strategy .....	.73
Table 4.3 Definitions of Threats to Internal Validity.....	.74
Table 4.4 Classification of the Sampling Techniques.....	.76
Table 4.5 Definitions of the Threats to External Validity.....	.78
Table 4.6 A Summary of the Quantitative Validity Results.....	.80
Table 4.7 Pair-Wise Comparisons of the Validity Estimates for PBDIs, SBIs and SIs.....	.85
Table 4.8 Comparisons of Validity Estimates Yielded in Three or More Studies .....	.87



# **Chapter 1**

## **Introduction**

### 1.0 Introduction

This chapter is designed to offer a brief introduction to the present study. The introduction will provide a background to the study, offer a rationale for the investigation, state the research problem and explicate the benefits of the study. In addition, an overview of the chapters that comprise this investigation will be provided.

### 1.1 Background to the Study

Employment interviews have been identified as the most popular and widely used selection tool of all time. In this study the main focus is on the validation research that has been conducted to ascertain the effectiveness of the structured employment interview as a selection tool. In this section, the author will provide international and local South African perspectives on the complexities inherent in the selection process, trace the employment interview's prodigious rise to fame and provide an argument for why South African selectors should employ structured employment interviews as their preferred selection tool. This discussion will offer the reader an insight into the context in which employment interviews, in general, and structured employment interviews, in particular, are used. These insights will facilitate the contextualisation of validation research on structured and unstructured employment interviews in terms of the environments in which they have been conducted and the realities by which they have been influenced. Unstructured employment interviews served as the precursor to their structured counterparts. Therefore, the background to this study will encompass an exploration of unstructured employment interviews as selection tools in addition to the more structured variants of employment interviews that emerged during the course of employment interview history.

#### 1.1.1 The Complex Nature of Selection: An International Perspective

Selection is an indispensable and critical component of the human resource management systems that operate within organisations (Cascio, 1991; Gerber, Nel and van Dyk, 1995). The selection process may be regarded as a system in its own right, which may be likened to a filter that regulates the individuals that are employed into organisations on the basis of organisational-specific and job-specific criteria. Human resource management systems, particularly selection systems, are informed by

internal organisational dynamics such as organisational culture and by external environmental factors such as the social, political, cultural, economic and legal climates within which organisations exist and function. These systems are extremely sensitive to and deeply influenced by their internal and external environmental circumstances. They, in turn, impact on the external and organisational environments, thereby 'providing the feedback loop to complete the system', in the language of systems theory (French and Bell, 1995).

This relationship of mutual interaction between, amongst others, the selection system and organisational and external environmental influences, poses many challenges to organisations. The advent of globalisation and the birth of multinational organisations serve to exacerbate these difficulties by creating a fairly unique business scenario that requires an adaptation of organisational practices, in general, and human resource management practices in particular, including the age-old process of selection (Ryan, McFarland, Baron and Page, 1999).

#### 1.1.1.1 The Human Element

At the best of times, the selection of individuals into organisations is arduous. One of the major reasons for this is that the selection process is executed by human beings. Hence, it is not infallible, since it hinges on human judgement from its conceptualisation and inception through to its execution and evaluation (Lewis, 1992). A major counter-argument to this sentiment may be that there are several systematic and scientific procedures in existence for the identification and selection of suitable candidates. These are typically oriented towards fulfilling the following functions:

- Identifying the criteria that are to be assessed in the selection process (this is usually based on the requirements for the job);
- Assessing the extent to which job applicants fulfil pre-specified selection criteria;
- Evaluating the outcomes of the assessment tools that are used for selection; and
- Integrating the information derived from the selection process to render selection decisions.

However, despite the scientific origins of these procedures, whether or not they are implemented is dependent on the choices that selectors make and whether or not they are used correctly is dependent on the levels of competence that selectors possess.

### 1.1.1.2 Choosing Effective and Appropriate Selection Instruments

Wood and Payne (1998) suggested that the most critical consideration in the selection process pertains to which selection method or methods to use to assess job applicants. The passage of time has yielded many significant developments in selection 'technology'. One such development is the impressive range of selection tools, which may have served to compound the dilemmas associated with the key selection decisions for which selectors are responsible. Some of the most common instruments used to derive information about job applicants include: interviews (unstructured and structured interviews, one-on-one and panel interviews), cognitive ability tests, personality questionnaires, assessment centres, application forms, biographical data forms, recommendations from references, work sample tests and honesty and integrity assessments (Muchinsky, 1993; Schultz and Schultz, 1986; Whetzel and Wheaton, 1997).

Typically, the considerations for the choice of the selection instruments to be included in selection processes have revolved around job-related factors and instrument-related factors. In effect, therefore, selection instruments were chosen if they were regarded as being capable of assessing the criteria identified as critical for effective job performance and if they were scientifically proven to be psychometrically sound (i.e. they were shown to be valid and reliable).

#### (i) Globalisation and Multinational Corporations

However, the advent of globalisation has necessitated the consideration of further factors, at the national level, that impact on the choice of selection instruments. An exploration of international selection practices revealed considerable variability in the selection instruments that are used across national boundaries (Ryan et al., 1999). This tendency may pose unique problems to multinational organisations that are striving towards the standardisation of selection practices in their international subsidiaries. The use of selection instruments that are not familiar to selectors or applicants may require more intensive instruction and training for the former and more practice with sample items for the latter (Ryan et al., 1999). In addition, the use of unfamiliar selection instruments may affect applicants' perceptions of employers and consequently impact on the recruitment process (Ryan et al., 1999).

#### (a) Cultural Factors

Ryan et al. (1999) proposed that national variability in the use of selection practices could be attributed to cultural differences on Hofstede's dimensions of uncertainty avoidance and power distance.

Uncertainty avoidance refers to the extent to which individuals that exist within a particular cultural environment feel threatened by uncertainty in situations whereas power distance refers to the extent to which less powerful people in a cultural environment accept that power is unequally distributed (Ryan et al., 1999). Although their investigation yielded mixed results, the researchers concluded that cultural differences in uncertainty avoidance, and power distance to a lesser extent, did explain some of the national differences in terms of the extensiveness with which selection methods were used on an international scale (Ryan et al., 1999).

Thus, multinational organisations should strive to ensure congruence between the cultural factors that inform and shape their international subsidiaries and the selection practices that are employed within these subsidiaries. This is imperative for both the selection process and the specific selection practices that are employed under its auspices to be perceived as credible by local applicants and selectors. A perceived lack of credibility could lead to the ineffective conceptualisation, execution and evaluation of the selection process by selectors and an unwillingness on the part of applicants to join the organisation. Hence, regardless of the cultural factors that inform the operations of multinational organisations in the “mother-country”, organisational practices should be oriented to accommodate cultural nuances on an international level in order to attain the universally sought after outcome of organisational effectiveness.

#### (b) Legal Factors

In addition to highlighting the importance of cultural differences on the choice of selection instruments, globalisation has yielded a complex legal environment that transcends national boundaries (Terpstra, Mohamed and Kethley, 1999). This implies that multinational organisations with operations in different countries are increasingly being compelled to conform to the specific legal environments of the countries in which they operate. For example, Terpstra et al. (1999) cited Schuler and Jackson who reported in 1996 that a number of foreign companies operating in the United States of America had been sued for discriminatory employment practices by their American employees. Thus, in the context of selection, multinational organisations should consider the legal implications of the selection techniques that they choose to employ in different national contexts (Terpstra et al., 1999).

#### (ii) The Global Challenge in the Realm of Selection

Multinational organisations, therefore, are charged with the intricate and complex challenge of establishing an amicable balance between the cultural and legal factors that operate at a national level to inform the choice of selection instruments in the selection process. Their ultimate aim is to use

selection practices that conform to both the cultural and legal environments within the operating countries.

In this discussion, the author has attempted to highlight some of the difficulties associated with the selection process, particularly with regard to the choice of appropriate and acceptable selection instruments within and outside of the context of globalisation. This discussion is not meant to be comprehensive or exhaustive. Instead, its primary aim is to highlight the complex nature of selection in the international arena and to explicate some of the challenges that confront the major role-players in the selection process.

### 1.1.2 In Search of a Universal Selection Instrument

These global realities in the context of selection have created an urgent need for the identification of universal selection practices (Ryan et al., 1999). The employment interview may be perceived as one of the prime contenders for the esteemed title of 'universal selection instrument' on the basis of its popularity, on an international scale, during the last century.

#### 1.1.2.1 The Popularity of the Employment Interview

The employment interview has been cited as one of the most popular and pervasive selection tools of the 20<sup>th</sup> century (Harris, 1989; Mayfield, 1964; Moffatt, 1969; Schmitt, 1976; Ulrich and Trumbo, 1965; Wagner, 1949; Whetzel and McDaniel, 1997; Wright, 1969). Evidence of its widespread use has been chartered through the course of time by several commentators. This has been captured in survey research aimed at ascertaining the frequency with which employment interviews have been used as selection tools by organisations.

Initially, this evidence was confined to the United States of America. Ulrich and Trumbo (1965) cited the research of Spriegel and James, which was conducted from the 1930s through to the late 1950s. Their findings over this period of time revealed that the overwhelming majority of American organisations used the employment interview for the purposes of hiring employees.

In a later survey reported by the Bureau of National Affairs in America in 1988, the frequency with which various selection techniques were used for screening individuals for five different job types was assessed (Dipboye, 1992). The results revealed that 82% of the companies surveyed reported that an interview with the prospective employees would be required (Dipboye, 1992). In addition, these companies were less likely to report the use of other selection tools (Dipboye, 1992).

However, the acclaim of the employment interview has not been confined to organisations in the United States of America. Shackleton and Newell (1997) conducted surveys to compare the selection techniques used to recruit managers in Europe. Initially, their research focused only on large companies in the United Kingdom and France (Dipboye, 1992). These researchers subsequently extended their research to include other European countries, namely, Belgium, Germany and Italy (Shackleton and Newell, 1997). Their findings were oriented to ascertain the frequency with which these select European countries made use of interviews, application forms, references, personality tests, cognitive tests, handwriting analysis, biodata and assessment centres. Their analyses were categorised in terms of the extent to which these selection tools were used to recruit managers either always or sometimes (Shackleton and Newell, 1997). The findings revealed that, with the exception of Germany, the interview was always used by the majority of the British (91%), French (94%), Italian (96%), Belgium-Flemish (92%) and Belgium-French (100%) companies in comparison with the other selection techniques that were always used (Shackleton and Newell, 1997). The majority of German companies (83%) always made use of the application form when recruiting managers in comparison with the 60% of German companies, which always used the interview (Shackleton and Newell, 1997). A similar trend (as that observed for British, French, Italian and Belgian companies) was noted for American (99%) and Australian (91%) companies (Shackleton and Newell, 1997).

Ryan et al. (1999) surveyed organisations in twenty countries around the world in order to ascertain which selection method was used most frequently. Their findings revealed that the face-to-face interview was the most widely used selection tool in comparison with a range of other tools that included application forms, educational qualifications, personal references, group or panel interviews, tests and questionnaires and biodata forms (Ryan et al., 1999). The five countries that made the most extensive use of the face-to-face interview (in descending order) were Greece, France, Sweden, the United States of America and South Africa (Ryan et al., 1999).

#### 1.1.2.2 Understanding the Dominance of Employment Interviews

Thus, on the basis of this impressive international evidence, the employment interview is probably the most ubiquitous selection tool of all time. In the light of its tremendous popularity, one would expect to find sound justification for its widespread use. However, the psychometric evidence that exists to support the use of employment interviews for selection purposes does not appear to be congruent with or justify their extensive global proliferation. Perhaps, other explanations exist. In this section, some of the reasons that have been offered for the continued use of interviews in organisations will be explored.

### (i) The Psychometric Evidence

A logical assumption that may be derived from the dominance of the employment interview during the 20<sup>th</sup> century is that this tool has been subjected to scientific scrutiny and has emerged from this inquiry as an effective, efficient and, consequently, psychometrically sound selection device. Investigations of the validity and reliability of employment interviews have constituted one of the most prevalent research streams during the 20<sup>th</sup> century. Two distinct findings emerged from these research efforts. These pertained to unstructured and structured variants of employment interviews.

The earliest employment interviews were unstructured in nature. The validity and reliability evidence for these interviews was extremely disappointing (Mayfield, 1964; Moffatt, 1969; Ulrich and Trumbo, 1965; Wagner, 1949; Wright, 1969). This constituted the first distinct finding in employment interview research. In an effort to enhance their psychometric qualities, it was proposed that unstructured interviews evolve into more structured selection devices. Although this proposition was made relatively early in the 20<sup>th</sup> century, in one of the earliest reviews of the research on employment interviews by Wagner (1949), it only appears to have been heeded in the 1980s when structured variants of the employment interview appeared (Eder and Ferris, 1989). Two of the more prominent variants that emerged during this period, were patterned behaviour description interviews and situational interviews (Arvey and Campion, 1982; Harris, 1989). Investigations into the validity and reliability of structured employment interviews yielded more promising results (Dipboye, 1992). This represented the second distinct finding in employment interview research.

Evidence of the preponderance of the employment interview as a selection tool predated the development of structured interviews and the more encouraging validity and reliability results that they yielded. Hence, the unstructured employment interview provided the impetus for the interview's ascension to international acclaim. The fascinating question that this reality poses is why the unstructured employment interview attained this unparalleled fame as a selection tool despite its disappointingly low validity and reliability results.

A significant reason for the continued use of employment interviews, despite the overwhelmingly negative psychometric evidence that accrued to discredit them, may be attributed to the perception that they really are valid (Arvey and Campion, 1982; Dipboye, 1992). One of the arguments that is presented in this regard suggests that the employment interview can be used to make valid assessments of the interpersonal behaviour that is demonstrated during the interview (Arvey and Campion, 1982; Dipboye, 1992). This, however, does not imply that it is a valid predictor. It may attain this status, however, if the interpersonal behaviour that is evaluated during the interview is directly related to

effective job performance and if a valid assessment of this behaviour by the interviewer (or interviewers) results in effective and accurate predictions of job performance.

A second argument, based on the premise that employment interviews really are valid, suggests that the psychometric techniques that have been used to evaluate them are incapable of detecting their true validity due to a range of statistical and methodological constraints. These include restriction of range, small sample sizes and homogeneous applicant populations, which lead to attenuated correlation coefficients (Arvey and Campion, 1982).

These arguments are, indeed, compelling and fascinating. However, in the light of limited and inconclusive evidence to support them, they remain theoretical conceptualisations of very real possibilities. In order to prove their plausibility and confirm their accuracy, substantial research evidence is required.

#### (ii) Beyond Selection: Extending the Utility of Employment Interviews

The continued use and popularity of employment interviews have been attributed to the notion that they really are valid instruments, by some commentators. There is an alternative perspective, which concedes that the psychometric qualities of unstructured interviews are poor. However, they have many other uses that serve to perpetuate their pervasiveness (Dipboye, 1992).

The face-to-face interaction and two-way dialogue opportunities that interviews engender enable them to fulfil other functions, apart from the identification of suitable job applicants for the purpose of selection. Firstly, interviews can serve as a recruitment device (Dipboye, 1992; Gatewood and Field, 1994; Rynes, 1989). Interviewers are able to attract applicants to an organisation and influence their decisions about whether or not they would like to be employed within it by portraying the organisation in a positive light and by glorifying its merits during the dialogue (Rynes, 1989).

Secondly, interviews are useful for the purpose of information dissemination since they allow applicants the opportunity to ask questions and obtain clarity on issues pertaining to the job and the organisation (Gatewood and Field, 1994).

A third function that interviews can fulfil is to offer career advice and guidance to applicants (Dipboye, 1992). In the context of this function, interviewers play an advisory role to applicants in order to assist them to make optimal career decisions (Dipboye, 1992). Their goal is to help applicants to evaluate



whether or not the job and the organisation are best suited to their needs and abilities towards the ultimate aim of assisting them to find suitable employment (Dipboye, 1992).

Fourthly, interviews can serve a symbolic function (Dipboye, 1992). They can be used as a vehicle to communicate the values of the organisation to applicants (Dipboye, 1992). For example, if an organisation wishes to be perceived as being concerned with its employees on a personal level, it may include interviews as part of its selection process. Interviews offer employers the opportunity to interact personally with applicants, in comparison with a battery of tests, for instance, which does not allow for as much interpersonal contact (Dipboye, 1992).

Interviews generally constitute the first contact between employing organisations and applicants (Dipboye, 1992). They are an ideal opportunity for employers to inform applicants about the nature of the job, the responsibilities that they will be required to fulfil and the expectations that the organisation has of them (Dipboye, 1992). In addition, applicants can also use interviews as a forum for outlining their needs and expectations to the organisation. Thus, interviews provide the forum for the initial establishment of the psychological contract between applicants and organisations (Handy, 1993; Schein, 1980). Therefore, the fifth function of interviews is the initiation of the psychological contract. If applicants are eventually employed into the organisation then the foundation for the psychological contract will already be in place.

In fulfilling each of these five functions, interviews may also serve as public relations tools for organisations (Arvey and Campion, 1982; Dipboye, 1992; Gatewood and Field, 1994). Thus, interviews are extremely versatile (Dipboye, 1992). Perhaps, it is the versatility of unstructured interviews rather than their psychometric qualities that contributed to their exceptional popularity during the 20<sup>th</sup> century.

### (iii) Into the Realm of Irrationality

Dipboye (1992) suggested that in addition to the more rational reasons for the prominence of unstructured employment interviews, some of which have been captured in the previous discussion, a few less rational reasons for their popularity have been offered. Firstly, the popularity of employment interviews has been attributed to the fact that they have become so ingrained in organisations that they continue to be used to perpetuate a norm (Dipboye, 1992).

Secondly, it has been argued that the prolific use of employment interviews within organisations may be attributed to the belief that they fulfil some personal needs (e.g. the need to feel personally responsible for selection outcomes) for the individuals who conduct them (Dipboye, 1992).

Initial surveys that reported the popularity of employment interviews emanated from the United States of America. A misinterpretation of the American Civil Rights Act of 1964 led to the assumption that the use of unscored selection interviews would significantly reduce legal rebuttals on the count of discrimination against minority groups (Arvey and Campion, 1982). This could have led to the perpetuation of unstructured employment interviews by organisations in their attempt to avoid being charged for unfair discrimination. It is possible that this interpretation contributed to their popularity and pervasiveness in the United States of America. Their popularity in the international arena could have been attributed to other countries following the example set by American organisations in the sphere of selection. It is not clear, however, if their popularity waned in the light of the fact that unstructured interviews have been the most frequently challenged selection device on the count of discrimination in America's legal arena (Terpstra et al., 1999).

While this discussion is not exhaustive, it does shed some light on some of the less legitimate perspectives that could have contributed to the popularity of these selection instruments.

#### 1.1.2.3 Difficulties with the Popularity of Employment Interviews

Thus, there are many possible explanations for why employment interviews have attained international recognition and acclaim. It is difficult, if not impossible, however, to pinpoint the actual reason or combination of reasons that accounts for their pervasiveness. This is, unfortunately, not the only difficulty associated with the popularity of employment interviews.

The surveys that have demonstrated the popularity of employment interviews seem to suggest that there is a single entity called the employment interview, which is used extensively in organisations all over the world. Apart from a distinction between one-on-one and panel interviews which was noted by Ryan et al. (1999), the majority of these surveys do not take cognisance of the reality that employment interviews have evolved and diversified over time to yield two very broad categories of interviews (viz. unstructured and structured interviews). Although it has been established that the unstructured variant sparked off the employment interview's immense popularity, these surveys do not offer a definition of employment interviews. As a result, it is unclear whether the dominance of these instruments may be attributed to the unstructured variants, the structured variants or to both. Recent survey research by Shackleton and Newell (1997) suggested that structured interviews such as behavioural interviews and

situational interviews are gaining in popularity in some countries such as the United States of America and the United Kingdom. In the light of this reality, the extent of the contributions that the different variants of employment interviews have made to their overall popularity is confounded.

#### 1.1.2.4 The Employment Interview: A Universal Selection Instrument

The international reputation that employment interviews have earned, irrespective of the variants that have informed it or the legitimacy of the reasons on which it is grounded, render them relatively universal selection instruments. Their utility in the international arena, however, will be informed by the unique interplay of social, political, economic, legal and cultural factors in different national contexts. Furthermore, their continued popularity and pervasiveness will ultimately depend on whether or not the assumptions and philosophies on which employment interviews are grounded are congruent with the interaction of these crucial factors at a national level.

#### 1.1.3 Selection in South Africa

In this section, the nature of selection in South African organisations will be explored. This discussion will culminate in a motivation for the use of structured employment interviews in the South African context and a rationale for this recommendation.

##### 1.1.3.1 The Employment Interview: The Most Frequently used Selection Tool

The one-on-one variant of the employment interview was reported as the most frequently used selection tool in the South African context (Ryan et al., 1999). Other selection tools, which were also used frequently by the sample of 54 South African organisations in the survey, were educational qualifications, application forms, employer references, personal references and panel interviews (Ryan et al., 1999). Although these findings pertain only to the handful of organisations that served as respondents in the survey, the use of employment interviews, in general, appears to be rife in South African organisations.

##### 1.1.3.2 The Implications of the Pervasiveness of Employment Interviews

The unique macro-level factors, that impact the organisational environments within which selection occurs in the South African context, have created an extraordinary set of national circumstances that have dire consequences for organisational practices, in general, and selection, in particular, in South African organisations.

(i) The Macro Environment: Focusing on Employment Equity

The broader macro context in South Africa is characterised by social, political, economic, cultural and legal factors at the national level. South Africa's history has been riddled with conflict and adversarialism as a result of the apartheid regime's perpetuation of blatant discrimination against black people. This has created vast rifts between black people and white people in every sphere of life encompassed in the social, political, economic and legal realms of society. The inequities in the labour market and the distinct disparities in the distribution of jobs, occupations and incomes (Society for Industrial Psychology, 1998) coupled with the grossly high prevalence of unemployment is indicative not only of the rift between black people and white people but also of the huge inequities between men and women and between people with disabilities and the complement of South Africa's able-bodied workforce and prospective employees.

Economically active South Africans and those floating around in our saturated labour market have had to contend with the triple-bind effects of the aftermath of racial discrimination, the subtle but ever-present consequences of covert patriarchy and the social stereotypes and prejudices held against people with disabilities. This depicts a bleak scenario for individuals in employment or seeking employment in South African organisations. However, the institution of the Employment Equity Act (EEA), No.55 of 1998 has offered employees and prospective employees a glimmer of hope.

The EEA came into being in a fervent attempt to actively promote the fundamental principles of our fledgling democracy in South African organisations. The essential purpose of the Act is to ensure fairness and equality in the employment context. It proposes to achieve this aim in two ways:

- By advocating the eradication of unfair discrimination in employment practices and procedures; and
- By instituting affirmative action measures to counteract the disadvantages in employment experienced by previously disadvantaged groups (referred to as designated groups in the EEA and which by definition include black people, women and people with disabilities).

The scenario envisaged by the EEA, therefore, is an employment context that is devoid of unfairness, in which all South Africans, from a multitude of diverse backgrounds, enjoy equal representation across occupational categories and levels in the workforce. In order to realise this desired state of affairs, the EEA clearly stipulates the measures that will be taken to monitor and enforce the provisions of the Act, together with the responsibilities of the parties who will be actively involved in the

monitoring and evaluation processes. The penalties for non-compliance with the Act's provisions are also clearly delineated.

## (ii) The Employment Equity Act and Employment Interviews

In the EEA it is apparent that psychological testing and other similar assessments are regarded as having the potential to result in unfair discrimination. In an effort to prohibit unfair discrimination, towards the ultimate end of ensuring employment equity, the Act stipulates, in section 8, that:

“Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment being used –

- (a) has been scientifically shown to be reliable and valid;
- (b) can be applied fairly to all employees; and
- (c) is not biased against any employee or group.”

Although this provision does not explicitly prohibit the use of employment interviews, it may be argued that employment interviews, in conjunction with a range of other selection tools are encompassed in the broad term “other similar assessments of an employee”. In accordance with this interpretation of the provision, employment interviews may not be used to assess employees in the South African context unless they are scientifically valid and reliable, can be applied fairly to all employees and are not biased against any employee or group of employees. In terms of section 9 ‘employees’ include applicants for employment.

Earlier in this discussion, it was established that employment interviews are extremely popular selection tools in the South African context. However, due to a lack of detail, it is not known whether this popularity is a result of the widespread use of unstructured interviews, structured interviews or both. It is clear, however, in the light of South Africa’s legal stance on the issue of employment equity, that the continued use of employment interviews is dependent on the extent of their compliance with the provisions of the EEA, in general, and with the above-mentioned provision, in particular.

International evidence on the psychometric properties of employment interviews has revealed extremely disappointing validity and reliability estimates for unstructured interviews and more promising validity and reliability evidence for their structured counterparts. In order to conform with the first condition of section 8 of the EEA, therefore, it is logical for South African organisations to use structured employment interviews for the purpose of selection and, whenever relevant and appropriate, for making other critical organisational decisions.

The EEA further stipulates in section 6(2b) that “It is not unfair discrimination to – distinguish, exclude or prefer any person on the basis of an inherent job requirement”. This provision has a direct bearing on the practice of selection, which by its very nature, is a process of discrimination and differentiation between more effective and less effective job applicants. It attempts to prevent unfair discrimination by proposing that differentiation between individuals may be fair if it occurs on the basis of inherent requirements for the job. This necessarily has implications for the tools that are used to discriminate between more effective and less effective individuals. All the tools that are used for making such distinctions in the South African context should be oriented towards assessing relevant job-related criteria so that organisational decision-makers can comply with this provision and select or reject individuals on the basis of inherent job requirements. Structured employment interviews are generally premised on detailed job analyses. For example, the patterned behavioural description interviews, structured behavioural interviews and situational interviews are all based on the critical incident job analysis technique (Janz, 1982; Whetzel and McDaniel, 1997). Thus, the use of structured interviews in the South African context should also ensure compliance with this provision of the EEA.

On the basis of the international evidence that has been cited, it seems that structured employment interviews may be able to withstand legal challenge, certainly on the basis of their psychometric properties and their job-relatedness, in the South African context.

## 1.2 The Rationale for the Study

There is a gross incongruence between the tremendous popularity of employment interviews in the South African context and the severe lack of published and documented research on their use for selection purposes. In the light of the provisions of the EEA, employers have a legal obligation to prove that the psychological tests and other similar assessments that they use to assess employees, are valid and reliable (amongst other conditions that have been outlined above) in order for the use of these tools to be legally justifiable and defensible. The psychometric properties of validity and reliability can only be ascertained in empirical research. Therefore, the continued use of employment interviews in the South African context is dependent on the derivation of positive validity and reliability evidence from scientific inquiries into this popular selection tool.

The international evidence that has been yielded on the structured variant of employment interviews seems to suggest that it may be well suited to South Africa’s unique legal climate. Hence, South African employers and researchers should contemplate the use of this variant of employment interviews for the purpose of selection. Notwithstanding the positive validity and reliability evidence

that has been obtained for structured employment interviews in the international arena, employers that choose to use them are compelled by the law to prove their validity and reliability, among other things, in the South African context.

In the light of these realities in South Africa, the author believes that organisations that continue to use employment interviews for the purpose of selection will opt to use structured variants of these selection instruments. In the absence of local validity evidence for these tools, there is likely to be an increasing need for organisations to initiate and fund validation research on the structured employment interviews that they use to render crucial selection decisions.

In this investigation, the threats that appear to have impacted on the internal and external validity of international validation research on structured employment interviews will be explored. The key objectives in this study are to highlight these threats for South African researchers and to offer recommendations regarding how they could safeguard their validation efforts from them. The author believes that the insights that are yielded in this study will be beneficial to South African researchers who attempt to offer local validation evidence for structured employment interviews. South African researchers could incorporate these insights on the threats to internal and external validity and the recommendations regarding how they may be constructively managed, into their research endeavours in the quest to yield sound validation results.

### 1.3 Statement of the Research Aims

In this investigation, the main aim was to explore the potential threats to internal and external validity that international validation research on structured employment interviews may have been prone to. The literature on research methodology alludes to an interdependence between internal and external validity. As a result of this interdependence, research investigations may only possess external validity if they possess internal validity. It is not the purpose of this study to explicate the link between internal and external validity in the context of international validation research on structured employment interviews. Therefore, in this investigation independent explorations of the threats to internal and external validity were engaged in.

In addition to an exploration of the threats to internal and external validity in the international research on structured employment interviews, this study was oriented towards the attainment of two secondary aims:

- To comment on how the potential threats to internal validity could have been effectively managed; and
- To offer an insight into whether or not the potential threats to external validity appear to have been actualised in the context of international validation research efforts.

In this investigation the qualitative explorations of the potential threats to internal and external validity were based on the insights yielded by Stern and Kalof (1996) and Dooley (1995) respectively. A meta-analytic technique, proposed by Rosenthal (1991), was used to compare the findings across multiple studies, which investigated the same variables, in order to comment on whether or not the threats to external validity appear to have been actualised in the international studies under consideration.

#### 1.4 The Value of the Research

The primary impetus for embarking on this analysis was to assist South African researchers to conduct sound validation research on structured employment interviews. However, the author believes that the benefits of this study extend beyond the South African context. The insights on the threats to internal and external validity and the suggestions that were made regarding how these could be managed are likely to be relevant to all researchers who are involved in the validation of structured employment interviews. In order to benefit the South African researchers, for whom this study was specifically intended, however, the implications of the suggestions for controlling threats to internal and external validity in validation research in South Africa, were explored in the final chapter.

#### 1.5 Outline of the Chapters

This investigation consisted of six chapters. In this section, a brief synopsis of each chapter is provided.

##### (i) Chapter 1 – Introduction

In this chapter the author provided international and local perspectives on the contexts in which employment interviews are used. This commentary included an acknowledgement of the popularity of employment interviews as selection tools and a motivation for the benefits of structured variants of employment interviews both internationally and locally. This discussion served as the backdrop



against which a rationale was offered for the present study, the research aims were stated and the value of the investigation was outlined.

#### (ii) Chapter 2 – Literature Review

In the second chapter, the history of the research that was conducted on employment interviews during the 20<sup>th</sup> century was traced. Three dominant trends were identified in the realm of employment interview research. These were explored and the major criticisms that related to research within each of these trends were highlighted.

#### (iii) Chapter 3 – Methodology

This chapter encompassed a discussion of the methods and procedures that were employed to analyse the international research included for analysis in the study. A qualitative approach was adopted. The exploration of the threats to internal and external validity conformed to this approach. However, in the attempt to establish if the perceived threats to external validity did, in fact, manifest in the studies under investigation, a quantitative meta-analytic procedure was employed.

#### (iv) Chapter 4 – Results

The results that were derived from the analysis of international research efforts using the methods proposed in the previous chapter, are presented in this chapter.

#### (v) Chapter 5 – Discussion

In this chapter an exploration of the potential threats to the internal and external validity of the international research, that was analysed, was engaged in. In relation to the threats to internal validity, suggestions were made regarding how they could have been overcome and managed. The findings that were derived from the use of the meta-analytic technique for comparing the results of studies was used to comment on whether the perceived threats to external validity appeared to have manifested in the studies under investigation.

(vi) Chapter 6 – Conclusions and Recommendations

This chapter outlined the major conclusions that were derived from the study. On the basis of these, recommendations regarding how validation research on structured employment interviews may be effectively conducted in organisational contexts, in general, and in the South African environment, in particular, were made.

## **Chapter 2**

### **Literature Review**

#### 2.0 Introduction

The popularity of the employment interview was heralded, in academic circles, by a surge of research initiatives on this prolific selection tool during the 20<sup>th</sup> century. Some of the earliest research efforts have been traced back to the early 1900s (Mayfield, 1964; Ulrich and Trumbo, 1965; Wagner, 1949). Although the emphasis of the research has altered over time, the interest that has been afforded the investigation of the employment interview over the past 100 years has been phenomenal.

Several commentators (Arvey and Campion, 1982; Eder and Ferris, 1989; Harris, 1989; Jelf, 1999; Mayfield, 1964; Moffatt, 1969; Schmitt, 1976; Ulrich and Trumbo, 1965; Wagner, 1949; Wright, 1969) have tracked the research developments in the realm of employment interviewing during the course of the past century. Their reviews cumulatively offer a holistic insight into the total complement of the research on this ubiquitous selection tool. Using the holistic perspective derived from these reviews as a fundamental point of departure, the author intends to explore the dominant trends in employment interview research during the 20<sup>th</sup> century. In addition, the major criticisms that have been levelled against the research within each of these trends will be highlighted.

#### 2.1 An Overview of the Dominant Trends in Employment Interview Research

Three significant trends appear to have characterised research on employment interviews over the past 100 years. These trends were located within three sequential yet capacious time periods. In the author's view, the first trend extended roughly from the early 1900s to the early 1950s and was typified by investigations of the validity and reliability associated with unstructured or traditional employment interviews. The second trend, which was characterised by an influx of research on decision-making processes within employment interviews, spanned roughly from the early 1950s to the early 1980s. The final trend was typified by a resurgence of interest in the psychometric aspects of validity and reliability in relation to the employment interview. However, in this trend, the emphasis was on the investigation of specific variants of structured employment interviews (such as patterned behaviour description interviews, situational interviews, highly structured interviews and structured behavioural interviews) and meta-analytic investigations, which cumulated the findings of these specific interview

variants to yield general insights on structured employment interviews. The third trend extended from the early 1980s to the end of the 20<sup>th</sup> century.

The author has conceptualised these significant trends as having been dominant during specific time periods in the history of employment interview research. It is acknowledged that they cannot be confined exclusively to these time periods, since research on each of these major trends has been conducted, in one form or another, throughout the 20<sup>th</sup> century. In the author's opinion, these three trends aptly capture the essence of the research that was conducted on the employment interview over the past 100 years.

### 2.1.1 A Theoretical Conceptualisation of the Major Trends

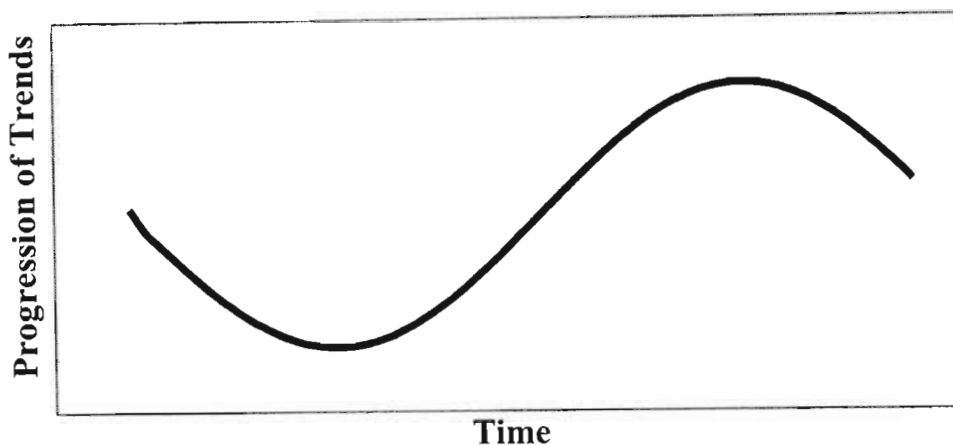
The author perceives trends as dynamic phenomena that evolve over time. This evolution occurs in the context of a specific and distinctive pattern that conforms to the contours of the sigmoid curve. The concept of the sigmoid curve was popularised by Charles Handy (1994) in his book entitled "The Empty Raincoat" in which it was identified as a mechanism with which the paradoxes that characterise the multiple facets of human life and society may be managed. According to Handy (1994, p. 50) "the sigmoid curve sums up the story of life itself". This sentiment captures the immense potential that a seemingly simple concept has for making sense of the characteristic ups and downs and highs and lows of practically every facet of life. In the light of its far-reaching applicability, the author envisages the concept of the sigmoid curve to be invaluable in tracing, exploring and understanding the progression of the major trends that have been identified in employment interview research. The waxing and waning of these trends are, in the author's opinion, judiciously captured in the ascending and descending crests of the sigmoid curve. An illustration of the sigmoid curve is provided in Graph 2.1. This graph has been adapted from an illustration offered by Handy (1994).

### 2.1.2 A Curvilinear Perspective of the Major Trends

The author conceptualised the three major trends that appear to have emerged from the vast array of research on the employment interview as a series of sequential sigmoid curves. These are depicted in Graph 2.2.

The first trend in employment interview research is characterised by investigations into the validity and reliability associated with unstructured interviews. This wave of research was born out of a need to scientifically affirm the effectiveness of these interviews in the selection arena. However,

Graph 2.1 The Sigmoid Curve

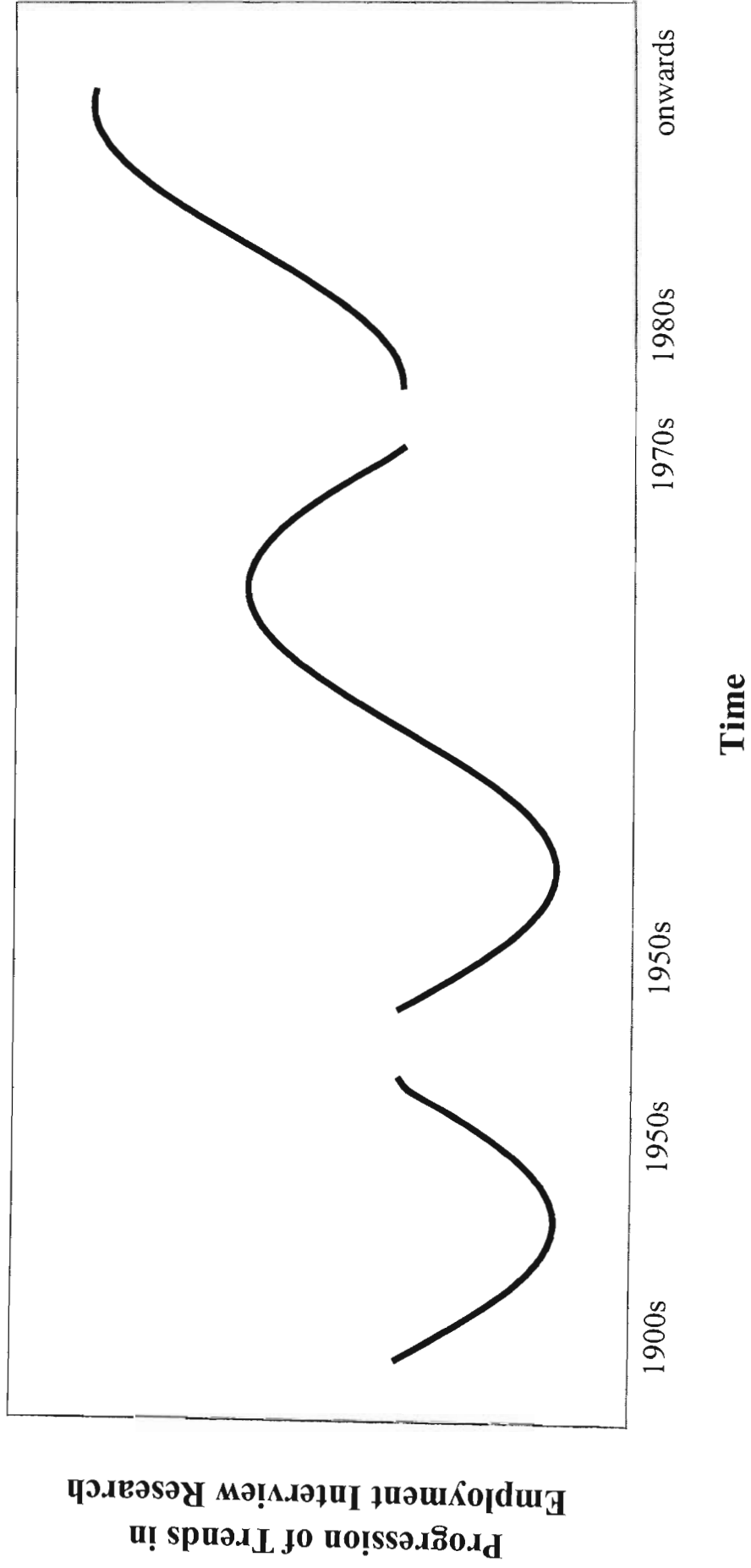


(Handy, C. (1994) The Empty Raincoat: Making Sense of the Future. Sydney: Arrow Business Books, p. 50).

the validity and reliability estimates yielded by these studies were contradictory, questionable, and all in all, disappointing (Wagner, 1949). These contentious findings are captured in the initial downward swing of the sigmoid curve that pertains to this trend. In the context of this disillusionment, however, numerous recommendations were made on how to improve the psychometric characteristics of reliability and validity associated with the employment interview. Unfortunately, it seems that instead of systematically incorporating and integrating these recommendations into subsequent research efforts, employment interview researchers shifted their attention away from research in this domain due to these discouraging findings. Thus, the sigmoid curve that depicts the first trend in employment interview research remains incomplete.

Employment interview researchers directed their efforts towards the investigation of the interview process instead, with particular emphasis on how decisions are made in employment interviews. Thus, the second major trend in employment interview research was born. This trend was characterised by investigations that touched on a wide range of issues pertaining to interview process and decision-making. Initially there was great uncertainty regarding the value of research of this nature (Wright, 1969). However, investigations on interview process and decision-making appear to have flourished between the early 1950s and the early 1980s. This interest seems to have continued through to the 1990s (Jelf, 1999). From the 1980s onwards, though, the second trend in employment interview research seems to have been over-shadowed by a renewed interest in the investigation of the validity and reliability associated with structured employment interviews.

Graph 2.2 A Curvilinear Perspective of the Dominant Trends



This is illustrated in Graph 2.2 as a decline in the sigmoid curve that pertains to the second trend. One could argue that this representation is not entirely accurate, since research pertaining to this trend did, in fact, continue into the 1990s. However, in the author's conceptual scheme, the downward swing of the sigmoid curve represents a decline in the dominance of this trend rather than a complete abandonment of research pertaining to it. Research on the validity and reliability associated with structured employment interviews appears to have dominated during the 1980s and 1990s. These studies appear to have yielded consistently positive and promising findings. Hence, the sigmoid curve for this trend is not characterised by an initial downward motion. It is evident from earlier reviews (Mayfield, 1964; Moffatt, 1969; Ulrich and Trumbo, 1965; Wagner, 1949) that a pertinent recommendation, which emerged from research on the reliability and validity associated with unstructured interviews, was to enhance their structure. In the third trend, this recommendation seems to have been heeded and incorporated into research efforts on the validity and reliability associated with structured employment interviews. Therefore, the third trend may be conceived of as a continuation of the first trend in employment interview research. The sigmoid curves that illustrate the first and third trends are designed to convey this impression. However, for the sake of conceptual clarity, chronology and as an aid to the present discussion and this research endeavour in its entirety, the author will regard them as two distinct trends.

## 2.2 An In-Depth Exploration of the Dominant Trends

In this section, a holistic review of the investigations that characterised each of the trends in employment interview research during the 20<sup>th</sup> century will be provided. The primary objectives of this discussion are three-fold. The first aim is to capture the nature of the research that was conducted in the context of these trends. Secondly, the major conclusions that were yielded by these investigations will be highlighted. Thirdly, the major criticisms that may be levelled against the research within each trend will be explored.

The discussion of the first trend will be based on research on the validity and reliability of unstructured interviews. In the discussion of the second trend a brief commentary on the research on decision-making processes in employment interviews will be offered. The discussion of the third trend will encompass an exploration of the validity and reliability research on structured employment interviews.

## 2.2.1 The First Trend: Research on Unstructured Interviews

In this discussion a definition for unstructured employment interviews will be sought, the main highlights of research in the first trend will be reviewed and the major criticisms of research on unstructured employment interviews will be explicated.

### 2.2.1.1 In Search of a Definition for Unstructured Interviews

Despite the substantial interest that was afforded to unstructured employment interviews in research circles, there appears to have been no attempt on the part of employment interview researchers to offer a unified definition of the concept that constituted the pivotal focus of their investigations. Thus, there is no singular and widely accepted definition for the unstructured employment interview. The author will attempt to define this concept by highlighting the distinguishing characteristics of the unstructured interviews investigated in some of the studies that constituted the first trend in employment interview research.

Employment interviews, in general, have been defined as interpersonal dialogues between one or more organisational representatives and one or more job applicants (Dipboye, 1992; Eder, Kacmar and Ferris, 1989). They provide a forum within which the individuals engaged in the interpersonal interaction can evaluate each other in the process of making significant employment decisions. Employment interviews allow employers and applicants alike to elicit vital information about each other in order to assist the former to decide whether or not the latter will suit their organisational needs and to assist the latter to decide whether or not the organisation will suit their personal and career goals. This generic conception of the employment interview is all-embracing and accommodates all its variants at a very fundamental level. However, the existence of different variants of employment interviews is indicative of an evolution in the nature of these selection tools. Thus, while unstructured employment interviews conform to the parameters of this generic definition, they possess very specific qualities that distinguish them from other variants of employment interviews.

An employment interview may be regarded as unstructured if it is not executed and evaluated in a standardised manner by interviewers, among all the individuals that have applied for a job. Typically, this lack of standardisation is characterised by a lack of consistency in the questions that are posed to different applicants for the same job and by the absence of a standard format for the implementation and evaluation of the interview itself (Taylor and O'Driscoll, 1995). The questions that are posed to applicants in unstructured employment interviews are usually based entirely on the discretion of the interviewers. These questions are typically guided by interviewers' initial impressions of applicants on



the basis of preliminary information, their personal preferences, beliefs and values (Taylor and O'Driscoll, 1995). Their individual outlooks on the positions in question and the characteristics and qualities they believe that applicants should possess in order to be successful within them, may also serve as guides to the questions that interviewers pose. This suggests that the questions, which are ultimately included in these interviews, are, in general, not based on the results of systematic and scientific job analysis procedures. Hence, the content they are designed to elicit is usually not job-related. Instead, employment interviewers appear to be preoccupied with information pertaining to the personal attributes of applicants such as their personality, physical characteristics and non-verbal behaviour. This subjective information, which usually has no direct bearing on the position in question, constitutes the basis on which interviewers render their employment decisions in unstructured interviews. These decisions are generally made in the absence of a standard method or format for evaluating the suitability of applicants. In the light of these distinguishing features of unstructured employment interviews, it is not surprising that they require minimal preparation and are well suited to being conducted in an ad hoc manner.

#### 2.2.1.2 The Main Highlights of Research in the First Trend

The author identified the following as significant issues in the research on unstructured interviews, which characterised the first trend in employment interview research:

- The primary focus of the research on the psychometric qualities of validity and reliability;
- An insight into what unstructured interviews were originally designed to measure;
- The major insights derived from research initiatives on unstructured interviews;
- The emergence of peripheral research interests during the context of the first trend; and
- Further developments in research on unstructured employment interviews outside the context of the first trend.

##### (i) The Main Focus of the Research

Unstructured interviews constituted the focal point of the majority of research efforts on the employment interview during the first half of the 20<sup>th</sup> century. The principal impetus for these investigations was to prove the value and usefulness of the unstructured interview in the selection arena (Wagner, 1949). To this end, employment interview researchers oriented their investigations towards the estimation of the validity and reliability associated with these interviews.

## (ii) What Employment Interviews Originally Measured

In early investigations of the reliability and validity associated with unstructured employment interviews, the primary purpose of the interview was to assess specific constructs such as intelligence and personal characteristics and traits (Wagner, 1949). In general, these constructs did not appear to be directly related to the requirements of the jobs in question (Wagner, 1949). For example, in the selection of teachers, applicants were assessed on personal constructs such as beauty and conceit amongst others (Corey, as cited in Wagner, 1949). In the author's opinion, these qualities certainly do not seem to have any direct bearing on an individual's ability to be effective as a teacher.

While the assessment of apparently non job-related constructs appears to have been the norm in initial studies on the employment interview, Wagner (1949) noted the assessment of job-related constructs in a study reported by O'Rourke. This researcher evaluated the traits of judgement and resourcefulness, which were believed to be vital for and relevant to the job of a prohibition officer (Wagner, 1949). The relevance of these traits was established on the basis of the results yielded by a job analysis procedure (Wagner, 1949). The manner in which the employment interview, used in this study, was designed and conducted and the nature of the questions that were included in it, resembled a specific structured variant of employment interview, known as the situational interview, in the author's opinion. Thus, it seems that this study was prophetic insofar as it held the key to the future of the employment interview. Although the study was not validated and the value of the principles and procedures it espoused was not recognised or acknowledged by the majority of researchers at the time, it is undeniable that it yielded critical notions pertaining to employment interviews. Thus, the concepts of job-relatedness and structure appear to have been contemplated from very early on in the 20<sup>th</sup> century. Despite evidence of such progressive ideas early in the history of employment interview research, the process of evolution that this selection tool has undergone has been slow.

In an attempt to optimise the value of unstructured employment interviews, researchers investigated the constructs that could be assessed most effectively in them, in their investigations of the reliability and validity of these selection tools. Mayfield (1964) noted that the construct of intelligence could be validly and reliably estimated from employment interviews. However, Wagner (1949) postulated, in an earlier review, that this construct could be assessed more accurately and effectively using other mechanisms such as intelligence tests. This line of inquiry yielded two constructs that researchers believed could be effectively evaluated using employment interviews (viz. personal relations and motivation to work) (Ulrich and Trumbo, 1965). These constructs were perceived as directly relevant to a wide range of jobs and were believed to be well suited to being examined in the context of the employment interview (Ulrich and Trumbo, 1965).

However, the unstructured employment interviews evaluated in early research were not confined solely to the assessment of specific constructs. Wagner (1949) cited early studies in which interviews were used to ascertain applicants' overall ability or proficiency to be successful on the job. For example in 1915 and 1916 respectively, Scott investigated the reliability and validity associated with employment interviews which were designed to assess the sales ability of interviewees (Mayfield, 1964; Wagner, 1949). In these interviews, interviewers were charged with the task of formulating global estimations, on the basis of information that they deemed to be job-related, in order to render decisions about applicants' general abilities to perform effectively in the job in question (Mayfield, 1964).

### (iii) Significant Insights on Research on Unstructured Interviews

The most significant finding that emerged from research on unstructured employment interviews was that the validity and reliability estimates associated with them were disappointingly low (Mayfield, 1964; Moffatt, 1969; Schmitt, 1976; Wagner, 1949). This discouraging finding ought to have cast serious doubts on their value as selection tools. However, notwithstanding this seemingly serious blow to their reputation, these employment interviews retained their popularity in the selection arena. Wagner (1949) suggested that a possible reason for their continued popularity is the unique opportunity they provide for employers and applicants to interact on a person-to-person level with each other and in so doing to appease a fundamental human curiosity.

Despite the fact that early research on unstructured employment interviews did not yield promising results, these studies did lay the groundwork for valuable recommendations pertaining to how employment interviews could have been improved upon to enhance the psychometric results associated with them. In Mayfield's (1964) review of the research on the employment interview, it was noted that more structured and standardised interviews would yield improved validity and reliability coefficients. However, some reviewers (Moffatt, 1969; Wright, 1969) proposed that improvements in structure only lead to significant improvements in the reliability results associated with employment interviews. Moffatt (1969) argued that the mere act of structuring interviews was not adequate to raise the validity estimates associated with them to acceptable levels. He proposed that the scope of the information that interviews are designed to elicit is a determinant of validity (Moffatt, 1969). Ulrich and Trumbo (1965), on the other hand, did not comment on the relationship between structured interviews and reliability but maintained that the highest validities were observed with structured interviews. Thus, although there does not appear to have been consensus pertaining to the exact benefits of structured interviews, they were recognised as tools that possessed the potential to initiate advancements in the realm of employment interviewing.

A structured interview format is characterised by a standard set of interview questions that are arranged and posed to applicants in a consistent manner and is accompanied by a standard rating methodology to evaluate applicants' responses. This format is likely to enhance the reliability results associated with employment interviews, since the use of standardised techniques could assist interviewers to derive consistent information from all their interviewees and to evaluate this information in a consistent manner.

The author subscribes to the sentiment, advocated by Moffatt (1969), that merely enhancing the structure of employment interviews is not sufficient to raise the validity coefficients associated with them. Unlike reliability, which is deemed to be an intrinsic property of employment interviews and other selection tools, (Gatewood and Field, 1994; Muchinsky, Kriek and Schreuder, 1998), validity is a property of the inferences made on the basis of the information derived from the use and evaluation of employment interviews, in particular, and selection tools, in general (Cascio, 1991; Gatewood and Field, 1994). In the selection arena, validity is determined by correlating the results derived from selection tools with appropriate and relevant job performance data. In order for the correlation exercise to be worthwhile, the variables that are correlated should bear a direct relationship to the job in question. It is logical to conclude, therefore, that if employment interviews are designed to elicit information that offers insights into the potential that applicants possess to function effectively within a job, then the inferences that interviewers make about the suitability of applicants, on the basis of this information, are likely to be more valid than the inferences that are likely to be made in the absence of this information.

Wagner (1949) latched onto the importance of the notion of job relevance when he proposed that constructs should only be included for evaluation in employment interviews if they are directly related to the success of applicants in the jobs for which they are being considered. This could have been accomplished by conducting systematic job analyses of the positions being selected for and formulating interview questions on the basis of the job-related information that was derived from them. In general, however, this advice was not incorporated into investigations by mainstream employment interview researchers at the time. It was only much later on in the 20<sup>th</sup> century that researchers began to incorporate the notion of job-relevance into their work on employment interviews.

In the discussion of what employment interviews were originally designed to measure, the author introduced a belief that was advocated by many researchers (Moffatt, 1969; Ulrich and Trumbo, 1965; Wright, 1969), which suggested that two specific constructs (viz. personal relations and career motivation) could be assessed most effectively in employment interviews. In addition, the evaluation of these constructs appeared to have contributed significantly to improvements in interviewer decisions

and to the validity coefficients associated with employment interviews (Moffatt, 1969; Ulrich and Trumbo, 1965). This led researchers to conclude that by circumscribing the scope of employment interviews to these specific constructs, improvements in validity would be noted. In the author's view improvements in validity could also be attributed to a range of other factors. These include the job relevance of the constructs that are assessed and the questions that are used to assess them.

A further blow to the validity associated with unstructured employment interviews was dealt when researchers noted that the contribution of the information derived from interviews was negligible in employment decisions when accompanied by information derived from other sources such as cognitive ability tests and credentials (Ulrich and Trumbo, 1965). This finding inevitably cast doubt on the value of the interview in general and the potential it possessed for yielding useful information for the purpose of basing valid inferences about applicants upon. This finding suggested that when used in conjunction with other selection tools, the incremental validity of the employment interview was low.

#### (iv) Peripheral Research Interests During the First Trend

The second trend in employment interview research was characterised by investigations on how decisions were made in employment interviews. However, in literature on the employment interview that predated this trend, there is evidence of an interest in decision-making. In one of the earliest reviews of the literature on employment interviews, Wagner (1949) posed the question of whether or not interviewers could effectively integrate and synthesise all the information they gleaned from employment interviews in order to render optimal employment decisions that could rival the decisions rendered on the basis of statistical manipulations of interview data. Wagner's (1949) review of the research, in relation to this issue, did not permit him to arrive at an unequivocal conclusion regarding the capabilities of human beings to effectively integrate the information they derived from employment interviews in order to make critical employment decisions. This early interest in the realm of decision-making paved the way for the second trend in employment interview research, which will be explored later on in this chapter.

In the research on the validity and reliability associated with unstructured employment interviews, the notion of panel or board interviews was introduced in studies reported by Fearing in 1942, Bingham in 1939, Driver in 1944 and Rundquist in 1947, which were cited in Wagner (1949). The validity and reliability findings yielded by these studies did not appear to differ significantly from the findings obtained for unstructured employment interviews in general. Typically, the unstructured interviews investigated in the first trend in employment interview research consisted of one-on-one interactions between interviewers and applicants. Hence, the use of panels of interviewers, in research during the

first half of the 20<sup>th</sup> century, constituted a drastic deviation from the norm. However, in subsequent years, during what the author has termed the third trend in employment interview research, there was a renewed interest in panel interviews in the research on structured employment interviews.

From as early as the 1930s, research on standardised interviews was underway and promising validity and reliability results were yielded (Wagner, 1949). Despite this breakthrough in the realm of employment interviewing, mainstream researchers seemed to pay little attention to the real benefits that standardised or structured interviews had to offer to the selection process. Thus, notwithstanding the introduction of the notion of standardised interviews early in the first half of the 20<sup>th</sup> century, it took employment interview researchers over four decades to realise their value in the selection arena. It was only in the early 1980s that systematic attempts were noted to devise and investigate structured employment interviews. These research efforts will be highlighted in the discussion of the third trend in employment interview research.

#### (v) Later Research on Unstructured Interviews

In the early days of employment interview research during the first half of the 20<sup>th</sup> century, Mayfield (1964) noted that more studies were oriented towards the analysis of reliability than validity. Later studies, in which unstructured employment interviews featured, were primarily oriented towards the investigation of the validity estimates associated with a range of alternative predictors of job performance. The main purpose in these studies was to comment on how the validity results of these alternative predictors compared with the validity results associated with cognitive ability tests, which were regarded as the most effective tools for predicting future job and training success (Reilly & Chao, 1982; Hunter and Hunter, 1984; Schmidt and Hunter, 1998). In addition, these investigations included insights into the notions of fairness (Reilly and Chao, 1982) and utility (Hunter and Hunter, 1984; Schmidt and Hunter, 1998) as they pertained to a range of alternative predictors. All these studies employed an interesting statistical technique known as meta-analysis. Using this technique, these researchers cumulated the validity findings yielded by several independent studies to provide combined estimates for the validities associated with a range of selection techniques including employment interviews.

The average validity estimates that these meta-analytic investigations yielded for employment interviews were disappointing. Reilly and Chao (1982) calculated an average validity estimate of 0.19 while Hunter and Hunter (1984) calculated an average validity coefficient of 0.14. In both of these investigations, supervisory ratings were used as the criterion measure against which employment interview results were validated (Hunter and Hunter, 1984; Reilly and Chao, 1982). Although the later

meta-analysis conducted by Schmidt and Hunter (1998) yielded a more promising average validity for the unstructured employment interview (0.38), these researchers conceded that this bolstered result could have been attributed to carefully conducted interviews. They argued that had these unstructured interviews been conducted carelessly, the average validity result would undoubtedly have been considerably lower (Schmidt and Hunter, 1998).

Schmidt and Hunter (1998) estimated that when used in conjunction with cognitive ability tests, unstructured employment interviews contributed a 0.04 increment in validity, which translated into an 8% increase in validity. In comparison with the incremental validity reported for the use of structured interviews in conjunction with cognitive ability tests (0.12 or 24%), the incremental validity of unstructured employment interviews is low.

Cumulatively, the results yielded by these later investigations corroborated the results of earlier investigations, which exposed the poor validity estimates associated with unstructured employment interviews and their low incremental validity.

### 2.2.1.3 An Appraisal of Research on Unstructured Interviews

Early in the 20<sup>th</sup> century, the majority of the studies that focused on employment interviews were not empirical and scientific in nature. Instead, they took the form of 'how-to' literature and presented a range of unsubstantiated and often conflicting opinions on how to conduct interviews. Wagner (1949) reported that of the 106 articles he identified for the purpose of his review of the literature on employment interviews, only 25 were experimental in nature. Mayfield (1964) noted the same ratio, of approximately one experimental study to four articles that presented opinions, when he reviewed over 300 articles pertaining to the employment interview. This state of affairs suggests that employment interviews were used extensively during the first half of the 20<sup>th</sup> century despite limited empirical evidence of their utility as selection tools.

The empirical studies, which focused primarily on estimating the reliability and validity results associated with unstructured employment interviews, were plagued by a number of methodological constraints. These flaws were a manifestation of the way in which the individual studies, within this broad category of studies, were designed.

On the basis of reviews of the literature on employment interviews (Mayfield, 1964; Wagner, 1949) a general tendency was discerned with regard to the sizes of the samples that were used to yield the validity and reliability estimates associated with unstructured interviews. Essentially, the samples used



in research of this nature were small. On the subject of sample sizes, Dyer (1995) proposed that larger samples are preferable in order to obtain accurate research results. He argued that in every sample there will be a few individuals that contribute extreme scores to the data amongst a majority of individuals that contribute moderate scores (Dyer, 1995). In a large sample, these extremities can be cancelled out by the moderate scores to yield an accurate view of the typical results that were obtained (Dyer, 1995). In smaller samples, however, the impact of extreme scores is pronounced and a skewed average or typical result is obtained (Dyer, 1995). In the context of research on the unstructured employment interview, therefore, small sample sizes could have resulted in undue and inaccurate emphasis being placed on extremely low reliability and validity coefficients, thereby, rendering disappointingly low overall estimates.

The research oriented towards ascertaining validity of the unstructured interview, is prone to the effects of the methodological phenomenon of restriction of range. Wagner (1949) derived this insight in his early review of the research on employment interviews. Typically, when employment interviews are validated using actual job applicants, the individuals in the samples under investigation, are screened using other less resource-intensive tools before being exposed to the employment interview in the selection process. This implies that the individuals ultimately interviewed, are those that have performed satisfactorily on previous selection devices during the process of selection. In effect, the range of the initial sample of applicants is restricted by using other selection tools as hurdles. This leads to a situation in which only those applicants that are perceived in a positive light by selectors are exposed to employment interviews. In order to validate the scores obtained by applicants in the interview, these scores are correlated with a criterion measure, which represents a snapshot of actual job performance. This correlation exercise, for the purpose of validation, in a sample whose range is restricted, will typically yield a weak association between the variables in question (Heiman, 1998). This phenomenon could invariably have contributed to the low validity coefficients that have been yielded for unstructured employment interviews.

Ulrich and Trumbo (1965) commented on the methodological flaws that characterised the studies that Wagner (1949) included in his review of the literature on the employment interview. They argued that some of these studies included unrealistic experimental conditions. For example, twelve sales managers were required to interview and rank 57 applicants according to their suitability for the job in question in a study reported by Hollingworth (as cited in Wagner, 1949). Ulrich and Trumbo (1965) suggested that it was unrealistic to expect interviewers to rank 57 applicants. Wilkins (as cited in Ulrich and Trumbo, 1965) offered evidence that suggested that interviewer predictions based on three-minute interviews did not contribute constructively to the identification of recruits that were destined to make little or no contribution to the army. In the author's opinion, it does appear unrealistic and



unreasonable to expect interviewers to render predictions after a mere three minutes of interviewing applicants. In general, unrealistic experimental conditions can impact significantly on the interview results that are obtained and this, in turn, could impact on the validity and reliability estimates that are calculated on the basis of these results.

In their critique of the studies reviewed by Wagner (1949), Ulrich and Trumbo (1965) noted a second methodological flaw, which pertained to criterion contamination. They argued that this contamination arose when criterion ratings were made after knowledge was obtained about interviewers' ratings. The result of this flaw is erroneous and inaccurate criterion ratings, which invariably impact on the validity coefficients that are rendered. It is clear, therefore, that studies that use contaminated criterion measures do not yield accurate validity findings.

The author identified a further criticism that may be levelled at the criterion measures used in some of the validation studies on unstructured employment interviews, on the basis of a review of relevant research by Wagner (1949). The actual criterion measures used in the studies that were commented on, were often not specified. For example, Rundquist (as cited in Wagner, 1949) did not divulge the criterion that he used to conduct his validation study. In circumstances such as this one, where insufficient information about the criterion is provided, it is often not known whether these measures were statistically valid and reliable. Further, readers are not offered the opportunity to formulate independent opinions of the face validity of these measures in the course of their critical appraisals of validation studies of this nature. In the absence of statistical and general credibility, these criterion measures may be called into question together with the validity results based on them.

Mayfield (1964) argued that the criterion measures used in validation research on the unstructured employment interview had the potential to lead to difficulties in the interpretation of the results that were ultimately yielded. These difficulties could essentially be attributed to the practices of validating trait ratings with overall measures of job performance instead of with criterion measures that pertained to these specific traits and of using criterion ratings assigned by friends instead of objective observers (Mayfield, 1964). The use of these questionable criterion measures invariably casts doubt on the validity results yielded by investigations of the unstructured employment interview.

A final methodological criticism that Ulrich and Trumbo (1965) levelled against the research that Wagner (1949) reviewed was interview contamination. According to this methodological transgression, interviewers obtained information about their interviewees' performance on the criterion measures before they rendered their interview decisions (Ulrich and Trumbo, 1965). This implies the

possibility that the prior information about performance on the criterion measure could have biased the interview results and could, consequently, have tainted efforts to estimate validity and reliability.

In the experimental designs within which unstructured employment interviews were investigated, the interview itself played a significant role in the final results that were obtained. The main criticisms that were levelled against unstructured interviews have been touched on in previous sections and will be mentioned only briefly in this discussion. Firstly, the employment interviews used in early research were criticised for their lack of organisation and structure (Mayfield, 1964; Moffatt, 1969; Ulrich and Trumbo, 1965; Wagner, 1949). Secondly, they were denounced because they were devoid of job-relevant content (Wagner, 1949) and thus, were not well equipped to elicit job-related information from interviewees (Ulrich and Trumbo, 1965).

During the exploration of the literature on unstructured employment interviews, the author identified general criticisms that may be levelled against the entire body of research on these interviews. In particular, these criticisms are pertinent to validation research. Mayfield (1964) proposed that one of the most significant shortcomings of studies on the unstructured employment interview was that they were difficult to compare. This lack of comparability was attributed to various factors including the heterogeneous purposes these studies were designed to fulfil, the diverse populations they tapped to obtain their samples, their unique settings and contexts and the vastly different information they elicited (Ulrich and Trumbo, 1965). The realities were exacerbated by the absence of a unifying definition of the unstructured employment interview. Cumulatively, these factors led to a situation in which it was difficult, if not impossible, to obtain information, pertaining to the unstructured employment interview, that was sufficiently generalisable to contribute to the database of knowledge on these selection tools. In effect, therefore, the findings yielded by research on this early variant of employment interview was situationally-specific and, by implication, could not be generalised to all unstructured employment interviews (Mayfield, 1964; Moffatt, 1969; Ulrich and Trumbo, 1965; Wagner, 1949).

One of the most fundamental and, perhaps, most radical criticisms that was levelled against validation research on unstructured employment interviews is that traditional research designs tended to underestimate the validities associated with these interviews (Dreher, Ash and Hancock, 1988; Gehrlein, Dipboye and Shahani, 1993). Essentially, Dreher et al. (1988) argued that in order to obtain average validity coefficients in traditional validation research, the data yielded by multiple interviewers is collapsed without offering consideration to the inevitable reality that individual interviewers differ in their abilities to render effective judgements and in their constant tendencies to make favourable and unfavourable ratings. This approach to the investigation of validity clearly does not take cognisance of

individual differences in interviewer competence and rating styles, thereby yielding biased, erroneous and underestimated validity results when aggregated data is used to ascertain the utility of these interviews in the selection arena (Dreher et al., 1988). However, these researchers did not ground their conclusions in empirical evidence.

Zedeck, Tziner and Middlestadt (1983) embarked on an empirical investigation to examine individual differences among interviewers. Their findings revealed individual differences in relation to the manner in which interviewers used and evaluated information in order to render their employment decisions (Zedeck et al., 1983). On the basis of this finding, these researchers proposed that it is reasonable to anticipate variability in the success with which interviewers make decisions or interviewer validity (Zedeck et al., 1983). This finding was corroborated in a later study by Dougherty, Ebert and Callender (as cited in Harris, 1989) in which the aggregated analysis of data across three interviewers did not yield significant results while an individual analysis of the results obtained by these interviewers did render significant correlations. This study provided the most substantial support for the proposition that some interviewers are more valid than others (Harris, 1989). On the basis of these findings, Arvey and Campion (1982) argued that traditional validation studies had focused on the wrong theory. Instead of ascertaining the validity associated with interviewers, these studies were oriented towards the determination of the validity associated with interviews (Arvey and Campion, 1982).

Gehrlein et al. (1993) conducted a study using interviews with unstructured formats, which were evaluated according to structured rating forms. These researchers demonstrated that the validity coefficients yielded by correlating interview scores with criterion results using traditional validation techniques were not significant when compared to the statistically significant results obtained via the use of alternative statistical techniques known as the Dreher and Kenny techniques. On the basis of these somewhat harsh criticisms, the validation strategies that were used to yield validity estimates of the unstructured employment interview have been called into question. However, the empirical evidence in support of these criticisms is extremely limited. Hence, it is not feasible to conclude that traditional validation strategies and techniques are as deficient as these contentions suggest they are.

## 2.2.2 The Second Trend: Research on Decision-Making in Interviews

In the light of the disappointing results yielded by macroanalytic research that pertained to the validity and reliability of employment interviews, researchers turned their attention to microanalytic research efforts (Wright, 1969). In so doing, they hoped to derive useful information regarding the use of employment interviews in the selection arena (Wright, 1969). The main thrust of microanalytic

research was on understanding the interview process with a primary emphasis on how interviewers render employment decisions. An in-depth discussion of the specific findings that were yielded by research that characterised this trend would be too voluminous in this discussion. Therefore, the author will merely touch on some of the key issues that were investigated within this trend and will highlight the main criticisms that were levelled against research of this nature.

Wright (1969) noted that a comprehensive research effort was undertaken by Webster and his colleagues on decision-making in the context of employment interviews at the McGill University. This initiative served as the major impetus for the proliferation of related research endeavours during the second trend in employment interview research. The primary focus of the research in this trend is on how interviewers use the information (both visual and aural) that they glean from interviews in order to render decisions. In the decision-making process, several factors may affect the ultimate judgements that are rendered. These include interviewer perceptions and biases, the manner in which interviewers process information, interview process factors (such as the length of the interview and the manner in which the interview is conducted) and active attempts by applicants to influence the perceptions of interviewers (impression management).

Due to the deviation of research on decision-making from the notions of validity and reliability, and the focus of the present study on an analysis of the research pertaining to the validity and reliability findings that were obtained for structured variants of employment interviews, the author will not elaborate on the major findings that investigations pertaining to decision-making have yielded. For a general insight into the research on decision-making processes, the reader is referred to the reviews of the literature on employment interviews by Arvey and Campion (1982), Harris (1989), Jelf (1999), Schmitt (1976) and Wright (1969).

A fundamental criticism that may be levelled against research on decision-making processes, in the context of employment interviews, is that the findings yielded by these investigations have essentially been fragmented and unintegrated. Wright (1969) argued that microanalytic research was so fragmented that it tended towards being meaningless. In an attempt to rectify this lack of integration and gross fragmentation, Schmitt (1976) proposed the first model that logically accommodated and associated the factors that impacted on the outcomes of employment interviews. This model was causal in nature. However, Schmitt (1976) could only speculate about the causal relations that he identified in the absence of empirical evidence to substantiate his claims. This initial attempt to impose some order on the information pertaining to the manner in which decisions are made in employment interviews sparked off several other attempts to model the decision-making process and the factors that influenced it (Arvey and Campion, 1982; Eder, Kacmar and Ferris, 1989).

Like the research on unstructured employment interviews, the research on decision-making processes extended beyond the period within which the second trend was conceptualised for the purpose of this discussion. The reviews of the literature on employment interviews by Harris (1989) and Jelf (1999) attest to this. Jelf (1999) noted some of the issues that subsequent research efforts have focused on, in addition to the conventional issues pertaining to decision-making processes in employment interviews. These included:

- The use of decision-making theories and models to understand and explore the decision-making task that selection interviewers undertake; and
- The practice of rendering recruitment decisions based on the fit between the individual and the organisation as an adjunct to the examination of the decision processes used to render employment decisions on the basis of the applicant's suitability for the position in question.

### 2.2.3 The Third Trend: Research on Structured Employment Interviews

This discussion will encompass early perspectives on structured interviews, an insight into early research on these interviews, an attempt to define structured interviews and an insight into the major contributions of and criticisms against research on these interviews.

#### 2.2.3.1 Early Perspectives on Structured Interviews

The 1980s ushered in an entirely new era in employment interview research during which investigations on structured employment interviews predominated. However, some early commentaries by Janz, Hellervik and Gilmore (1986), Mayfield (1964), Moffatt (1969), Ulrich and Trumbo (1965), and Wagner (1949), to name only a few, recognised the existence of the notion of structured employment interviews in early investigations during the first half of the 20<sup>th</sup> century. These interviews were described in a variety of different ways in early research efforts, ranging from patterned and standardised to systematic, designed, guided and structured (Ulrich and Trumbo, 1965; Wagner, 1949).

The early research on structured employment interviews appeared to have encompassed interviews that varied in terms of the level of structure they possessed (Yonge, 1956). In addition, there does not seem to have been a singular, well thought out definition for structured employment interviews. In the author's opinion, despite these difficulties, the mere fact that the notion of an employment interview variant, which deviated from the popular unstructured interview variant at the time, was tolerated,

signalled an openness and a commitment to understanding and improving employment interviews in the selection arena. The favourable results and distinct advantages associated with structured employment interviews may have been overshadowed by their unstructured counterparts during these early years (Janz et al., 1986; Wagner, 1949). However, from the 1980s through to the end of the 20<sup>th</sup> century, the impact and contribution of structured interviews in the realm of employment interview research appears to have been recognised and appreciated.

### 2.2.3.2 Early Investigations of Structured Interviews

The early research initiatives on structured employment interviews were primarily oriented towards the determination of validity and reliability. In general, the validity and reliability estimates associated with these interviews were promising. Hovland and Wonderlic (cited in Wagner, 1949) reported a reliability coefficient of 0.71 and promising validity results in an investigation of a standardised interview that they developed. McMurray (cited in Wagner, 1949) researched the patterned interview and reported validity estimates of 0.68 and 0.61 (these estimates were derived from the correlation of interview scores and foremen's evaluations). Yonge (1956) explored the validity associated with the patterned interview. This study yielded positive validity evidence for these interviews with correlations between supervisors' ratings and interviewers' scores ranging from 0.45 to 0.99 across the five groups of applicants that were interviewed.

The structured employment interviews, used in the studies outlined above, provided valuable insights into the key features of these early structured interviews. Firstly, these interviews appeared to have been designed to cover pre-determined topics. For example, the standardised interview proposed by Hovland and Wonderlic (cited in Wagner, 1949) encompassed work history, family history, social history and personal history. In relation to patterned interviews, McMurray (cited in Wagner, 1949) noted that interviewers had a plan, which guided the questions they posed to applicants. A second characteristic of these interviews was that they bore a direct relationship to the requirements of the job. In the standardised interview that Hovland and Wonderlic (cited in Wagner, 1949) researched, the items on the interview guide were scored in terms of their relevance to the job in question. In the context of patterned interviews, McMurray proposed that interviewers proceeded on the basis of definite job specifications (Wagner, 1949). The third distinguishing characteristic of early structured interviews was their reliance on pre-determined scoring systems for the evaluation of the information that interviewers gleaned from applicants during the interview. Hovland and Wonderlic used a simple scoring system to evaluate the responses obtained on their standardised interview (Wagner, 1949). This system required interviewers to assign a positive or negative value to each item on the interview guide depending on its relevance to the job before tallying them to obtain overall scores (Wagner,

1949). In McMurray's validation studies on the patterned interview, applicants were rated on a four-point scale (Wagner, 1949). In Yonge's (1956) investigation of the validity associated with a patterned interview, a three-point rating scale was used. On each of these counts, from the consistency of structured interview guides and their job-orientedness through to the systematic techniques used in their evaluation, early structured employment interviews differed fundamentally from their unstructured counterparts.

In the light of the drastic changes in employment interviewing introduced by structured employment interviews, it is not surprising that interviewers required training in how to conduct and evaluate them effectively and proficiently. Thus, although interviewer training does not constitute a feature of structured employment interviews per se, it is a valuable and necessary accompaniment to the implementation of these interviews in the selection arena.

Although it is not the main focus of this discussion to provide a comprehensive insight into the methodological flaws that have plagued early research on structured employment interviews, the author believes that it is important to note that there is evidence of small sample sizes and restriction of range in these studies (Wagner, 1949). Despite these and other criticisms that may be levelled against early research on the structured interview during the first half of the 20<sup>th</sup> century, the insights yielded by these research efforts were, indeed, invaluable in the realm of employment interviewing. Subsequent research on structured interviews during the 1960s and 1970s yielded further critical insights.

The early structured employment interviews that emerged during the first half of the 20<sup>th</sup> century, demonstrated a definite orientation towards job relevance. However, on the basis of some of the commentaries and reports that were written about them, there does not seem to have been any evidence of a tendency for them to capture or be evaluated in terms of actual job-related behaviours (Wagner, 1949; Yonge, 1956). However, Dudycha (cited in Janz et al., 1986) claimed to have successfully assessed the trait of punctuality in the employment interview. Dudycha accomplished this feat by asking applicants what they would do in situations that required punctual behaviour to be demonstrated (Janz et al., 1986). On the basis of this evidence, Janz et al. (1986) asserted that this study constituted the first behaviour-based approach to formulating interview questions. In a study conducted by Maas (1965) the idea of a behavioural orientation in structured employment interviews was investigated. This study and its findings appear to have made significant in-roads in employment interview research (Janz et al., 1986). Maas (1965) essentially proposed a patterned scaled expectation interview procedure, which was grounded in and evaluated with a rating mechanism that was based on scaled examples of actual on-the-job behaviour. The idea of evaluating applicants on the basis of actual job

behaviour, seemed to signal a new direction in research on structured employment interviews. However, not only did Maas' (1965) study emphasise the importance of relating interviewees' responses to on-the-job behaviours, it also illustrated the value of assessing interviews with structured and more differentiated rating mechanisms. The significantly higher and superior inter-rater reliability results obtained for interviews evaluated in terms of scaled expectation rating techniques, when compared with the reliability results for interviews evaluated with traditional adjectival scales, bears testament to this latter contribution (Maas, 1965).

The promise of structured employment interviews was recognised very early in the 20<sup>th</sup> century. Early commentators proposed that improved reliability and validity results could be attained by using structured interviews for the purpose of selection (Berman, 1997; Mayfield, 1964; Moffatt, 1969; Ulrich and Trumbo, 1965; Wagner, 1949). However, by 1964, no systematic attempts had been made to assess the impact of varying levels of structure in employment interviews (Mayfield, 1964). Schwab and Heneman (1969) attempted to rectify this situation by investigating the effects of three degrees of interview structure on inter-interviewer reliability. These researchers found that the amount of agreement among interviewers and the degree of interview structure were positively related. This implied that as the latter increased, so did the former. The researchers deemed this positive relationship to be significant (Schwab and Heneman, 1969). Despite these promising results, Schwab and Heneman (1969) did caution readers to some of the inherent flaws in their research including the effect of small sample sizes and the impact of the degree of homogeneity among applicants (which was manipulated in this study) on the results that they obtained.

In a few subsequent research endeavours, the impact of the degree of structure in the interview on validity and reliability was investigated in conjunction with other factors such as:

- The format of the rating scales and the similarity of interviewers' ratings (Hakel, 1971); and
- The presence or absence of biographical information and interviewee order (Heneman, Schwab, Huet and Ford, 1975).

Although these initiatives did not yield positive reliability and validity results, the importance of investigating the impact of varying levels of interview structure was established before the onslaught of the third trend in employment interview research.



### 2.2.3.3 In Search of a Definition for Structured Interviews

The author was unable to locate a general and widely accepted definition of the structured employment interview. During the third trend in employment interview research several variants of structured interviews emerged. Using some of the most well known of these variants (viz. behaviour description interviews, situational interviews and highly structured or comprehensive interviews) as a fundamental point of departure, a general definition of structured employment interviews will be attempted by capturing the essence of the qualities that these interviews share.

In the light of the orientation of the present study to the examination of research on situational and behaviour description interviews, these specific variants will be explored extensively in this discussion. While this in-depth exploration will not detract from the discussion of other variants of structured employment interviews, the author will not explore these in as much detail.

#### (i) The Qualities of Structured Interviews

In this discussion the main characteristics of three types of structured employment interviews (viz. situational interviews, behaviour description interviews and highly structured interviews) will be explored.

#### (a) Situational Interviews

The situational interview concept was developed and operationalised by Latham, Saari, Pursell and Campion (1980). Situational interviews have essentially been premised on goal-setting theory (Latham, 1989). According to this theoretical orientation, the future behaviour of individuals may be predicted on the basis of their future intentions and goals (Locke and Latham, 1990). This premise is grounded in empirical research on the relationship between intentions and performance. Locke (1966) concluded, on the basis of three experimental investigations, that the higher the level of intention, the higher the level of performance. In the light of this theoretical foundation, situational interviews are oriented towards eliciting information pertaining to how applicants intend to behave in specific hypothetical job-related situations (Whetzel and McDaniel, 1997).

A necessary preliminary requirement for the formulation of situational interviews is to conduct a job analysis (Gatewood and Field, 1994). Typically, situational interviews are based on the information yielded by the critical incident job analysis technique (Dipboye, 1992). The critical incidents generally depict examples of exceptionally good and poor work behaviour that have been observed by

employees and their managers (Gatewood and Field, 1994). These critical incidents are assessed in terms of their relevance to the job in question and only those that are deemed most representative of the critical dimensions of job performance, by the subject matter experts that participate in the job analysis and interview formulation process, are utilised as the basis for constructing situational questions (Dipboye, 1992).

In order to elicit information pertaining to the intentions of individuals in specific circumstances, situational interview questions assume a unique format. The relevant critical incidents yielded by the job analysis process are transformed into situational questions, which provide brief descriptions of important work-related scenarios. This process is followed by asking applicants what they would do if they were confronted with similar sets of circumstances in their future work contexts (Gatewood and Field, 1994).

In conducting situational interviews, interviewers are required to adhere rigidly to the situational questions included in the interview guide. They are not allowed to probe applicants for further information nor are they allowed to deviate from the questions in the guide (Dipboye, 1997).

The uniqueness of situational interviews extends beyond the format of their questions and the manner in which they are conducted, to the way in which the information that they yield is evaluated. In situational interviews, applicants' responses are evaluated using highly structured rating guides (Harris, 1989). These are essentially comprised of behaviourally anchored five-point rating scales for each situational question, which include good, average and poor examples of answers (Dipboye, 1992; Gatewood and Field, 1994). The examples are based on the judgements of the supervisors for the jobs in question, and constitute either the actual behaviours they observed or the responses they encountered during situational interviews (Gatewood and Field, 1994). The benchmarked examples of good, average and poor responses are not disclosed to applicants but are used by interviewers in the scoring process (Gatewood and Field, 1994). An example of a situational interview question together with a scoring scale was provided by Latham and Saari (1984, p. 571):

“For the past week you have been consistently getting the jobs that are the most time consuming (e.g. poor handwriting, complex statistical work). You know it's nobody's fault because you have been taking the jobs in priority order. You have just picked your fourth job of the day and it's another “loser”. What would you do?

- |             |  |
|-------------|--|
| 1 (low)     | Thumb through the pile and take another job  |
| 2 (average) | Complain to the co-ordinator, but do the job |
| 3 (high)    | Take the job without complaining and do it.” |

The manner in which situational interviews are scored is determined by a range of factors, which include the needs of the selectors and the purpose of the interview in the selection process. If a total interview score is required then the ratings assigned to applicants for each question are merely tallied and averaged (Gatewood and Field, 1994). If, however, selectors require an insight into the performance of applicants on each of the dimensions assessed during the interview, then the scores for all the questions pertaining to these dimensions are cumulated to yield composite scores per dimension (Gatewood and Field, 1994).

#### (b) Behaviour Description Interviews

Behaviour description interviews are premised on the principle of behaviour consistency, which advocates that the best predictor of future behaviour is past behaviour (Harris, 1989; Janz et al., 1986). Janz et al. (1986) qualified this statement by suggesting that past behaviour is only truly predictive of future behaviour in the context of a similar set of circumstances. Janz (1989) further, proposed two conditions that enhanced the predictive power of past behaviour. Firstly, the more recent the behaviour the greater its predictive potential and secondly, the more long-standing the behaviour, the greater its predictive power (Janz, 1989; Janz et al., 1986). On the basis of the fundamental principle upon which behaviour description interviews are grounded, it is evident that they are designed to elicit information about how applicants have behaved in specific work situations in the past in order to render predictions and inferences about how they are likely to behave in similar situations in the future (London, 1995).

There are essentially two types of behaviour description interviews that have emerged from the literature and research on structured employment interviews. These are patterned behaviour description interviews, which were developed and researched by Janz (1982) and structured behavioural interviews, which were developed and researched by Motowidlo, Carter, Dunnette, Tippins, Werner, Burnett and Vaughan (1992). The development of both of these variants of behaviour description interviews is premised on the critical incident job analysis technique (Janz et al., 1986; Motowidlo et al., 1992). Pertinent critical incidents that are identified during the job analysis process are transformed into behaviour description questions. These questions elicit information about how applicants actually behaved in specific job-related situations in order to render insights into the typical behaviour (as opposed to maximal or optimal behaviour) that they are likely to engage in in similar future circumstances (Gatewood and Field, 1994; Harris, 1989).

In both patterned behaviour description interviews and structured behavioural interviews, behaviour description questions essentially assume the same format. These questions provide applicants with general job-related scenarios and ask them to elaborate on a time that they encountered a similar situation. With the appropriate use of probes, interviewers elicit specific and relevant details of the situation including how the applicant behaved and the implications or consequences of their actions or behaviours (Gatewood and Field, 1994). Janz et al. (1986) offered a range of patterned interviews for various positions. For the purpose of this discussion, an example of a behaviour description question that was offered for the position of a personnel officer will be cited (Janz et al., 1986, p. 185):

“When a group of people work closely together it is inevitable that conflict will arise. Tell me about the most serious disagreement that you have had with a co-worker.

- When did this happen?
- What led to the disagreement?
- How did you attempt to solve the problem?
- What was your co-worker’s reaction?
- How was the situation resolved?
- What is your relationship with that person today?
- How often in a period of six months did you find yourself in this type of situation?”

In both patterned behaviour description interviews and structured behavioural interviews, interviewers are encouraged to use discretionary probing questions to elicit specific information regarding work-related behaviour (Janz et al., 1986; Motowidlo et al., 1992) and to take notes during the interview (Dipboye, 1997; Motowidlo et al., 1992).

Patterned behaviour description interviews and structured behavioural interviews do differ in some respects. Motowidlo et al. (1992) suggested that the latter are more structured than the former. Their assertion is based on the design of structured behavioural interviews, which necessitates that interviewers pose the same questions to all their applicants. In patterned behaviour description interviews, however, one gets the impression that interviewers are required to cover specific information pertaining to the critical behavioural dimensions identified in the job analysis and are presented with a pattern from which they can work in order to do so (Janz et al., 1986). However, they are not required to ask the same questions of applicants. Using the pattern and their knowledge of the job in question as their guides, interviewers are allowed to exercise discretion in the manner in which they conduct patterned behaviour description interviews.

These two variants of behaviour description interviews also appear to differ in the manner in which they are scored. Janz et al. (1986) proposed a procedure for evaluating applicants in the context of patterned behaviour description interviews. They recommended that the responses for each behavioural dimension should be examined and that applicants should be placed into one of five categories on the basis of their combined input throughout the interview for each behavioural dimension. If applicants received a ranking of one on the rating scale, they were regarded as falling into the bottom 20% of all applicants. If they received a ranking of three, they were perceived as falling into the middle 20% of all applicants and if they received a ranking of five, they were regarded as existing in the top 20% of all applicants. Rankings of two and four represented the 20% brackets that existed between the first and third ranks and the third and fifth ranks respectively. On the basis of this information, it appears that applicants were assigned rankings on the basis of how they compared with each other. Structured behavioural interviews, on the other hand, are evaluated in a manner that is comparable to the evaluation of situational interviews (Motowidlo et al., 1992). In structured behavioural interviews, interviewees are rated on behaviourally anchored rating scales (Motowidlo et al., 1992). This suggests that in these interviews, applicants are evaluated in relation to concrete behavioural criteria instead of in relation to each other, as appears to be the case with patterned behaviour description interviews.

In the realm of scoring and evaluation a further difference may be discerned between patterned behaviour description interviews and structured behavioural interviews. In the former, the dimensions being assessed in the interview may be differentially weighted in the process of obtaining a total interview score (Gatewood and Field, 1994; Janz et al., 1986) even though it is advised that weightings not be assigned unless some dimensions are at least two to three times more important than other dimensions (Gatewood and Field, 1994). In the structured behavioural interviews developed by Motowidlo et al. (1992) total interview scores were attained by mechanically combining ratings as unweighted sums.

### (c) Highly Structured Interviews

The highly structured interview was developed by Campion, Pursell and Brown (1988). This variant of structured employment interview was deemed to be one of the most structured interviewing formats (Dipboye, 1997). It was grounded in the principles of fairness, non-discrimination and a firm belief in the value of conforming to legal guidelines and requirements in the selection arena. Highly structured interviews were premised on the information yielded by job analysis tools (Campion et al., 1988). All interviewers asked applicants exactly the same questions and no probing or deviations from the questions in the interview guide were allowed (Dipboye, 1997). The highly structured interview

essentially encompassed four different types of questions: job knowledge questions, situational questions, job simulation questions and worker requirements questions (Campion et al., 1998). These interviews were designed to be evaluated in accordance with anchored rating scales which offered examples of good, average and poor responses in order to ensure consistency across interviews (Campion et al., 1988; Harris, 1989; Dipboye, 1997). The developers of highly structured interviews envisaged that they be conducted and evaluated by a panel of interviewers (Campion et al., 1988). In an effort to facilitate the evaluation process, it was further proposed that panel interviewers take extensive notes during the interview to prevent memory decay and that they refrain from previewing the applications of interviewees before interviews to reduce the possibility of bias in their evaluations (Dipboye, 1997). The final interview scores for highly structured interviews were obtained by calculating averages in the absence of weightings across the interviewers in the panel (Campion et al., 1988; Dipboye, 1997). The developers of highly structured employment interviews intended for them to be administered consistently to all applicants with no variations or deviations (Dipboye, 1997). All the crucial qualities of highly structured employment interviews assisted in ensuring that they conformed to the fundamental principles of fairness and non-discrimination upon which they were based. The final distinguishing feature presented in this discussion, which pertained to consistency and standardisation, in addition to an emphasis on the documentation of relevant job analysis information, interview development procedures, candidate responses and scores, evidence for content and criterion-related validity and adverse impact analyses, were instrumental in ensuring the adherence of highly structured interviews to these essential principles (Campion et al., 1988).

#### (ii) The Common Elements in Structured Interviews

On the basis of a comparative analysis of the structured interview variants that were explored in the previous section, the author identified the following general characteristics of structured employment interviews:

- Structured interviews tend to be job-related because they are premised on job analysis procedures and consequently, possess some degree of content validity (Gatewood and Field, 1994);
- They tend to be oriented towards eliciting specific information regarding the behaviour of applicants in work-related contexts in addition to other job-related information;
- Structured employment interviews appear to be conducted in a consistent manner even though the extent of this consistency may differ from one variant to another. Interviewers, in

situational and highly structured interviews, are required to ask applicants exactly the same questions without probing or deviating from the structured interview guides. In structured behavioural interviews, they are required to ask all applicants the same questions but are allowed to probe for further information. In patterned behaviour description interviews, however, interviewers are merely required to cover the same topics and follow the same pattern with all their applicants which implies that the questions each applicant is asked may differ and probes are allowed and encouraged; and

- These interviews tend to be evaluated with some degree of consistency although the extent of this consistency may differ from one variant to another. Situational interviews, highly structured interviews and structured behavioural interviews are scored using behaviourally anchored rating scales while patterned behaviour description interviews are evaluated by comparing candidates with one another.

The author’s exploration of structure in situational, behaviour description and highly structured interviews yielded four common characteristics of structured employment interviews. Campion, Palmer and Campion (1997) identified fifteen contributors to structure in employment interviews. They divided these components into aspects that impacted on the content of the interview or the nature of the information that was elicited and those that influenced the evaluation process or the manner in which the information that was elicited was assessed (Campion et al., 1997). An insight into these fifteen components of structure is presented in Table 2.1. This table was derived from a table that these writers presented in their research report.

Table 2.1 The Contributors to Structure in Employment Interviews

CONTENT	EVALUATION
1. Base questions on a job analysis	8. Rate each answer or use multiple scales
2. Ask exact same questions of each candidate	9. Use detailed anchored rating scales
3. Limit prompting, follow-up questioning and elaboration on questions	10. Take detailed notes
4. Use better types of questions	11. Use multiple interviewers
5. Use longer interview or larger number of questions	12. Use same interviewer(s) across all candidates
6. Control ancillary information	13. Do not discuss candidates or answers between interviews
7. Do not allow questions from candidate until after interview	14. Provide extensive interviewing training
	15. Use statistical rather than clinical prediction

(Campion, M.A., Palmer, D.K., and Campion, J.E. (1997). A Review of Structure in the Selection Interview. *Personnel Psychology*, 50, p. 657).

In the author's opinion, the distinction between the components of structure that affect the content of the interview and those that impact on the manner in which interviews are conducted, may be generalised to the four features of structure that were identified earlier in this discussion. Job relevance, the nature of the information elicited from applicants and the consistent manner in which interviews are conducted may be classified as factors that impact on the content of these interviews, while the consistency with which they are evaluated may be deemed to influence the manner in which they are assessed or judged.

#### 2.2.3.4 Major Contributions of Research in the Third Trend

The nature of the research on structured employment interviews essentially followed the same path as research on unstructured interviews with an emphasis on the notions of validity and reliability. This body of research consisted of studies that were oriented towards ascertaining the reliabilities and validities associated with specific variants of structured interviews. These investigations were geared towards identifying the factors that impacted on validity and reliability and meta-analyses that cumulated the findings across a range of studies in order to render general conclusions about them.

One of the most significant contributions that this body of research has made to employment interview research as a whole is renewed faith in the interview as a selection tool. The common call for enhancing the level of structure in employment interviews was echoed by several commentators throughout the 20<sup>th</sup> century (Mayfield, 1964; Moffatt, 1969; Schmitt, 1976; Ulrich and Trumbo, 1965; Wagner, 1949; Wright, 1969). This call was systematically operationalised in research on structured employment interviews, which predominated from the early 1980s to the end of the 20<sup>th</sup> century. The results that these empirical investigations yielded were, indeed, imbued with promise and potential for the use of structured employment interviews in the selection arena.

The research on structured employment interviews yielded several different variants of employment interviews. Three of these (viz. situational interviews, behaviour description interviews and highly structured interviews) were elaborated on earlier in this discussion. In the author's opinion, the validity and reliability results associated with these three structured interview variants lend considerable credibility to structured employment interviews in general. The results yielded by these investigations encompassed criterion-related (predictive and concurrent) validity estimates, inter-rater reliability estimates and internal consistency data. Construct validity findings were reported in the studies by Motowidlo et al. (1992) on the structured behavioural interview and test-retest reliability estimates were noted in the investigations on patterned behaviour description interviews (Janz, 1982; Orpen, 1985).



The predictive validity coefficients noted by Campion et al. (1988) for the highly structured interview were 0.34 (uncorrected) and 0.56 (corrected for unreliability in the criterion and restriction of range). The reliability evidence yielded for this structured interview variant was positive with estimates of 0.72 for internal consistency and 0.88 for inter-rater agreement (Campion et al., 1988). In general, the validity and reliability results obtained in relation to behaviour description and situational interviews were also positive. However, the author will not offer a detailed discussion of these findings in this section. Due to the focus of the present investigation on an evaluation of the validation research on situational and behaviour description interviews, the validity findings that were yielded for these two structured interview variants will be presented in detail in the Results chapter of this study.

The positive validity and reliability findings associated with structured employment interviews may have served to convince selectors of their utility and effectiveness as selection tools. However, this knowledge appears to have done little to appease employment interview researchers who were concerned with identifying the factors that affected validity and reliability results. In this section, the author will comment on some of the research efforts that were oriented towards identifying these factors in the context of employment interviews.

#### (i) Factors Affecting Validity

Huffcutt and Arthur (1994) concluded that structure was a major moderator of validity in employment interviews. In their analysis, structure was defined in terms of the extent to which the procedures followed by interviewers were allowed to be varied. Structure, therefore, was dependent on the degree of discretion with which interviewers were endowed. The two dimensions of structure that related directly to the degree of interviewer discretion were the standardisation of interview questions and the standardisation of response scoring. The results of their meta-analytic investigation revealed that mean validity estimates tended to increase with increasing levels of structure. Moreover, the increases in interview structure did not yield improvements in validity beyond a point, thereby suggesting that structure has a ceiling effect on validity (Huffcutt and Arthur, 1994).

In a review of structure in selection interviews, Campion et al. (1997) assessed the impact of three types of validity information on the fifteen components of structure they identified. The two components of structure that Huffcutt and Arthur (1994) identified may, in the author's opinion, be mapped onto the two main components of structure (*viz.* factors that influence the content of interviews and factors that impact on the evaluation of interviews) proposed by Campion et al. (1997). Campion et al. (1997) defined validity in terms of job-relatedness (the extent to which the interview is related to the content of the job); reduced deficiency (the degree to which measurement deficiency is

reduced and a large amount of useful information is elicited from the interview) and reduced contamination (the degree to which the interview is designed to prevent faking and irrelevant information from contaminating the interview process). A summary of their findings is presented in Table 2.2. This table constitutes an excerpt of a table that these authors presented in their article (Campion et al., 1997, p. 657).

Table 2.2 The Impact of the Fifteen Components of Structure on Validity

STRUCTURE	VALIDITY		
	Job-relatedness	Reduced deficiency	Reduced contamination
<b>CONTENT</b>			
1. Base questions on a job analysis	+	+	+
2. Ask exact same questions of each candidate		+	+
3. Limit prompting, follow-up questioning and elaboration on questions		-	+
4. Use better types of questions	+		+
5. Use longer interview or larger number of questions		+	
6. Control ancillary information		-	+
7. Do not allow questions from candidate until after interview		-	+
<b>EVALUATION</b>			
8. Rate each answer or use multiple scales		+	+
9. Use detailed anchored rating scales	+	+	+
10. Take detailed notes	+	+	+
11. Use multiple interviewers		+	+
12. Use same interviewer(s) across all candidates			+
13. Do not discuss candidates or answers between interviews			+
14. Provide extensive interviewing training	+	+	+
15. Use statistical rather than clinical prediction		+	+

+ denotes a positive impact and - denotes a negative impact (Campion, M.A., Palmer, D.K., and Campion, J.E. (1997). A Review of Structure in the Selection Interview. *Personnel Psychology*, 50, p. 657).

Although they did not conceive of note-taking as constituting a component of structure in the way that Campion et al. (1997) did, Burnett, Fan, Motowidlo and Degroot (1998) concluded that the effects of note-taking behaviour and the content of interview notes impacted on the validity of employment interviews. In particular, they found that voluntary note-taking resulted in note-takers rendering more valid ratings than their non-note-taking counterparts and that when the notes were behavioural in nature, greater validity was noted for interviews (Burnett et al., 1998).

Motowidlo and Burnett (1995) conducted an intriguing study on the impact of aural and visual sources of information on the validity associated with structured employment interviews. They argued that in empirical investigations oriented towards ascertaining the validity estimates associated with situational and behaviour description interviews, interviewers were required to base their judgements solely on the

content of interviewees' responses to questions pertaining either to their actual past behaviour or their intended future behaviour (i.e. aural sources of information). In these interviews, interviewers were strictly required to ignore all other sources of information such as the physical characteristics of interviewees, their appearance, mannerisms, dress, grooming and all other non-verbal characteristics (Motowidlo and Burnett, 1995). However, Motowidlo and Burnett (1995) argued that these visual sources of information could also exert an influence on interviewers' judgements in the interviews that advocated a reliance on aural sources of information. Using interviews that assumed a behaviour description interview format, these researchers concluded that interviewer judgements, that pertained to the potential of interviewees to be effective within a management position, based solely on aural cues correlated substantially with the same judgements when they were based solely on visual cues (Motowidlo and Burnett, 1995). These researchers further suggested that interview judgements based solely on visual cues could effectively predict supervisors' ratings of job performance. On the basis of their findings, Motowidlo and Burnett (1995) concluded that visual sources of information played a vital role in the estimation of interview validity when the performance criterion was assessed through supervisors' ratings.

On the basis of this discussion, it is clear that different facets of structure including note-taking content and behaviour and visual sources of information impacted on the validity of structured employment interviews.

#### (ii) Factors Affecting Reliability

Conway, Jako and Goodman (1995) conducted a meta-analysis of the inter-rater and internal consistency reliabilities associated with selection interviews. They proposed that the following factors impacted on the inter-rater reliability estimates obtained in employment interviews:

- Study design (panel interviews versus individual interviews);
- Interview structure (which implied standardisation in three areas: (a) interview questions, (b) the manner in which responses were evaluated and (c) the method that was used for combining ratings);
- Job analysis; and
- Interviewer training.

Their analysis revealed that study design, interview structure and interviewer training did, in fact, moderate interview reliability. Further, they found that the standardisation of questions was more strongly related to inter-rater reliability when coefficients were based on individual as opposed to panel interviews and that multiple ratings (which represented one of the ways in which responses were evaluated) were useful when combined mechanically or actuarially as opposed to subjectively (Conway et al., 1995). These findings were corroborated by Campion et al. (1997) in their review of structure on employment interviews in which they commented on the impact of the fifteen components of structure that they identified on different aspects of reliability.

Thus, research efforts that were aimed at identifying the factors that influence the validity and reliability estimates associated with structured employment interviews have yielded useful insights. However, in the author's opinion, more systematic research undertakings are required in this area in order to obtain clarity on the factors that impede and contribute to the validity and reliability of structured employment interviews.

During the course of the third trend in employment interview research, which spanned roughly from the early 1980s to the end of the 20<sup>th</sup> century, several meta-analytic studies emerged. These investigations cumulated the validity results obtained for employment interviews from several independent studies using statistical techniques in order to yield an overall perspective on the validity estimates associated with employment interviews, in general, and structured interviews, in particular. The meta-analytic studies that emerged during this period typically offered positive validity results for structured employment interviews.

Wiesner and Cronshaw (1988) proposed that structured interviews (0.62) yielded substantially and significantly higher predictive validity coefficients than their unstructured (0.31) counterparts. Although they predicted that board or panel interviews would yield higher validity estimates than individual interviews, their findings did not support this hypothesis. This led to the conclusion that interview structure moderated validity while interview format did not appear to influence this psychometric characteristic (Wiesner and Cronshaw, 1988). In a meta-analysis of structured employment interviews by Wright, Lichtenfels and Pursell (1989) an estimated validity of 0.39 was obtained. In their meta-analysis of the validity of employment interviews, McDaniel, Whetzel, Schmidt and Maurer (1994) concluded that:

- Interviews with situational content (0.50) yielded a higher validity estimate than job-related interviews (0.39) and psychological interviews (0.29);
- Structured interviews (0.44) were more valid than their unstructured (0.33) counterparts;
- Structured individual interviews (0.46) were more valid than structured board interviews (0.38);
- Unstructured individual and board interviews were associated with almost equivalent mean validity coefficients of 0.34 and 0.33 respectively; and
- The nature of the criterion would impact on validity. Their analysis revealed that similar validity results were yielded for interviews in which job performance (0.37) and training performance (0.36) were used as the criterion measures while the validity that was obtained by using tenure (0.20) as the criterion was lower (Whetzel and McDaniel, 1997).

#### 2.2.3.5 Major Criticisms of Research on Structured Interviews

The third trend in employment interview research was characterised by investigations of structured employment interviews. This body of research does not seem to have proceeded from a unanimous definition of the term “structure”. However, the reference to structured interviews does generally encompass the common features of consistency and standardisation with regard to the information that is elicited and the manner in which this information is evaluated. Notwithstanding, these key similarities, the details of exactly what the term “structure” encompasses may differ from one investigation of so-called structured interviews to another. Thus, comparisons of the findings yielded by these studies may be difficult, if not impossible, to undertake. In the author’s opinion, the formulation of a unanimous definition of structure would benefit future research on structured employment interviews insofar as it will render them directly comparable.

In the individual studies of different structured interview variants, small sample sizes were noted. This criticism was also levelled against research on unstructured employment interviews. The implications of this methodological flaw are similar to those outlined in the section on the major criticisms of unstructured employment interviews which was included in the discussion of the first trend in employment interview research.

The introduction of meta-analytic investigations in the research on structured employment interviews constituted an attempt to overcome the problem of small sample sizes and to generally improve research efforts on these interviews by cumulating the findings of individual studies. However, the

investigations that were engaged in on the basis of this technique, were characterised by difficulties of their own. One of the most prevalent concerns in the context of meta-analyses is the potential for sampling bias (Rosenthal, 1991) or availability bias (Hunter and Schmidt, 1990). According to this argument the studies that are available for analysis in meta-analytic investigations typically constitute a biased sample of all available studies (Hunter and Schmidt, 1990). Generally, studies have a greater probability of being published if they yield statistically significant results (Rosenthal, 1991). Published research is usually most accessible to researchers. The range of published studies that are available to meta-analytic researchers may, therefore, not be representative of the entire population of studies that have been conducted on a specific topic due to the tendency of research publishers to exclude investigations that do not yield statistically significant findings (Hunter and Schmidt, 1990; Rosenthal, 1991).

Rosenthal (1991) argued that a second criticism, which may be levelled against meta-analytic research is the tendency for information to be lost during the course of analysis. One way in which information may be lost is through the use of a single value to summarise an entire research domain (Rosenthal, 1991). The primary purpose of meta-analytic research is to summarise independent studies that pertain to a research topic in a quantitative manner. However, Rosenthal (1991) stated that one of the consequences of summarising data is to gloss over details. This constitutes the second way in which information tends to be lost in meta-analyses. This criticism may also be levelled against narrative reviews that aim to summarise the findings of a range of individual investigations (Rosenthal, 1991).

A third criticism of meta-analytic research is that it tends to cumulate studies that have been conducted according to vastly different methodologies. Some commentators proposed that such studies may be perceived as not comparable (Hunter and Schmidt, 1990). In effect, therefore, the cumulation of such studies in the context of meta-analytic investigations seems to represent a futile comparison of apples and oranges which would not yield any useful information (Hunter and Schmidt, 1990; Rosenthal, 1991). In response to this criticism, however, Glass (as cited in Rosenthal, 1991) noted that the comparison of apples and oranges are indeed a useful point of departure when one's aim is to generalise to fruit.

There are many more criticisms that may be levelled against meta-analytic research. However, it is not within the scope of this study to highlight all of them. In this discussion the most common criticisms that have been levelled against meta-analyses have been highlighted in order to offer an insight into some of the problems that may be inherent in the meta-analytic research on structured employment interviews. The aim of this exploration was to emphasise that while meta-analytic research appears to have contributed positively to the entire body of research on structured employment interviews, they

may have also introduced some fundamental problems that could have a significant impact on the outcomes of this research.

### 2.3 Conclusion

On the basis of this discussion, some general criticisms may be levelled against the entire body of research on selection interviews, which emerged during the course of the 20<sup>th</sup> century. Researchers and reviewers of research tended to dichotomise interviews in terms of the level of structure they possessed. In other words, the interviews investigated were either structured or unstructured. The author agrees with Huffcutt and Arthur (1994) who proposed that structure is a complex phenomenon that cannot merely be reduced to the extremities on a bipolar continuum. Instead, they proposed the existence of varying degrees of structure in employment interviews (Huffcutt and Arthur, 1994). If researchers proceed with investigations on employment interviews in the 21<sup>st</sup> century, they should rectify the simplification of the notion of structure that has inhered in research on employment interviews for approximately 100 years.

The unique dynamics of the selection process seems to contribute to the pervasiveness of two methodological constraints in research on employment interviews and other selection tools. These are small sample sizes and restriction of range. Due to their resource intensity, employment interviews are typically situated late in the selection process. Selectors rely on other less expensive, time-consuming and human resource-intensive selection devices to evaluate large batches of applicants in order to sift out those that are likely to be most suitable in terms of the job requirements. This may be accomplished by a series of successive hurdles. Once the applicant pool has been sufficiently narrowed down, selectors expose the remaining applicants (i.e. those individuals that have been identified by a range of other devices as most promising for the position in question) to more resource-intensive procedures such as employment interviews and assessment batteries. In order to accommodate the capitalistic ideal of profit-making and optimisation, therefore, employment interviews are only conducted on small groups of individuals towards the latter part of the selection process. Hence, the prevalence of small sample sizes in the bulk of the research in this arena.

The problem of restriction of range is likely to be exacerbated by the use of a multiple-hurdle approach to selection. The progressive downscaling of applicant groups on the basis of their suitability to the requirements of the job implies that selectors are ultimately left with a group of individuals that demonstrate promise. It is likely that the range of scores that these applicants will achieve on subsequent selection devices, to which they are exposed in the selection process, will not possess a

great deal of variability. Hence, the range of these scores may be restricted. Due to their intimate relationship with the unique dynamics of the selection process, the methodological flaws of small sample sizes and restriction of range may be difficult to overcome practically in the research on employment interviews. However, developments in the realm of statistics have offered mathematical controls for these problems.

As a function of the context to which the use of employment interviews is confined in organisations, it is often difficult to obtain access to actual applicants and interviews in actual employment contexts for research purposes. In order to overcome the first difficulty, some researchers have constructed paper-people to serve as candidates (Arvey and Campion, 1982; Harris, 1989). In these investigations no actual interviews occurred but the information provided to “interviewers” about these imaginary candidates was evaluated and selection decisions were made. The findings obtained in investigations of this nature may not be generalisable to the decision-making process in actual interview contexts (Arvey and Campion, 1982). In response to the lack of actual interview settings in which to conduct and evaluate employment interviews, some researchers simulated interview situations using students as interviewers and willing participants (not necessarily individuals seeking employment) as interviewees. Arvey and Campion (1982) noted that this practice could also impede the generalisability of the results that were obtained to actual interview situations. However, the research that these writers identified in relation to this issue suggested minimal differences in the evaluations rendered by students and those rendered by experienced interviewers (Arvey and Campion, 1982). Notwithstanding this evidence, the author remains sceptical of the use of paper-people and students as interviewers in the process of investigating employment interviews.

In this chapter, the three dominant trends that emerged in the context of research on employment interviews during the 20<sup>th</sup> century were explored. The third trend, which focused on the validity and reliability associated with structured employment interviews, constitutes the basis for the present study. In this investigation, the author intends to critically evaluate the impact of the methodological or research design features of a group of studies on two specific variants of structured employment interviews (viz. behaviour description interviews and situational interviews) on their internal and external validity. The main purpose of this exploration will be to alert structured employment interview researchers to the threats they are likely to encounter in their research initiatives and to recommend how these threats may be managed effectively. These initial generic recommendations will ultimately be tailored to suit the unique needs and demands of the South African context.

General and methodological criticisms have been levelled against research on employment interviews throughout the 20<sup>th</sup> century. However, the author has not identified any commentary that provided an



in-depth evaluation of the methodological or research design features of studies on employment interviews in order to make recommendations regarding how these studies could be improved upon. It is possible that such efforts were conducted as preliminary analyses by the employment interview researchers who have contributed to the large base of empirical evidence on this crucial selection tool. However, the details of these analyses, if any were, in fact, engaged in, do not appear to have been documented in the literature on employment interviews. In the present study, a critical evaluation of a group of studies on behaviour description and situational interviews will be engaged in with a view to highlighting the methodological factors that hindered and contributed to the internal and external validity of these studies. In so doing, a perceived gap in the literature on employment interviews, in general, and on structured employment interviews, in particular, will be bridged.

## **Chapter 3**

### **Methodology**

#### 3.0 Introduction

In this chapter, the procedures that were employed in the execution of the investigation will be highlighted. This discussion will commence with a statement of the primary and secondary aims of the study and will outline the manner in which these aims were attained.

#### 3.1 A Statement of the Research Problem

The primary aim of the study was to explore the potential threats to internal and external validity that plagued international validation research efforts on structured variants of employment interviews. In addition, there were two secondary research aims:

- To comment on how the potential threats to internal validity could have been effectively managed; and
- To offer an insight into whether or not the potential threats to external validity appear to have been actualised in international validation research efforts.

#### 3.2 Exploring the Concepts of Internal and External Validity

In this section, the concepts of internal and external validity will be discussed. The author will demonstrate how these concepts pertain to the study and will highlight the vital link between them.

##### 3.2.1 Internal Validity

The theory on research methods espoused that internal validity refers to “the degree to which the mathematical relationship we observe between subjects’ scores actually and only reflects the relationship between the variables of interest” (Heiman, 1998, p. 61). Shaughnessy and Zechmeister (1997) proposed that the internal validity of research studies may be called into question when there is evidence of confounding. Confounding occurs when the findings obtained may be attributed to

extraneous variables that interact with the independent variables to influence the outcomes of research efforts (Shaughnessy and Zechmeister, 1997).

International validation research on structured employment interviews (viz. behaviour description interviews and situational interviews) constituted the primary focus of this investigation. The main purpose of validation research on employment interviews and other selection tools is to ascertain whether or not the use of the selection tool in question, allows for the accurate prediction of job performance. In the context of validation research in the selection arena, the independent variable is typically the employment interview or the selection tool in question and the dependent variable is usually the criterion measure, which is oriented towards the assessment of job performance. When there are strong positive correlations between the scores that subjects obtain on the predictor and criterion measures then high validity results are observed.

The author explored the potential threats that could have impacted on the internal validity of the validation studies included for analysis in this investigation. This exploration was engaged in by investigating whether the validity results yielded were due to the ability of the predictor (i.e. the structured employment interview variant in question) to accurately and effectively predict the job performance of subjects on the criterion measure or whether these estimates were the result of interactions of extraneous variables with the predictors (independent variables), which served to confound the correlations obtained between the scores on the predictor and criterion measures.

### 3.2.2 External Validity

External validity refers to the extent to which the findings from individual investigations generalise to other individuals, contexts and time periods (Dooley, 1995; Heiman, 1998). Stern and Kalof (1996) proposed that the most effective way to determine whether or not studies possess external validity is to compare them with other research endeavours that are oriented towards the investigation of the same basic variables, but in different ways. They argued that if similar findings emerge from studies that are conducted in different settings, with different populations and using different observers (in the case of this investigation, interviewers and job performance evaluators), then these studies may possess external validity (Stern and Kalof, 1996).

In this study, the research on patterned behaviour description interviews, structured behavioural interviews and situational interviews were clustered into three separate categories. In the context of each category, an exploration of the threats to external validity was engaged in. In order to comment on external validity, studies that investigated the relationship between the same independent (i.e. the

specific structured interview variant in question) and dependent (i.e. job performance) variables using different research procedures and processes, in different settings and using different groups of individuals as both subjects and evaluators of interview and job performance, were compared. The type of structured employment interview was held constant in the exploration of external validity in order to ensure that the comments made were a function of differences in the research processes and procedures as opposed to differences in the type of structured employment interviews that were used.

### 3.2.3 The Link Between Internal Validity and External Validity

Stern and Kalof (1996) proposed that the concept of external validity only becomes meaningful once the internal validity of research investigations is ascertained. Therefore, it is clear that the determination of internal validity is a crucial pre-requisite for the exploration of external validity. This investigation is not oriented towards explicating the link between the internal and external validity of the international validation studies that were reviewed. Thus, the inter-relatedness of these concepts was not incorporated into this investigation and the threats to internal and external validity were, consequently, explored independently.

The main aim of the present study was to explore the threats to internal and external validity to which international validation research on structured employment interviews appear to have been exposed. The purpose of this exploration was to alert future researchers, especially those in the South African context, to the potential difficulties that they could encounter in their validation research attempts on structured employment interviews and to offer suggestions regarding how these threats could be overcome. Therefore, in the exploration of the threats to internal validity in each of the three categories of structured employment interviews, the possible threats were considered and suggestions were made on how they could have been addressed. In the exploration of threats to external validity, a qualitative exploration was engaged in and a meta-analytic technique was used to compare the findings of multiple investigations. On the basis of this latter analysis, the author attempted to comment on whether or not the potential threats to external validity did, in fact, manifest in the studies under investigation. The exploration of the threats to external validity also occurred in the context of the three separate categories of structured employment interviews that were identified in this study (viz. patterned behaviour description interviews, structured behavioural interviews and situational interviews).

### 3.3 The Sample

In this investigation the sample was comprised of published research initiatives, which were oriented towards the validation of structured variants of employment interviews. Research reports that investigated two of the most popular types of structured employment interviews (viz. behaviour description interviews and situational interviews) were selected for analysis. A total of 14 studies were located in nine research reports that appeared to focus on the validation of these structured interview variants. These studies are listed in Table 3.1.

Table 3.1 The Studies Considered for Inclusion in this Investigation

RESEARCHERS	TITLES
<b>BEHAVIOUR DESCRIPTION INTERVIEWS</b>	
<i>PATTERNED BEHAVIOUR DESCRIPTION INTERVIEWS</i>	
1. Janz (1982)	Initial Comparisons of Patterned Behaviour Description Interviews Versus Unstructured Interviews
2. Orpen (1985)	Patterned Behaviour Description Interviews Versus Unstructured Interviews: A Comparative Validity Study
<i>STRUCTURED BEHAVIOURAL INTERVIEWS</i>	
3. Motowidlo, Carter, Dunnette, Tippins, Werner, Burnett and Vaughan (1992)	Studies of the Structured Behavioural Interview
<b>SITUATIONAL INTERVIEWS</b>	
4. Latham, Saari, Pursell and Campion (1980)	The Situational Interview
5. Latham and Saari (1984)	Do People Do What They Say? Further Studies on the Situational Interview
6. Weekley and Gier (1987)	Reliability and Validity of the Situational Interview for a Sales Position
7. Robertson, Gratton and Rout (1990)	The Validity of Situational Interviews for Administrative Jobs
8. Stohr-Gilmore, Stohr-Gilmore and Kistler (1990)	Improving Selection Outcomes With the Use of Situational Interviews: Empirical Evidence from a Study of Correctional Officers for New Generation Jails
9. Gabris and Rock (1991)	Situational Interviews and Job Performance: The Results in One Public Agency

These research reports were located using computerised searches and the conventional technique of isolating relevant material in the reference lists of related literature. Once these research reports had been identified, they were evaluated in terms of the following criteria in order to decide on which studies to include for analysis in this investigation. Only those studies that conformed to all the pre-determined criteria were included for analysis. These criteria are presented in the form of questions. Those studies, for which it was possible to offer positive responses to all these questions, were ultimately analysed in this investigation.

- Is the research report oriented towards the determination of the validity associated with a structured employment interview in the behaviour description interviews or situational interviews categories?
- Does the research endeavour yield at least one quantitative validity estimate for the structured employment interview under investigation?
- Did the definitions of the specific variants of structured employment interviews conform to the general definitions of behaviour description interviews, patterned behaviour description interviews, structured behavioural interviews and situational interviews advocated in employment interview literature?
- Were the studies that comprised these research reports independent instances of primary research investigations into the validities associated with specific variants of structured employment interviews as opposed to meta-analytic investigations or other secondary instances of research?

These criteria were used for the following reasons. Firstly, in order to explore the potential threats to internal and external validity that validation research on structured employment interviews is likely to be plagued by and to comment on how these may be managed, it was imperative to select a group of validation studies on structured variants of employment interviews for the purpose of this investigation.

Secondly, it was ensured that each of the investigations considered for analysis yielded a quantitative validity estimate for a specific variant of structured employment interview. The rationale for using this criterion was to ensure that a meta-analytic technique, for comparing the findings yielded within each category of structured employment interview, could be used to comment on whether or not the potential threats to external validity appear to have influenced the generalisability of the findings. The author acknowledges that the applicability of these insights into external validity is dependent on whether or not the studies in question possess internal validity.

Thirdly, the aim in ensuring that the structured interviews which were investigated all conformed to the general theoretical definitions provided for them in employment interview literature, was to ensure that the studies included in each category were all investigating the same interview type. The independent variables investigated in the context of each category were, thus, consistent. In the light of this consistency it was possible to comment on the threats to internal and external validity that appear to have plagued the studies in each category, with the confidence that emanated from the knowledge that these commentaries pertained to the same type of structured employment interview.

Fourthly, the reason for including studies that constituted independent instances of primary research initiatives was to facilitate the use of the meta-analytic techniques proposed by Rosenthal (1991) for comparing the findings of independent studies. Rosenthal (1991) defined independent investigations as those that were conducted using different research participants. In this study, these meta-analytic techniques were used to comment on whether or not the threats to external validity appear to have been actualised in the research within each category of structured employment interview.

The author offered a positive response to all these questions for the majority of the research reports listed in Table 3.1, except for the last one by Gabris and Rock (1991). Thus, most of the studies included for analysis in this investigation did conform to the pre-defined criteria. At face value, the title of the article by Gabris and Rock (1991) seemed to imply an orientation towards the determination of the validity associated with the situational interview. However, a closer analysis revealed that this study was qualitative in nature. Consequently, it did not yield a quantitative validity estimate. Secondly, the manner in which situational interviews were defined in this investigation, did not conform to the original definition of situational interviews that was envisaged by its developers and chief proponents in the literature on employment interviews. In this investigation the researchers envisaged the use of in-basket exercises as part of the situational interview (Gabris and Rock, 1991). This practice does not conform to traditional situational interview formats. Therefore, the definition of situational interviews used by these researchers did not conform to the generally accepted definitions espoused in the literature on employment interviews. On the basis of these deviations from the specific pre-defined criteria outlined above, this report was excluded from consideration in this investigation.

The use of eight research reports and the inclusion of 13 studies in the final sample for this investigation, implies that some of the research reports in question contained more than one study. In Table 3.2 an insight into the number of studies, located within each category of structured employment interview, is provided. In so doing, the frequencies of the studies within each category of structured employment interview, are offered.

### 3.4 The Method

In this study a qualitative investigation of secondary data was engaged in. Consequently, the approach that was adopted was primarily qualitative in nature. The author undertook an exploration of the threats to the internal validity of the studies included in the sample in the context of the three categories of structured employment interviews identified in this study. This exploration was based on the guidelines outlined by Stern and Kalof (1996). Their methods for classifying studies in terms of the

types of research design strategies that they employed served as a fundamental point of departure for the analysis of the potential threats to internal validity. These classification methods are captured in a decision-tree, proposed by Stern and Kalof (1996). This decision-tree, which is presented in Figure 3.1, was used to identify the research design strategies employed in the studies analysed in this investigation.

Table 3.2 Frequency of Studies Per Structured Interview Variant

TYPE OF STRUCTURED INTERVIEW	FREQUENCY OF STUDIES	STUDIES
Patterned Behaviour Description Interviews	2	Janz (1982) Orpen (1985)
Structured Behavioural Interviews	3	Motowidlo et al. (1992) [Studies 2, 3 & 4]
Situational Interviews	8	Latham et al. (1980) [Studies 1, 2 & 3] Latham and Saari (1984) [Studies 1 & 2] Weekley and Gier (1987) Robertson et al. (1990) Stohr-Gilmore et al. (1990)
Total	13	

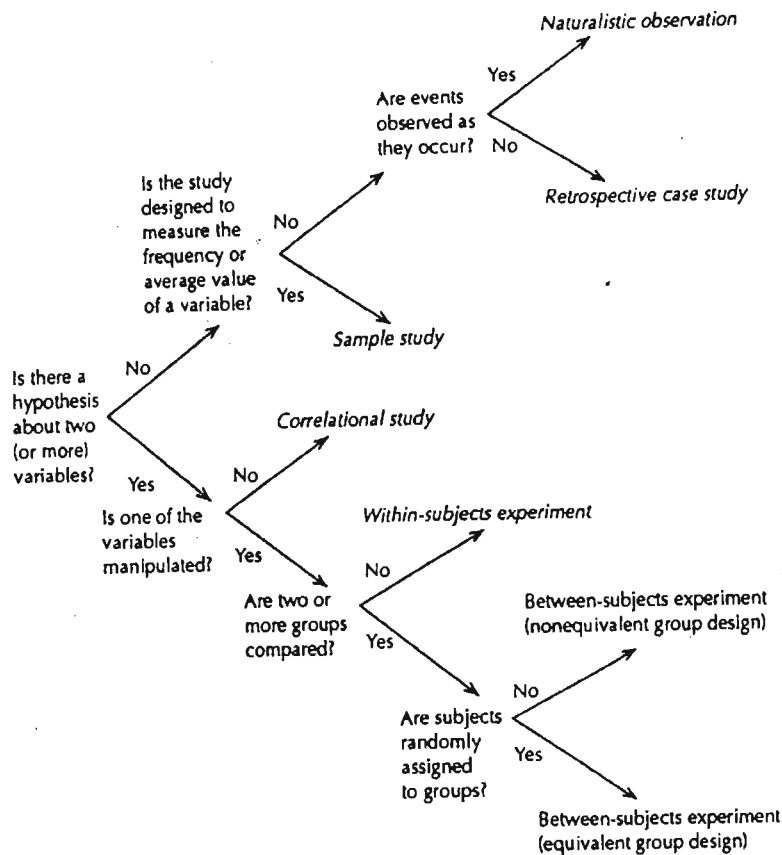
Once the studies in the sample were analysed and classified according to the research design strategies that they employed, the author used Stern and Kalof's (1996) guidelines, on the unique threats to internal validity that different research design strategies are prone to, to explore the potential threats to the internal validity of these studies. In this discussion, commentary was offered on whether or not these threats were controlled for by the researchers. In the instances where they were not controlled, suggestions, regarding how they could have been controlled, were proposed.

The notion of external validity, in relation to the research investigations in question, was explored by comparing the results obtained in the studies within each category of structured employment interview. This exploration was undertaken in two ways. Firstly, using the threats to external validity identified by Dooley (1995), as a point of departure, a qualitative exploration of the potential threats to the external validity of the studies in question, was engaged in. This exploration yielded an insight into the threats that could have influenced the external validity of the studies. In order to comment on whether or not these potential threats appear to have manifested in the studies under investigation, a second quantitative strategy was employed. This entailed the use of meta-analytic techniques for comparing the findings of independent research undertakings. However, the author is cognisant of the fact that in order to comment on whether or not the studies in question possessed external validity, an insight into their internal validity is required. In the absence of



conclusive information on the internal validity of the investigations in question, the author merely commented on whether or not the threats to external validity could have been actualised in the studies that were analysed.

Figure 3.1 A Decision-Tree for Identifying Research Design Strategies



(Stern, P.C. and Kalof, L. (1996). Evaluating Social Science Research (Second Edition). Oxford: Oxford University Press, p. 42).

Rosenthal (1991) proposed that the findings yielded by independent investigations may be compared or combined using meta-analytic procedures. In this investigation, comparisons of the findings yielded by the studies within each category of structured employment interview were engaged in to determine whether or not these studies were essentially telling the same stories (Rosenthal, 1991). On the basis of these comparisons the author believes that it is possible to comment on whether or not the comparability of the results, and consequently the studies, was affected by the potential threats to external validity that were identified in the qualitative exploration. External validity refers to the extent to which the findings of studies generalise to different contexts, different groups and different time

periods (Dooley, 1995). The use of meta-analytic procedures for comparing the findings of different research efforts, that are essentially oriented towards asking the same research questions about the same research variables, is likely to render an insight into their external validity. However, in this study, these techniques were employed to comment on whether or not the potential threats to external validity, that were identified in the qualitative analysis, were realised in the studies under investigation. This aim was achieved by commenting on whether or not the findings obtained in these studies were consistent or statistically significantly different from one another. If they were consistent, it was proposed that the threats to external validity probably did not manifest and that, provided these studies possessed internal validity, their findings could be generalised to the contexts and populations of the studies with which they were compared. If the findings were statistically significantly different, however, it was suggested that the potential threats to external validity, identified in the qualitative analysis, probably did manifest.

Rosenthal (1991) proposed that there are two major ways in which the results of research initiatives may be evaluated. Firstly, they may be evaluated in terms of their significance levels and secondly, they may be evaluated in terms of their effect sizes. In this investigation, the studies were evaluated in terms of their effect sizes. These were represented by the validity estimates yielded in the context of the research under investigation. A further consideration, when engaging in the meta-analytic procedures for comparing the effect sizes of research investigations, is the number of independent investigations being compared. Rosenthal (1991) proposed two different techniques for the comparison of studies in terms of their effect sizes depending on whether a set of two studies was being evaluated or whether a set of three or more studies was being evaluated.

In order to compare the results obtained in the context of sets of two studies, the following equation was used (Rosenthal, 1991, p. 63):

$$\frac{z_{r1} - z_{r2}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

In this equation,  $z_{r1}$  and  $z_{r2}$  refer to the Fisher's z transformation of the effect size estimates ( $r$ ) obtained from each study and  $N_1$  and  $N_2$  refer to the number of subjects included in the two studies being compared (Rosenthal, 1991).

The author mentioned earlier that, in this investigation, the effect size estimates assumed the form of the validity results that were yielded in the studies included for analysis. These estimates reflected the

extent of the association between the scores obtained on the predictor and criterion measures. Once these validity estimates had been extracted from the investigations, the author transformed them into their Fisher  $z_r$  equivalents. Rosenthal (1991) proposed two ways in which to perform this transformation. The first is by using a table of Fisher's  $z$  transformations of  $r$ . The second, is by using the following equation (Rosenthal, 1991, p. 63):

$$\frac{1}{2} \log_e [(1 + r) / (1 - r)]$$

Essentially, this equation could be substituted for the numerator in the above equation. In this study these equations were used to calculate a standardised score, which represented the comparability of the findings yielded in a set of two studies. This score is represented as  $Z$ . On the basis of these standardised scores the corresponding  $p$ -values were obtained using the table of standard normal deviates, which was located in Rosenthal and Rosnow (1984). The  $p$ -values represented the extent of the consistency or the differences between the findings being compared.

Paired comparisons of the findings of each study with every other study within each category of structured employment interview were conducted. The reason for engaging in this practice was to comment on whether or not the findings were comparable and consequently, to investigate whether or not the potential threats to external validity had been actualised. In the category of patterned behaviour description interviews, only two independent studies were located for analysis. In order to compare the findings yielded by these studies, the meta-analytic equation for comparing a set of two studies was used. However, more than two independent studies were located in the structured behavioural interviews and situational interviews categories. For these studies, the meta-analytic technique for comparing the findings in sets of two studies was used to compare the findings of each study with every other study within each category. A cumulative comparison of all the studies within these specific categories was also engaged in using the technique for comparing the findings in a set of three or more studies proposed by Rosenthal (1991).

The equation that was used to engage in this latter comparison is provided below (Rosenthal, 1991, p. 74):

$$\bar{z}_r = \frac{\sum (N_j - 3) z_{rj}}{\sum (N_j - 3)}$$

These comparisons were engaged in using the effect sizes (validity coefficients) yielded by the three or more independent investigations in the structured behavioural interviews and situational interviews categories. Rosenthal (1991) advocated that in this equation,  $z_{rj}$  is the Fisher  $z_r$  value that corresponds to any  $r$  and  $\bar{z}_r$  is the weighted mean of all the Fisher  $z_r$  values.  $N$  refers to the number of sampling units on which each  $r$  is based (Rosenthal, 1991).

Once the weighted mean of all the Fisher  $z_r$  values being compared is attained, it is necessary to calculate the  $\chi^2$  value associated with it and the associated degrees of freedom in order to comment on the statistical significance of the differences between the  $r$  values being compared. Rosenthal (1991, p. 74) proposed the use of the following equation to compute  $\chi^2$ :

$$\sum (N_j - 3)(z_{rj} - \bar{z}_r)^2 \text{ with } K - 1 \text{ df}$$

This equation yields a  $\chi^2$  value with  $K - 1$  degrees of freedom.  $K$  in this instance refers to the number of studies or  $r$  values that are being compared. On the basis of this information, it is possible to derive the corresponding  $p$ -value, which offers an insight into whether the findings are consistent or whether they differ significantly from one another.

It was noted earlier that this technique was used for comparing the results of more than three studies, in the context of structured behavioural interviews and situational interviews. The purpose of this comparison was to comment on whether the findings yielded by the studies within each of these categories were consistent or significantly different from one another and in so doing to comment on whether or not the potential threats to external validity, which were identified in the qualitative exploration, appear to have been realised. Thus, in relation to the investigations included in these categories, both the meta-analytic techniques discussed above were used to compare the findings that were yielded.

Using these meta-analytic procedures, the author commented on whether or not the findings of the studies within each category of structured employment interviews were comparable (i.e. whether or not the potential threats to the external validity of these investigations were realised). In addition, commentary was provided on whether these findings could be generalised to other studies within the same category by using the disclaimer that this generalization would only be valid if the studies in question possessed internal validity.

### 3.5 Conclusion

This chapter outlined the methodology that was envisaged for the execution of the present investigation. In the next chapter, the results that were obtained from the implementation of these methods and procedures are presented. These results allow comments on the threats to the internal and external validity of international validation research on structured employment interviews to be made. In addition, the information that is essential for understanding and contextualising these results will be highlighted. This information will include the main insights on which the qualitative explorations of the threats to internal and external validity, of the studies that were analysed within each of the three categories of structured employment interviews, were based.

## **C h a p t e r 4**

### **Results**

#### 4.0 Introduction

Internal validity refers to the extent to which the findings yielded by research endeavours are a function of the interaction of the variables under investigation with one another instead of with extraneous variables that may serve to confound the findings and introduce alternative explanations for them (Heiman, 1998; Shaughnessy and Zechmeister, 1997; Stern and Kalof, 1996). External validity refers to the generalisability of the findings from one study to other research settings, groups of people and time periods (Dooley, 1995; Shaughnessy and Zechmeister, 1997). The definition of internal validity suggests that the exploration of threats to the internal validity of research efforts necessarily requires probing at the level of individual investigations. The exploration of threats to external validity, on the other hand, necessarily involves the comparison of the findings that are yielded by a group of investigations in order to ascertain whether or not these results are generalisable to other contexts, populations and times.

In this chapter, information that was derived from the research on the structured employment interviews analysed in this investigation, will be presented in order to comment on:

- The potential threats to internal validity to which they are exposed and how these may be overcome; and
- The potential threats to external validity to which they are exposed and whether or not these threats have appear to have affected the comparability of the findings across research efforts.

#### 4.1 The Exploration of Threats to Internal Validity

In this section the information that was essential for exploring the threats to the internal validity of the international validation research on structured employment interviews will be outlined.

#### 4.1.1 Research Design Strategies

Stern and Kalof (1996) suggested that a range of research design strategies may be employed in research initiatives. They identified six major types of research design strategies (viz. naturalistic observation, retrospective case studies, sample studies, correlational studies, within-subjects experiments and between-subjects experiments [with either non-equivalent or equivalent groups]) (Stern and Kalof, 1996). Stern and Kalof (1996) proposed that the first step in identifying the threats to internal validity, that plague research endeavours, is to gain an insight into the research design strategies that they employ. In order to determine the research design strategies that were used in the studies included for analysis in the present investigation, the author drew on Stern and Kalof's (1996) decision-making procedure for identifying research design strategies. This analysis constituted the point of departure for the critical evaluation of the studies included in the sample. The results of this undertaking are provided in Table 4.1.

Table 4.1 Classification of Studies in terms of Research Design Strategies

STUDIES	RESEARCH DESIGN STRATEGIES
<b>BEHAVIOUR DESCRIPTION INTERVIEWS</b>	
<i>PATTERNED BEHAVIOUR DESCRIPTION INTERVIEWS</i>	
1. Janz (1982)	Within-subjects experiment
2. Orpen (1985)	Within-subjects experiment
<i>STRUCTURED BEHAVIOURAL INTERVIEWS</i>	
3. Motowidlo et al. (1992) [Study 2]	Correlational study
4. Motowidlo et al. (1992) [Study 3]	Correlational study
5. Motowidlo et al. (1992) [Study 4]	Correlational study
<b>SITUATIONAL INTERVIEWS</b>	
6. Latham et al. (1980) [Study 1]	Correlational study
7. Latham et al. (1980) [Study 2]	Correlational study
8. Latham et al. (1980) [Study 3]	Correlational study
9. Latham and Saari (1984) [Study 1]	Correlational study
10. Latham and Saari (1984) [Study 2]	Correlational study
11. Weekley and Gier (1987)	Correlational study
12. Robertson et al. (1990)	Correlational study
13. Stohr-Gilmore et al. (1990)	Between-subjects experiment, non-equivalent groups

In general, the findings yielded by studies that do not possess internal validity may be explained by extraneous variables that interact with the independent variables and confound their effect on the dependent variables. This implies that alternative explanations may exist for the findings obtained in the context of such investigations. The identification of the research design strategies employed in the validation research on structured employment interviews, reveals that, typically, the studies, which were analysed, employed two broad types of research design strategies (viz. correlational studies and experimental studies).

### (i) Correlational Studies

According to Stern and Kalof (1996), correlational studies measure two or more variables. They strive to assess the relationship between multiple variables without manipulating any of them (Stern and Kalof, 1996).

### (ii) Experimental Studies

Experimental studies measure the effect of one variable on another by manipulating the independent variable and observing the effect on the dependent variable (Stern and Kalof, 1996). There are two types of experimental research designs (viz. within-subjects and between-subjects experiments) (Stern and Kalof, 1996). In this section, the author will offer an insight into these two types of experimental research designs.

#### (a) Within-subjects Experiments

In within-subjects experiments, researchers measure an effect by comparing the behaviour or performance of the same group of subjects under different conditions of the independent variable (Stern and Kalof, 1996).

#### (b) Between-subjects Experiments

Researchers using between-subjects experiments measure an effect by comparing the performance of different groups of individuals under different conditions of the independent variable (Stern and Kalof, 1996). Stern and Kalof (1996) proposed that, in between-subjects experiments, researchers may randomly assign subjects to different conditions of the independent variable in order to attain comparability between the two groups of subjects under investigation. Studies of this nature make use of equivalent groups. However, they noted that it is not always possible to attain comparability between the groups of subjects used in between-subjects experiments through the process of random assignment (Stern and Kalof, 1996). In these specific instances, non-equivalent groups are used. These groups are presumed to be comparable but they may differ in significant ways that could affect the dependent variable (Stern and Kalof, 1996).



4.1.2 Threats to Internal Validity and Research Design Strategies

Stern and Kalof (1996) proposed that different research design strategies are prone to unique threats to internal validity. For the purpose of this investigation, the author will present the threats to internal validity, that have been identified for the research design strategies to which the studies included for analysis in the present investigation, conform, in Table 4.2. This table is an excerpt of a table proposed by Stern and Kalof (1996, p.104).

Table 4.2 Threats to Internal Validity on the Basis of Research Design Strategy

THREATS TO INTERNAL VALIDITY	RESEARCH DESIGN STRATEGIES		
	Correlational Studies	Within-Subjects Experiments	Between-Subjects Experiments, Non-Equivalent Groups
Time-tied variables	* (If subjects are compared to themselves)	***	**
Organismic variables	***	- (Held constant)	**
Invalid operational definitions	**	**	**
Sampling bias	*	*	*
Uncontrolled variation in information	*	*	*
Selective or distorted memory	A source of alternative explanation when research relies on memory		
Researcher distortion	*	*	*
Researcher selectivity	-	-	-
Incomplete access	*	*	*
More persistent changes due to research	*	*	*
On-stage effects	*	*	*

(Stern, P.C. and Kalof, L. (1996). Evaluating Social Science Research (Second Edition). Oxford: Oxford University Press, p 104).

According to Stern and Kalof (1996), the asterisks in this table denote common sources of alternative explanations in the context of specific research design strategies. The number of asterisks represents the severity of the threats in question with more asterisks denoting more severe threats and fewer asterisks representing less severe threats.

Table 4.2 reveals that organismic variables and time-tied extraneous variables constitute the most common and severe sources of alternative explanations in correlational research designs and within-subjects experiments respectively. These two sources of alternative explanations also feature as threats in between-subject experiments that use non-equivalent groups. However, they are less common and severe in this context. According to Table 4.2, they exist at the same level as the threat of invalid operational definitions.

Drawing on the work of Stern and Kalof (1996), the author captured the definitions for each of the sources of alternative explanations in Table 4.3. In this investigation, the aim is to explore these threats in the context of the investigations that have been included for analysis. In order to do this effectively, it is imperative to have an insight to what the threats are in order to speculate about how they manifest in research endeavours and how they may be managed effectively. The author has had the benefit of access to the work by Stern and Kalof (1996). In an attempt to offer the reader an insight into how these threats were defined by these writers, the author embarked on this undertaking. It is hoped that these definitions will assist the reader in the next chapter when the manifestations of the threats to internal validity will be explored in relation to the research that was included for analysis in the present investigation.

Table 4.3 Definitions of Threats to Internal Validity

THREATS TO INTERNAL VALIDITY	DEFINITIONS
Time-tied variables	When subjects are observed over a period of time any changes that are observed in them could be confounded by the passage of time. In other words, in addition to being outcomes of the manipulation of the independent variable, any changes that are observed in relation to the subjects may be a function of time-tied extraneous variables.
Organismic variables	This threat is introduced into the research setting by the subjects that participate. In this threat to internal validity, factors such as race and gender could serve as sources of alternative explanations for the findings that are obtained.
Invalid operational definitions	The existence of this threat suggests that the variables being investigated may be measuring something other than that which they claim to measure.
Sampling bias	Biased samples refer to unrepresentative samples that contain systematic error, which render them consistently different from the populations from which they are drawn.
Uncontrolled variation in information	This difficulty arises because it is not always possible to collect the same information from or about different people in the context of research efforts. When data is collected about several people on different occasions it is possible that the results obtained could be a function of who collected the information, how, when and where. Thus, these results may be attributable to extraneous variables that interact with the independent variable to yield an effect.
Selective or distorted memory	This threat is prevalent when both researchers and subjects are required to rely on their memory for the purpose obtaining data for research investigations. Stern and Kalof (1996) suggested that what someone remembers is not only incomplete. It is systematically incomplete. In other words, they proposed that memories may be systematically biased on the basis of the theories that subjects and researchers may have about the relationships between events. In the context of these theories everything that they perceive as unimportant tends to be forgotten. This leads to biased samples of information on the basis of which the results for research investigations are derived.
Researcher distortion	This threat is introduced by researchers who render distorted interpretations of information they gather in the context of research investigations, on the basis of their pre-conceived notions about the subject matter.

Table 4.3 (continued)

THREATS TO INTERNAL VALIDITY	DEFINITIONS
Researcher selectivity	This threat is introduced into research investigations by researchers when their individual biases influence the observations they make when they focus on aspects that they deem important to the exclusion of factors that they do not consider to be important. Consequently, the results, based on these observations, are flawed.
Incomplete access	This threat is manifest when researchers experience difficulty gaining access to the phenomena they wish to investigate. This threat is often problematic when the subjects in investigations wish to conceal information from or deceive researchers. As a result, researchers are not exposed to all the relevant information that they wish to investigate. The interpretations that they make, in the light of incomplete access to information, are likely to be biased impressions. This bias is introduced, not by the researcher, but by the subjects' efforts to conceal or deceive.
More persistent changes due to research	This threat exists as a function of the presence of an observer in a research situation but its effects are more pervasive insofar as it results in more long-lasting changes in subjects' behaviour which extend outside the context of the research situation. These persistent changes result from some extraneous variables such as the element of novelty, which is inadvertently introduced into the research situation. These variables serve to confound the impact of the independent variable and act as sources of alternative explanations for the behaviour of research subjects.
On-stage effects	This threat to internal validity occurs merely as a function of the presence of an observer in a research situation. When subjects are aware that they are being observed in research endeavours they may act in certain ways that may differ from the manner in which they typically behave. Therefore, the results that researchers arrive at on the basis of these observations are likely to be inaccurate reflections of subjects' actual behaviour. Consequently, these results are likely to be biased.

The threats to internal validity call the findings that are yielded by research endeavours into question. In the next chapter, equipped with this preliminary information, the author will explore the threats to internal validity that could have detracted from the accuracy of the findings yielded by the research on the three variants of structured employment interviews that were included for investigation (viz. patterned behaviour description interview, structured behavioural interviews and situational interviews) in this study.

In relation to the specific threat of sampling bias the author believes that information pertaining to how the samples were obtained is invaluable in commenting on the internal validity of research efforts. To this end, the sampling techniques used in each of the studies that were included in the present sample, were explored. These findings are presented in Table 4.4.

Table 4.4 Classification of the Sampling Techniques

STUDIES	DESCRIPTIONS OF HOW THE SAMPLES WERE OBTAINED	SAMPLING TECHNIQUES
<b>BEHAVIOUR DESCRIPTION INTERVIEWS</b>		
<i>PATTERNED BEHAVIOUR DESCRIPTION INTERVIEWS</i>		
1. Janz (1982)	A lottery for the amount of \$100 was offered to teaching assistants who were willing to participate as interviewees in the research undertaking. The sample consisted of individuals who were willing to participate in the research.	Convenience sampling
2. Orpen (1985)	The sample consisted of all the individuals from a large life insurance company that applied for a life insurance sales position within the company. None of these individuals had been pre-screened by any device prior to the interview.	Saturation survey
<i>STRUCTURED BEHAVIOURAL INTERVIEWS</i>		
3. Motowidlo et al. (1992) [Study 2]	The sample was comprised of individuals that occupied positions in various functional areas within the organisations that participated in the research effort. Although it is not clear how these individuals were selected for inclusion in the sample, the researchers did state that the interviewees were asked to pretend that they were applying for the positions they currently occupied, during the interview.	Unknown
4. Motowidlo et al. (1992) [Study 3]	In this study the sample consisted of applicants for entry-level management positions in a telecommunications company. It is not clear whether or not this sample included all the individuals that applied for these positions at the time of the research. In other words the exact criteria used for including subjects in the sample were not specified.	Convenience sampling
5. Motowidlo et al. (1992) [Study 4]	The sample was comprised of individuals that occupied marketing positions in various functional areas within the organisations that participated in the research effort. Although it is not clear how these individuals were selected for inclusion in the sample, the researchers did state that the interviewees were asked to pretend that they were applying for the positions they currently occupied, during the interview.	Unknown
<b>SITUATIONAL INTERVIEWS</b>		
6. Latham et al. (1980) [Study 1]	The sample in this study was randomly selected from the total population of hourly-unionised sawmill workers within a company.	Random sampling
7. Latham et al. (1980) [Study 2]	The sample was comprised of first-line foremen within a company. The researchers merely offered information on the composition and characteristics of the sample and neglected to comment on or offer insights into how the sample was obtained.	Unknown
8. Latham et al. (1980) [Study 3]	The sample consisted of applicants for entry-level work in a pulp mill. The researchers were not explicit about how they obtained the sample of applicants for their investigation. They did reveal, however, that all of the applicants interviewed were subsequently hired in order to render a predictive validation study possible.	Saturation survey
9. Latham and Saari (1984) [Study 1]	The sample in this study was comprised of all the office clerical personnel in a regional office of a major wood products company.	Saturation survey
10. Latham and Saari (1984) [Study 2]	The sample that was used to attain the predictive validity estimate associated with the interview in question consisted of the group of individuals (derived from a larger group of recruits) that was hired for entry-level utility work in a newsprint mill.	Convenience sampling
11. Weekley and Gier (1987)	The sample was derived from a larger group of applicants for the position of sales associate and consisted of the individuals that were ultimately hired for the position.	Convenience sampling

Table 4.4 (continued)

STUDIES	DESCRIPTIONS OF HOW THE SAMPLES WERE OBTAINED	SAMPLING TECHNIQUES
12. Robertson et al. (1990)	In this study the sample was comprised of internal candidates in clerical and administrative positions, which entailed a large degree of customer contact, who were being considered for selection and promotion within an organisation.	Convenience sampling
13. Stohr-Gilmore et al. (1990)	The sample was comprised of applicants for the position of correctional officer from three separate groups interviewed at different points in time between 1985 and 1987. The validation phase of this study was conducted with the total population of the individuals that had been hired during this period.	Convenience sampling

## 4.2 The Exploration of Threats to External Validity

Stern and Kalof (1996) suggested that the best way to evaluate external validity is to compare the findings of studies that deal with the same basic question in different ways. In this investigation, the author intends to take heed of this advice in an attempt to ascertain whether or not the findings yielded in the context of studies within the categories of patterned behaviour description interviews, structured behavioural interviews and situational interviews respectively, are generalisable. The author has decided to approach the evaluation of external validity in two ways. Firstly, a qualitative exploration of the potential threats to external validity, that the studies within each of the afore-mentioned categories could have been exposed to, will be engaged in. This discussion will be premised on the key threats to external validity that Dooley (1995) identified. These are presented with definitions in Table 4.5. Secondly, the validity estimates yielded by the studies within each category will be compared using a meta-analytic technique proposed by Rosenthal (1991) in order to ascertain whether they are statistically consistent or significantly different from one another. The author believes that if the results are statistically consistent then the studies may possess external validity. This rationale is based on the assumption that the manner in which research is conducted impacts on the findings that are yielded and that if the findings are comparable, then despite the differences that are evident in the research investigations, they do possess external validity.

Table 4.5 provides an insight into the threats to external validity advocated by Dooley (1995). These threats will be explored in the context of the investigations included in each category of structured employment interview in the present study.

Table 4.5 Definitions of the Threats to External Validity

THREATS TO EXTERNAL VALIDITY	DEFINITIONS
Setting-by-intervention interactions	<ul style="list-style-type: none"> <li>• Experimental settings: The observed effect of the independent variable is a function of its combination with some aspect of the experimental arrangements.</li> <li>• Context: The observed effect of the independent variable is a function of its interaction with some aspect of the social or physical environment.</li> </ul>
History-by-intervention interaction	The observed effect of the independent variable is due to its combination with a recent event or with a particular era.
Selection-by-intervention interactions	The observed effect of the independent variable is attributable to its interaction with some of aspect of the particular subject sample.

Using the threats to external validity proposed by Dooley (1995) as a guide, the author explored the potential for the existence of these threats in the investigations that were included for analysis in this

study. However, merely on the basis of this exploration, it was not possible to comment on the external validity of these investigations.

In the quest to offer commentary on whether or not the studies in the sample appear to possess external validity, the author used a meta-analytic technique, which was devised by Rosenthal (1991). This technique was designed to compare the quantitative results yielded by independent research efforts. On the basis of these comparisons, the author commented on whether or not the findings were consistent with each other. Rosenthal (1991) proposed that quantitative findings and the sizes of the samples that were used to attain them, were vital for the comparisons. This information is presented in Table 4.6 together with additional information about the type of criterion measure that was used in each investigation and the type of individuals that comprised the subject groups in each study.



Table 4.6 A Summary of the Quantitative Validity Results

STUDIES	SAMPLE SIZE	SAMPLE COMPOSITION	VALIDITY RESULTS			
			Predictive Validity	Concurrent Validity	The Criterion	Construct Validity
<b>BEHAVIOUR DESCRIPTION INTERVIEWS</b>						
<i>PATTERNED BEHAVIOUR DESCRIPTION INTERVIEWS</i>						
1. Janz (1982)	15	Teaching assistants	r = 0.54 [unstructured interviews r = 0.07]		student evaluations	
2. Orpen (1985)	19	The population of applicants from routine sales positions in a large life assurance company	r = 0.48 [unstructured interviews r = 0.08]		supervisor ratings	
			r = 0.61 [unstructured interviews r = 0.05]		value of sales expressed in dollars per year	
<i>STRUCTURED BEHAVIOURAL INTERVIEWS</i>						
3. Motowidlo et al. (1992) [Study 2]	164	Management incumbents drawn from a range of functional areas		r = 0.23	supervisor ratings	convergent validity = 0.25 discriminant validity = 0.20
4. Motowidlo et al. (1992) [Study 3]	195	Applicants for entry-level management positions	r = 0.17		supervisor ratings	convergent validity = 0.15 discriminant validity = 0.14
5. Motowidlo et al. (1992) [Study 4]	165	Marketing incumbents from a range of functional areas		r = 0.32	supervisor ratings [behaviour descriptions]	convergent validity = 0.18 discriminant validity = 0.12
				r = 0.21	supervisor ratings [activity statements]	convergent validity = 0.21 discriminant validity = 0.15



Table 4.6 (continued)

STUDIES	SAMPLE SIZE	SAMPLE COMPOSITION	VALIDITY RESULTS			
			Predictive Validity	Concurrent Validity	The Criterion	Construct Validity
<b>SITUATIONAL INTERVIEWS</b>						
6. Latham et al. (1980) [Study 1]	47	Unionised hourly saw mill workers		r = 0.46 [r = 0.41 when experience is partialled out]	supervisor ratings [total BOS scores]	
				r = 0.50 [r = 0.46 when experience is partialled out]	supervisor ratings [overall performance ratings]	
7. Latham et al. (1980) [Study 2]	61	First-line foremen		r = 0.30 [r = 0.29 when experience is partialled out]	supervisor ratings	
8. Latham et al. (1980) [Study 3]	56 [Blacks]	Applicants for entry-level work at a pulp mill	r = 0.33 [Blacks]		supervisor ratings	
	30 [Females]	Applicants for entry-level work at a pulp mill	r = 0.39 [Females]		supervisor ratings	
9. Latham and Saari (1984) [Study 1]	29	Office clerical personnel in regional office of a wood products company		r = 0.39 [past experience interviews r = 0.14]	supervisor ratings [ $\alpha = 0.96$ ]	
				r = 0.42 [past experience interviews r = 0.15]	peer evaluations [ $\alpha = 0.96$ ]	
10. Latham and Saari (1984) [Study 2]	157	Entry-level utility employees in a newsprint mill	r = 0.14 [for entire sample]		supervisor ratings	
	29	A randomly selected sample from the initial 157 employees	r = 0.40 [for randomly selected portion]			
11. Weekley and Gier (1987)	24	Applicants for sales associate position	r = 0.45 [under ideal conditions when corrected for attenuation r = 0.47]		sales productivity	

Table 4.6 (continued)

STUDIES	SAMPLE SIZE	SAMPLE COMPOSITION	VALIDITY RESULTS			
			Predictive Validity	Concurrent Validity	The Criterion	Construct Validity
12. Robertson et al. (1990)	63	Internal candidates in clerical or administrative jobs with a high degree of customer contact	r = 0.28 [corrected r = 0.38]		performance appraisal scores [ $\alpha = 0.88$ ]	
			r = 0.33 [corrected r = 0.43]		potential criterion measure (derived from above) [ $\alpha = 0.78$ ]	
13. Stohr-Gilmore et al. (1990)	33	The population of candidates hired in 1987. These individuals comprised the 2 <sup>nd</sup> group of subjects in the study.	r = 0.19 [without situational questions r = 0.15]		first performance evaluation	
			r = 0.39 [without situational questions r = -0.04]		second performance evaluation	

Typically in validation research, validity estimates of 0.3 and above are regarded as acceptable. In the studies that were included for analysis in the present investigation, the author noted that the estimates for patterned behaviour description interviews were higher than 0.3 and were significantly higher than the validity estimates obtained for unstructured interviews. Therefore, the researchers who conducted these investigations proposed that the use of patterned behaviour description interviews would yield more valid predictions of future job performance (Janz, 1982; Orpen, 1985).

Conventional wisdom in the context of employment interview literature led writers and researchers to conclude that greater degrees of structure in interviews implied higher validities (Huffcutt and Arthur, 1994). Based on this argument, the validity estimates associated with the structured behavioural interviews (SBIs), which represented a more structured variant of employment interview than patterned behaviour description interviews (PBDIs), are expected to be higher than those obtained in the research on PBDIs. However, the results of the validation studies by Motowidlo et al. (1992) revealed that this was not the case. The only acceptable validity estimate that these researchers obtained was in their fourth validation study using the criterion of supervisors' ratings of subjects, which was based on behaviour descriptions.

In the author's opinion, this discrepancy may be attributed to the use of comparatively larger samples in the investigations on SBIs and smaller samples in the investigations on PBDIs. Dyer (1995) suggested that when very small samples are used in research (like the ones used in the research on PBDIs) then the presence of extreme values in the data exert considerably more influence on the holistic picture that is presented by the sample data. In other words, in the context of employment interview validation research based on small samples, extreme scores in relation to the predictor and criterion tend to either inflate or deflate the average scores that are obtained. These, in turn, influence the validity estimates, which are based on the correlations between the scores on the predictor and criterion variables. These estimates may either be inflated or deflated depending on the direction of the extremity. When larger samples are used, Dyer (1995) suggested that extreme values tend to be tempered by a larger number of moderate scores. Hence, the validity estimates that are attained are likely to be less prone to distortion by extreme values in the predictor and criterion and consequently, are likely to be more accurate.

Based on this insight, it is possible that the validity estimates in the research on PBDIs were distorted by extremely high values attained in the context of the predictor and criterion variables. This distortion was possible due to the use of small samples by Janz (1982) and Orpen (1985). The author believes that the possibility of such distortion may have been reduced in the studies on SBIs, which drew on much larger samples. Hence, the comparatively lower validity estimates.

In general the validity estimates associated with situational interviews were in the vicinity of 0.3. Latham and Saari (1984) and Stohr-Gilmore et al. (1990) yielded estimates of 0.14 and 0.19 respectively. The sizes of the samples used in the validation efforts on situational interviews were larger than those used by Janz (1982) and Orpen (1985) in the research on PBDIs. However, with the exception of the first study by Latham and Saari (1984), all the samples consisted of less than 100 individuals. It is possible that the validity estimates associated with situational interviews could have been distorted by the sizes of the samples that were used to obtain them.

Whetzel and Oppler (1997) identified two factors that informed the appropriateness of the sample sizes used for validation research (viz. the nature of the statistics to be computed and the level of reliability that is desired for these statistics). They proposed that, typically, the larger the number of subjects included for analysis in validation studies, the more dependable or reliable the results will be, provided that the samples are representative of the target populations (Whetzel and Oppler, 1997). A technique called power analysis offers insights into the acceptability of sample sizes on the basis of the extent of the reliability that is required for research outcomes (Whetzel and Oppler, 1997). The author believes that it is prudent for employment interview researchers to implement this technique to ascertain appropriate sample sizes in their validation efforts on structured interviews.

Using the estimates derived from the studies analysed in the present investigation, the author compared the validities yielded in the context of each category of structured employment interview (viz. patterned behaviour description interviews, structured behavioural interviews and situational interviews). There were only two studies to compare in the PBDI category. Therefore, the author used the meta-analytic technique for comparing two studies (Rosenthal, 1991) to compare the validity estimates yielded by these investigations. In relation to the research on structured behavioural interviews and situational interviews, there were more than two studies whose estimates required comparison. In these instances the author used the meta-analytic technique for comparing the findings of three or more investigations (Rosenthal, 1991). In addition, the author engaged in pair-wise comparisons of the estimates of each study with every other study in these categories using the technique for comparing the findings of two studies mentioned above.

The results of the pair-wise comparisons of the validity estimates in the studies on PBDIs, SBIs and SIs are presented in Table 4.7. The results of the comparisons of the findings yielded by three or more investigations, which were conducted for the investigations on SBIs and SIs are presented in Table 4.8.

Table 4.7 Pair-Wise Comparisons of the Validity Estimates for PBDIs, SBIs and SIs

STUDY	1	3	4	6a*	6b*	7	8a	8b	9a*	9b*	10a	10b	11a*	11b*	12a*	12b*	12c*	12d*	13a*	13b*	
2a*	0.21 0.417																				
2b*	-0.27 0.394																				
4		0.59 0.278																			
5a*		-0.88 0.189	-1.50 0.067																		
5b*		0.19 0.425	-0.39 0.348																		
6a*						0.94 0.174	0.76 0.224	0.35 0.363	0.35 0.363	0.20 0.421	2.08 <b>0.019</b>	0.30 0.382	0.05 0.480	-0.05 0.480	1.06 0.145	0.78 0.218	0.49 0.312	0.19 0.425	1.29 0.099	0.36 0.359	
6b*						1.20 0.115	1.01 0.156	0.56 0.288	0.56 0.288	0.41 0.341	2.39 <b>0.008</b>	0.51 0.305	0.24 0.405	0.15 0.440	1.32 0.093	1.04 0.149	0.75 0.227	0.45 0.326	1.51 0.066	0.58 0.281	
7							-0.18 0.429	-0.44 0.330	-0.43 0.334	-0.59 0.278	1.09 0.138	-0.48 0.316	-0.69 0.245	-0.79 0.215	0.12 0.452	-0.18 0.429	-0.49 0.312	-0.82 0.206	0.52 0.302	-0.45 0.326	
8a								-0.29 0.386	-0.29 0.386	-0.44 0.330	1.27 0.102	-0.34 0.367	-0.55 0.291	-0.65 0.258	0.29 0.386	0.00 0.500	-0.30 0.382	-0.62 0.268	0.66 0.255	-0.30 0.382	
8b								0.00 0.500	-0.13 0.448	1.30 0.097	-0.04 0.484	-0.25 0.401	-0.34 0.367	0.54 0.295	0.30 0.382	0.05 0.480	-0.21 0.417	0.83 0.203	0.00 0.500		
9a*											1.28 0.100	-0.04 0.484	-0.25 0.401	-0.33 0.371	0.53 0.298	0.29 0.386	0.05 0.480	-0.20 0.421	0.82 0.206	0.00 0.500	
9b*											1.45 0.074	0.09 0.464	-0.13 0.448	-0.21 0.417	0.68 0.248	0.45 0.326	0.20 0.421	-0.05 0.480	0.95 0.171	0.13 0.448	
10a												-1.33 0.092	-1.48 0.069	-1.59 0.056	-0.96 0.169	-1.33 0.092	-1.70 <b>0.045</b>	-2.10 <b>0.018</b>	-0.26 0.397	-1.36 0.087	
10b													-0.21 0.417	-0.29 0.386	0.58 0.281	0.34 0.367	0.10 0.460	-0.15 0.440	0.86 0.195	0.04 0.484	
11a*															0.78 0.218	0.56 0.288	0.33 0.371	0.10 0.460	1.03 0.152	0.26 0.397	
11b*															0.88 0.189	0.66 0.255	0.43 0.334	0.20 0.421	1.12 0.131	0.35 0.363	
12a*																			0.43 0.334	-0.56 0.288	
12b*																			0.67 0.251	-0.31 0.378	
12c*																			0.93 0.176	-0.05 0.480	
12d*																			1.20 0.115	0.22 0.413	

The italicised values in the table reflect Z-scores. The values that are presented below the Z-scores reflect the corresponding p-values, which indicate significance levels. p-values that are significant at the 1% level are indicated in red and p-values that are significant at the 5% level are indicated in blue.

The author held the order of the studies and the order in which the estimates were presented, in the context of each study, constant in Tables 4.6 and 4.7. Therefore, the numbers of the studies in Table 4.7 correspond to those assigned to the studies in Table 4.6. In Table 4.6 it is evident that some studies yielded more than one validity estimate. These independent estimates are reflected as the number of the study followed by an 'a', 'b', 'c' or 'd'. Rosenthal (1991) proposed that findings are only comparable if they are obtained independently of each other. In Studies 2, 5, 6, 9, 11, 12 and 13 the multiple validity estimates that were obtained were derived using the same samples of subjects. Therefore, according to Rosenthal's (1991) condition, they were not comparable. The author used the symbol ♦ to indicate the estimates that would not be compared with other estimates within the same study number. In Studies 8 and 10 the researchers used larger samples of subjects to obtain one estimate and then drew sub-samples from these initial samples to yield further estimates. The author opted to engage in comparisons of the results that were obtained in these studies since the samples used to attain these results were altered by the researchers. Therefore, in the author's opinion, these estimates were independent since they were yielded in the context of different samples.

The multiple validity estimates in Studies 11 and 12 were yielded as a consequence of the researchers who conducted these studies, correcting the initial estimates they obtained. In Study 11 the initial estimate was 0.45 and the corrected estimate was 0.47. In Study 12 the researchers obtained two initial estimates for the same group of subjects using two different criterion measures. The author labelled these 12a and 12b in Table 4.7. The researchers corrected each of these estimates. The author labelled the corrected estimate, which corresponded to the initial value in 12a, 12c and the corrected estimate that corresponded to the initial validity value in 12b, 12d. As a result of the fact that these estimates were obtained using the same group of subjects, the author did not engage in pair-wise comparisons of the multiple estimates yielded in each of these investigations.

In Table 4.7 the author noted four significant z-scores. All of these were obtained in relation to the validity estimate that was yielded in Study 10a. In general, all the Z-scores that were attained in the comparisons of the estimate yielded in this study with all the other investigations on situational interviews, tended to be higher than those obtained in the pair-wise comparisons of other situational interview results.

Table 4.8 provides an insight into the comparisons of the validity estimates, which were undertaken in the context of SBIs and SIs. Rosenthal (1991) proposed that when the findings of more than three studies need to be compared, then an alternative meta-analytic procedure is necessary. In the

light of the reality that there were more than three investigations that required comparison in each of these categories, the author employed this technique.

Table 4.8 Comparisons of Validity Estimates Yielded in Three or More Studies

<b>CATEGORIES</b>	<b>Number of Studies</b>	<b>Mean z-scores</b>	<b>Chi-squared Values</b>	<b>Degrees of Freedom</b>	<b>p-values</b>
Structured Behavioural Interviews	3	0.24	2.26	2	0.32
Situational Interviews	10	0.31	7.86	9	0.55

The validity estimates by Motowidlo et al. (1992), which are presented in Studies 3, 4 and 5a, were compared in order to ascertain whether these estimates were consistent with or significantly different from one another. The purpose of this comparison was to offer a tentative comment on the external validity of the findings that were obtained for structured behavioural interviews. This comparison suggested that these findings are consistent with each other. This implies that any differences that may have existed between these investigations did not appear to have impacted significantly on the results that were attained. The author compared the estimates yielded in Studies 6a, 7, 8a, 8b, 9a, 10a, 10b, 11a, 12a and 13b in order to ascertain whether or not the findings for situational interviews were consistent or statistically significantly different and in so doing to once again, offer a tentative comment on external validity. The results of this comparison suggested that the findings were consistent. This implied that the studies may have been generalisable since the differences between them did not appear to have affected the comparability of the effect sizes that they yielded.

When a study yielded more than one validity estimate using the same group of subjects, the author decided on which estimate to include in the numerical comparisons that were engaged in for more than three studies in the categories of structured behavioural interviews and situational interviews, using the following criteria:

- Estimates obtained using supervisors' ratings as the criterion were included;
- In Study 5 the author selected the validity estimate that was derived using a criterion measure that was evaluated using a behaviourally oriented rating scale rather than a graphic rating scale. This choice was informed by Pulakos' (1997) contention that the former tends to render more accurate and reliable results when coupled with appropriate rater training than the latter type of rating scale;
- In Study 6 the author opted for the validity estimate that was derived using the criterion of supervisors' ratings on behaviour observation scales as opposed to the overall performance score. This estimate was included because the criterion that was used to derive it was more well-defined in the author's opinion;
- In Study 9, the second validity estimate (9b) was not included because it was derived using the criterion of peer evaluations rather than supervisors' ratings. The majority of the researchers used supervisors' ratings as the criterion. Therefore, to ensure comparability between the results, the author only included those estimates that conformed to this generally widely used criterion measure;
- In Study 12, the author included the estimate that was derived from the performance appraisal but did not consider the inclusion of the estimate that was derived using potential as the criterion. Actual performance is more tangible than potential. Therefore, the former is likely to be assessed with more ease than the latter. Validity estimates that are based on actual performance are likely to be more well founded than those that are based on the elusive concept of potential. This belief eliminated the estimate that was included under 12b and its corrected counterpart, which the author referred to as 12d above; and
- Corrected validity estimates were not included in the meta-analytic analyses. Therefore, the estimates that were included under 11b and 12c were excluded.

Using this preliminary information on external validity the author will comment on potential threats to external validity to which that the investigations analysed in this study are likely to be prone. In addition, the author will use the meta-analytically derived comparisons of the results to comment on whether or not these investigations appear to be generalisable to other settings, people and time periods.



### 4.3 Conclusion

The information presented in this chapter constitutes the foundation upon which the author will engage in the exploration of the threats to internal and external validity in the next chapter.

## Chapter 5

### Discussion

#### 5.0 Introduction

In this chapter, the threats to internal and external validity that have plagued the research on structured employment interviews will be explored. This exploration will be conducted in the context of the three categories of structured employment interviews into which the research studies, analysed in this investigation, were clustered (viz. patterned behaviour description interviews, structured behavioural interviews and situational interviews). The preliminary insights provided in the Results chapter of this study will inform the present discussion.

Stern and Kalof (1996) proposed that the research design strategies that are typically used as the framework for conducting research are susceptible to unique threats to internal validity. These threats were outlined in the previous chapter in Table 4.2 for each of the research design strategies that were used in the studies analysed in this investigation. This discussion is oriented towards highlighting the threats of time-tied extraneous variables, organismic variables, invalid operational definitions, sampling bias and uncontrolled variation in information. Stern and Kalof (1996) suggested that these threats are characteristic of quantitative research endeavours. The remaining threats in Table 4.2 (viz. selective or distorted memory, researcher distortion, researcher selectivity, incomplete access, more persistent changes due to research and on-stage effects) are typical in qualitative research endeavours. The studies evaluated in this investigation were all quantitative in nature. However, the mere fact that they required the collection of data by human beings in the context of interviews and performance evaluations rendered them susceptible to these threats as well. Although the exposure of all the studies in this investigation, to these threats, is minimal, the author believes that it is imperative that their potential to have influenced the internal validity of these research initiatives be acknowledged. However, the potential effects of the threats that are characteristic of qualitative research endeavours will not be elaborated on this discussion. Instead, the focus will centre on those threats that are more typical in quantitative investigations.

In the exploration of the threats to internal validity in the research within the three categories of structured employment interviews, a commentary will be provided on the efforts that the researchers have made to combat these threats. In instances where no overt measures were taken by researchers, recommendations will be offered on how specific threats may be overcome in validation research on the type of structured employment interview in question.

External validity will be commented on by exploring the potential threats to which the various investigations in each category of structured employment interview may have been exposed. Thereafter, with the assistance of meta-analytic techniques for comparing the findings yielded by different research efforts, the author will comment on whether or not these potential threats impacted on the generalisability of the findings, encapsulated in each of these investigations, to other settings, populations and time periods.

### 5.1 Research on Patterned Behaviour Description Interviews

Two investigations, which yielded quantitative validity estimates, were located for the patterned behaviour description interview (PBDI). The study by Janz (1982) yielded one estimate while the study by Orpen (1985) yielded two validity estimates based on independent criterion measures. In addition to offering insights into validity, these studies were both designed to render quantitative estimates of test-retest reliability (Janz, 1982; Orpen, 1985) and a further aim of the study by Janz (1982) was to analyse the content of the questions used and the responses derived in the context of PBDIs and unstructured employment interviews. However, for the purpose of the present investigation, the focus will be on the quantitative validity estimates yielded for the PBDIs and the manner in which the studies were designed to obtain these findings. The additional insights rendered by these investigations deviated from the primary aim of this study and were thus, excluded from consideration.

#### 5.1.1 An Exploration of the Threats to Internal Validity

It was noted in the Methodology and Results chapters that a pre-requisite for commenting on internal validity, according to Stern and Kalof (1996), is an insight into the research design strategies that are used to conduct research.

##### 5.1.1.1 Research Design Strategy: Within-Subjects Experiments

The first study which offered a validity estimate for the PBDI (Janz, 1982) inspired Orpen (1985) to engage in a replication effort. In this replication endeavour, Orpen (1985) shared Janz's (1982) interest in comparing the validity and reliability estimates obtained by exposing a common group of interviewees to two different interviewing conditions (viz. patterned behaviour description interviewing and unstructured interviewing). However, Orpen (1985) did not undertake to replicate the content analysis that Janz (1982) engaged in. In order to compare the validity estimates (i.e. the

accuracy of the predictions made by patterned behaviour description interviewers and unstructured interviewers regarding the ability of interviewees to perform effectively in their specific jobs) that were obtained by exposing the same group of interviewees to different interviewing conditions, both Janz (1982) and Orpen (1985) used a within-subjects experimental research design strategy (Stern and Kalof, 1996). In these studies, each interviewee was interviewed four times, twice by interviewers trained in patterned behaviour description interviewing and twice by interviewers trained in unstructured interviewing (Janz, 1982; Orpen, 1985).

#### 5.1.1.2 Threats to Internal Validity in Within-Subjects Experiments

In this section the threats that are likely to have impacted on these investigations will be explored using the insights provided by Stern and Kalof (1996) on the threats to internal validity that characterise within-subjects research design strategies.

##### (i) Time-Tied Extraneous Variables

According to Stern and Kalof (1996) the most significant threat to the internal validity of within-subjects experiments are time-tied extraneous variables. These writers proposed that when subjects are observed over a period of time any changes that they manifest could be confounded by the passage of time (Stern and Kalof, 1996). In other words, in addition to being outcomes of the manipulation of the independent variable, any changes that are observed in the subjects may be a function of time-tied extraneous variables (Stern and Kalof, 1996). The benefit of using this research design strategy (which necessitates the exposure of subjects to two conditions of the independent variable at different points in time in order to assess the effects of this manipulation) is that the organismic variables, introduced into the study by the characteristics that subjects bring with them (i.e. race, age, gender, experience, etc.), is held constant (Stern and Kalof, 1996). Thus, the use of between-subjects experiments ensures that the effects of the manipulated variable are not confounded by organismic or subject variables.

In the author's opinion, time-tied extraneous variables could have manifested in the studies by Janz (1982) and Orpen (1985) in the following manner. All the subjects (interviewees) in these investigations were exposed to both the treatment conditions (i.e. the patterned behaviour description interviewing condition and the unstructured interviewing condition) (Janz, 1982; Orpen, 1985). The interviewees were exposed to four interviews (two PBDIs and two unstructured interviews) at four different points in time. Essentially, these researchers were keen to establish if the type of interview that was used would impact on the validity of the predictions rendered by appropriately trained interviewers. In the author's opinion, it is possible that the exposure of the interviewees to one of the

interviewing conditions could have altered the responses they offered to the questions that were asked in the other interviewing condition to which they were subsequently exposed, given that this exposure occurred at different times. For example, if all the interviewees were exposed to PBDIs first then their responses to questions in unstructured interviews could have been affected by their insights into the unique dynamics of patterned behaviour description interview questions. A similar argument would apply if all the interviewees were exposed to the unstructured interviewing condition first and then to the patterned behaviour description interviewing condition. The timing or order of the interviewing conditions to which interviewees were exposed could also have resulted in them offering responses, in subsequent interviewing conditions, which were based on the subtle cues that previous interviewers may inadvertently have offered to their initial responses.

These time-tied extraneous variables could have impacted on the validity estimates obtained for PBDIs and unstructured interviews if they were left unchecked. In other words, the validity results could have been attributed to the order of the interviewing conditions to which the interviewees were exposed. In order to ensure that interviewees were not exposed to conditions that would have altered the responses they offered in systematic ways, the researchers (Janz, 1982; Orpen, 1985) chose to divide them into two groups and to alternate the interviewing conditions for each group of interviewees. They exposed the first group to PBDIs first and then to unstructured interviews while the second group, were interviewed by unstructured interviewers first and then by patterned behaviour description interviewers. The technique used by Janz (1982) and Orpen (1985), to control for the possible threats posed by time-tied extraneous variables in their investigations, is known as counterbalancing (Stern and Kalof, 1996).

#### (ii) Invalid Operational Definitions

Stern and Kalof (1996) suggested that a further threat to the internal validity of within-subjects experiments is the phenomenon of invalid operational definitions. Although this threat is not as severe as the threat of time-tied extraneous variables in this research design strategy, it does warrant consideration in order to fully explore the factors that could account for possible alternative explanations for the results obtained in the studies by Janz (1982) and Orpen (1985). Stern and Kalof (1996) proposed that the possibility that operational definitions may be invalid (i.e. may be measuring something other than that which they claim to measure) is present whenever they are used in research efforts. There is no research procedure that can counteract the effects of invalid operational definitions (Stern and Kalof, 1996). However, if such definitions are used, then there is a discrepancy between what researchers claim is being investigated and what is actually being investigated.

In essence validation research, in the realm of employment interviews, is designed to assess the extent to which the performance of a group of subjects in a specific type of interview is able to predict their performance on the job. This implies that the interview variant under investigation constitutes the independent variable and the criterion measure constitutes the dependent variable. Typically, the performance of subjects in specific interview variants is correlated with their performance on specific criterion measures, which are oriented towards assessing job performance. Ultimately, therefore, subjects' performance, in relation to a specific interview variant and the criterion measure that is used, is determined by the nature of the instruments (i.e. interview or performance evaluation instruments), in terms of their content, the manner in which they are conducted and the way in which they are scored or evaluated. Conventional wisdom derived from employment interview research and performance evaluation literature suggests that the more structured the instruments, the more consistent and objective the evaluations that are derived from their use.

In the investigations on PBDIs, the researchers were interested in comparing the validity estimates yielded by PBDIs and traditional unstructured interviews (Janz, 1982; Orpen, 1985). In order to accomplish their aim they varied the type of interview to which they exposed their subjects (Janz, 1982; Orpen, 1985). They compared the correlations between the validity results obtained for each interview variant and the criterion measures used to obtain job performance data in each of the studies (Janz, 1982; Orpen, 1985). In this investigation, the primary focus is on the validity estimates associated with structured variants of employment interviews and on the research procedures employed to obtain these estimates. Therefore, the author will only offer further consideration to this information. The validity information that was yielded for unstructured employment interviews will be ignored in this investigation.

In validation research, in general, and in the validation efforts of the PBDI that Janz (1982) and Orpen (1985) undertook, in particular, the potential exists for the confounding of the operational definitions of the variables in question. PBDIs are designed to elicit information about subjects' past experiences and behaviours in order to infer whether or not they will be effective in demonstrating the behaviours that are deemed necessary for the job being selected for. This interview variant is based on the underlying belief that past behaviour is the best predictor of future behaviour. This belief guides and informs the rationale for its use in employment situations. In the author's opinion, the factors that could confound the independent variable (i.e. the factors that could serve as sources of alternative explanations for the findings) in validation research are language proficiency, memory inaccuracies and interviewer proficiency and a significant factor that could confound the dependent variable is the proficiency of the evaluators of job performance.

#### (a) Language Proficiency

Interviews constitute verbal interactions between individuals. This implies that they are premised on language. In situations where interviews are conducted in a language with which interviewees are not familiar, it is possible that the outcome of the interviews could be confounded with the ability of the interviewees to understand and communicate in the language in question. In the investigations by Janz (1982) and Orpen (1985) the researchers did not provide explicit information pertaining to this issue. Therefore, it was not possible to comment on whether or not such confounding occurred in these studies. However, the author does acknowledge that the possibility for confounding as a result of this factor could have existed in these studies and consequently, could have affected the results that they ultimately yielded. In general, this threat can be controlled by ensuring that interviews are conducted in a language which interviewees are able to understand and express themselves effectively in.

#### (b) Memory Inaccuracies

In PBDIs, interviewees are required to draw on their memories to relate past experiences and behaviours in response to the questions that interviewers pose to them. It is possible for their responses and the evaluations that are subsequently rendered, on the basis of these, to be confounded with the ability of the interviewees' to remember their past behaviours and experiences accurately. This confounding is inevitable in the realm of PBDIs. However, the effects of confounding as a direct result of memory inaccuracies may be reduced if subjects are asked to think of the most recent experiences that they have had and to recall the behaviours that they engaged in in the context of these experiences. Janz (1982) and Orpen (1985) did not engage in practices that overtly avoided the potential for this source of alternative explanation. Therefore, memory inaccuracies could have influenced the validity estimates obtained in these investigations.

#### (c) Interviewer Proficiency

Interviews are evaluated by human beings. Earlier in this study, the infallibility of human beings was noted. This implies that the evaluations and judgements that interviewers make could be marred by their personal preferences and subjective biases. In order to counteract the confounding that could occur as a result of the infallibility of interviewers, Janz (1982) and Orpen (1985) exposed their interviewers to training programs in order to impart the skills that were necessary for them to conduct and evaluate interviews effectively and consistently. In so doing, these researchers controlled for the

threat of confounding the independent variable with the inconsistency and subjectivity that could be introduced by interviewers.

#### (d) Proficiency of Job Performance Evaluators

The information derived from the dependent variable (i.e. the criterion) in validation research, could also be confounded by a range of extraneous variables. In the author's opinion, the most significant extraneous variable is inconsistency in the application and evaluation of the criterion measure. This, once again, points to the infallibility of human judgements and may be overcome with the use of structured and standardised performance evaluation measures, which provide clear guidelines for implementation and evaluation. In the investigations on PBDIs, the researchers did not provide sufficient detail on their criterion measures to allow a commentary on whether or not they were designed and used in a manner that overcame this threat.

Performance evaluators could also suffer the fate of memory inaccuracies if they do not keep accurate and up-to-date records of the information that they are ultimately responsible for evaluating subjects on. These extraneous variables could result in a situation in which, in addition to assessing the job performance of the subjects in question, the criterion measure could also be assessing the proficiency of the performance evaluators to execute their responsibilities effectively and their ability to recall performance evaluation information accurately. The information provided by the researchers in their investigations on the PBDI is insufficient to determine whether the extraneous variable of performance evaluators' memory deficits constituted a threat to internal validity.

However, it is possible that both of the afore-mentioned threats could have featured in the investigations on PBDIs. In order to counteract them, researchers could expose performance evaluators to standardised training programs to impart the skills that are necessary to effectively assess job performance. In addition, researchers could monitor the evaluations they (performance evaluators) make and request that these be substantiated with actual examples of on-the-job behaviour. In these ways, researchers could strive to ensure that the criterion evaluations are a function of actual job performance and not of inconsistencies and subjective issues introduced by performance evaluators.



### (iii) Sampling Biases

One of the major considerations that researchers are confronted with in the realm of sampling is whether or not the sample is representative (Stern and Kalof, 1996). A representative sample may be defined as one “that is not systematically distorted by some extraneous variable” (Stern and Kalof, 1996, pp. 85-86). Unrepresentative samples are, thus, samples that contain systematic error and are deemed to be biased (i.e. they are systematically or consistently different from the target population) (Stern and Kalof, 1996).

Janz (1982) used a sample of teaching assistants as the interviewees in his investigation. He obtained this sample by offering a monetary incentive to the individuals in his target population. Those individuals from this population who volunteered to serve as interviewees in the study were included. Thus, Janz (1982) derived his sample of interviewees using the technique of convenience sampling. In describing his sample and the method that he employed to obtain them, however, Janz (1982) did not provide details about how many volunteers there were and how he decided on which volunteer teaching assistants to include in his investigation.

The use of the convenience sampling technique for obtaining the interviewees could have resulted in a biased sample. It is possible that those teaching assistants who volunteered were lured by the promise of a monetary incentive. This appeal could have been informed by specific realities in their lives, which were independent of their positions as teaching assistants. Therefore, these individuals could have volunteered as a result of a specific need for the monetary gain they were promised. In this way, they could have differed in a consistent and systematic way from the other teaching assistants in the target population who did not volunteer to be interviewees.

Stern and Kalof (1996) noted that a common source of sampling bias is the use of volunteers as subjects for experiments. They argued that individuals who volunteer to participate in research endeavours are more likely to want to please researchers by engaging in behaviours that they believe researchers are looking for. As a result, they may be unusually well motivated to perform effectively (Stern and Kalof, 1996). In conducting their investigations, researchers may inadvertently emit subtle cues, known as demand characteristics, that inform subjects about what they are looking for (Stern and Kalof, 1996). Subjects, especially volunteers, who may have a desire to please, may latch onto these cues and act in a manner that conforms to their interpretation of the researchers’ expectations. Thus, in research in which volunteers are used as subjects, the desire to please may constitute an alternative explanation for their behaviour (Stern and Kalof, 1996). In Janz’s (1982) study, demand characteristics

could have impacted on the responses that interviewees offered to the questions they were presented with in each interview and interviewing condition due to the fact that he used volunteers as subjects.

These are only some of the many possible explanations that could be offered to argue the possibility that the sample Janz (1982) used may have been biased. The ultimate point is that this possibility could have resulted in the systematic distortion of the validity findings that were derived from the responses subjects offered to the questions they were asked in the interviews to which they were exposed. This implies that the validity findings obtained using the teaching assistants that participated voluntarily in this study may not have been applicable to the population of teaching assistants from which they were drawn.

Orpen (1985) made use of all the applicants from within a life assurance company that applied for the life assurance sales position. These applicants were not screened prior to the interview and were all subsequently hired, regardless of their performance in the interview (Orpen, 1985). In this investigation the researcher drew on the entire population of applicants that was available. Sarantakos (1998) suggested that when all the units of a target population are included for analysis in an investigation, then the technique that is used to obtain the subjects (in this case, the interviewees) is known as a saturated survey. This is the method that Orpen (1985) appears to have used to obtain his group of interviewees.

Black (1993) suggested that when a saturation survey (i.e. the entire population) is used to obtain the subjects for a study, then the findings that are yielded will necessarily be applicable to the entire population. Consequently, the use of this technique for obtaining subjects implies representativeness. A disadvantage associated with the use of entire populations of subjects becomes prevalent when these populations are very small (Black, 1993). This is often the case when a conveniently available group of individuals is used (Black, 1993). In these instances the representativeness that is attained is only applicable to a handful of individuals. This state of affairs seems to defeat the purpose of striving for representativeness in relation to the subjects that are used to obtain research findings.

In Orpen's (1985) investigation, a small population of 26 conveniently available applicants served as the interviewees. Thus, this study was potentially prone to the disadvantage associated with using a small population of subjects outlined above. Notwithstanding this disadvantage, however, Black (1993) suggested if researchers drew on small populations and acknowledged and recognised this reality, then their findings and conclusions will be sound since these will be derived from an insight into the entire population instead of being inferred from the statistics obtained by using a sample of individuals in order to comment on a larger population. The validity findings that Orpen's (1985)

investigation yielded, were thus, applicable to the entire population of applicants that applied for the life assurance sales position. Of the 26 applicants, however, seven individuals left the organisation within the first year. Hence, criterion data was only available for 19 applicants. The sample was ultimately comprised of these remaining 19 applicants. On the basis of the information provided by the researcher, the erosion of the original group of applicants occurred through natural attrition. Therefore, the resulting group of applicants did not appear to be systematically biased in any significant way.

However, the use of extremely small samples in the investigations by Janz (1982) and Orpen (1985) could have distorted the validity results that were obtained. This argument was explored in more detail in the Results chapter of this investigation. The only manner in which this distortion can be minimised is if researchers utilise larger samples in their validation efforts. They could draw on the technique of power analysis to mathematically compute acceptable sample sizes according to the requirements of their specific research initiatives (Whetzel and Oppler, 1997). However, the use of larger samples necessarily implies higher costs. Researchers, therefore, have to balance the need for statistical reliability (which pertains to the accuracy of their research outcomes) with practical financial constraints.

Stern and Kalof (1996) proposed that the only way to ensure that a sample is representative is to use a truly random sample. In the studies by Janz (1982) and Orpen (1985) it is clear that the researchers did not make use of this sampling technique. However, in Orpen's (1985) investigation, the use of the random sampling technique would not have been feasible given the small size of the target population. Therefore, in this instance, the inclusion of all the individuals that applied for the life assurance sales position was probably the most effective strategy to ensure that the group of subjects was representative and that the findings obtained using this group pertained equally and directly to all the individuals that applied for the position. In Janz's (1982) study the researcher could have ensured the representativeness of his sample of teaching assistants in one of the following ways. Firstly, he could have used the technique of random sampling. This technique would not have been viable if the target population from which he derived his sample of teaching assistants was too small to ensure representativeness or if the majority of these individuals were not willing to participate in the research undertaking. A second technique that Janz (1982) could have employed is that of purposive sampling. Stern and Kalof (1996) suggested that when it is not possible to use the random sampling technique, it is acceptable to choose a sample on a convenient basis provided that the researcher ensures that the sample is equivalent to the target population with respect to the variables that are deemed significant in the context of the investigation. Janz (1982) did not offer details about how he obtained his sample or from which population they were derived apart from informing the reader that the sample was

comprised of a group of teaching assistants who volunteered to participate in the research in exchange for a monetary incentive. Thus, he did not offer an insight into whether or not the sample was representative of the population in terms of the factors that he perceived as relevant to the research question (Janz, 1982). A third technique that could have been used to achieve representativeness is the use of the entire population of teaching assistants as the subjects.

#### (iv) Uncontrolled Variation in Information

According to Stern and Kalof (1996), a further aspect that impacts on the findings yielded by quantitative investigations is uncontrolled variation in information. This difficulty arises because it is not always possible to collect the same information from or about different people (Stern and Kalof, 1996). In the investigations by both Janz (1982) and Orpen (1985) it is possible that interviewees offered different responses to interviewers on the basis of the interviewers' age, sex, race or other characteristics that could have served as extraneous variables (Stern and Kalof, 1996). In Janz's (1982) study, for example, all the unstructured interviewers were males. The females who were interviewed by interviewers within this group could have responded on the basis of the individual interviewers' sex. As a result, their responses to the questions that were posed to them could have been different if the interviewers who conducted interviews with them were females or other males. In this manner, the information on which the validity findings was based could have been distorted. The same argument could be made for the male subjects who may have been exposed to female interviewers in the patterned behaviour description interviewing condition. The responses that male subjects offered to male interviewers and those that female subjects offered to female interviewers, in these investigations, may also have been prone to the effects of uncontrolled variation in information.

In Janz (1982) and Orpen's (1985) investigations, each subject was interviewed twice by patterned behaviour description interviewers and traditional unstructured interviewers. The researchers were not explicit about whether or not both iterations of the same type of interview were conducted by the same interviewer for each subject. The fact that these studies yielded test-retest reliability estimates alludes to the possibility that the interviews within each interviewing condition were conducted by the same interviewer for each subject. If this was, in fact, the case one may argue that the researchers held the interviewer constant for each subject in each interviewing condition and in so doing, succeeded in controlling for the uncontrolled variation in information which could have featured had this control technique not been employed.

The concept of holding the interviewer constant for each subject may be extended to holding the interviewer constant for all the subjects. In the studies by Janz (1982) and Orpen (1985) this would have translated into using one interviewer to conduct all the patterned behaviour description interviews and another interviewer to conduct all the traditional unstructured interviews. This would necessarily have had implications for the time-frames within which these studies could have realistically been conducted. It would have taken one interviewer a significantly longer period of time to conduct two interviews each with every subject. Therefore, the threat of time-tied extraneous variables, which is prevalent in the within-subjects experimental research design strategies employed by Janz (1982) and Orpen (1985), may have been exacerbated by an extension of the period over which the data for these studies was collected.

Stern and Kalof (1996) argued that the practice of holding interviewers constant is not always an effective solution to the problem of uncontrolled variation in information. On the basis of the above discussion, it seems that this criticism holds true in the investigations by Janz (1982) and Orpen (1985). Stern and Kalof (1996) proposed more effective methods for controlling the effects of uncontrolled variation in information. One such method is to hold the procedures that are executed in research efforts constant (Stern and Kalof, 1996). This study is oriented towards an exploration of the impact of threats to internal validity on the validity results that are yielded for structured employment interviews. Therefore, although Janz (1982) and Orpen (1985) investigated the validity associated with both PBDIs and unstructured employment interviews, main focus is on the estimates yielded for the PBDIs.

In conducting the PBDIs the researchers held a range of procedures constant. Firstly, in both the studies the dimensions assessed in the interviews were derived from a systematic job analysis procedure known as the critical incident technique. Secondly, the patterned behaviour description interviewers received a standardised training program to equip them with the requisite skills and knowledge to conduct and evaluate these interviews effectively and consistently. Thirdly, the interviewers in both these studies were required to complete the interviews in an allocated time period (i.e. 30 minutes). Fourthly, the dimensions assessed in each study were consistent and the questions asked in the interviews conformed to and were drawn from a standardised pattern. Fifthly, the responses provided in the interviews were evaluated in terms of consistent rating scales in each investigation. Finally, the type of rating scales used to evaluate job performance in the context of the criterion measure, were consistent in each investigation. The literature on employment interviews (Huffcutt and Arthur, 1994; Wagner, 1949) recognises the impact of greater standardisation and consistency in the interview on the validity estimates that are associated with them. The relationship that was discerned between standardisation and validity is the greater the standardisation or structure the higher the validity estimate.

Despite the consistency that was attained by holding these procedures constant in the PBDIs used in these investigations, the following factors could have accounted for uncontrolled variation in information. Although the training to which the patterned behaviour description interviewers were exposed in each study was consistent and standardised, the manner in which the interviewers implemented the skills and knowledge that they gleaned from these programs could have varied. This could have impacted on the evaluations that interviewers made and consequently, on the validity estimates that were based on this information. The researchers could have controlled for this specific manifestation of the threat of uncontrolled variation in information by holding the interviewer constant. It was mentioned earlier that in the investigations by Janz (1982) and Orpen (1985), this practice could have exacerbated the threat of time-tied extraneous variables. Another way in which this difficulty could have been counteracted is if the researchers employed monitoring procedures to ensure that interviewers conducted and evaluated the interviews in accordance with the skills and guidelines imparted during the training programs to which they were exposed. One way in which this monitoring could have occurred is by videotaping the interviews that were conducted. This practice appears to be most suitable for effective monitoring. However, if the researchers did not have the resources to engage in this type of monitoring, audio-taped reproductions of the interview interactions would have sufficed. It is important to note that monitoring procedures, like the ones proposed above, may only be employed with the consent of applicants. This consent is not always forthcoming. This poses a significant difficulty to researchers since it precludes the verification of the information that is used for the purpose of validation. Nevertheless, armed with information about the possible ways in which interviewers could introduce uncontrolled variation in information into the results that are ultimately obtained, researchers could make informed decisions about the interview information that should be discarded as a result of this potential threat, even if they were only provided with written transcripts of the interviews.

A further contributor to the lack of standardisation in the manner in which these studies were conducted, is the semi-structured nature of the patterned behaviour description interview. In these interviews, interviewees are not all asked to respond to the same questions. PBDIs are conducted according to a pre-defined pattern, however, and the questions that are selected by interviewers from this pattern do cover the same essential dimensions. Invariably, this reality introduces the possibility of uncontrolled variation in information because the use of different questions for different interviewees does not allow interviewers the opportunity to collect exactly the same information from all the subjects. This implies that each PBDI could be regarded as an independent interview, which may not be comparable with other PBDIs, which follow the same pattern but use different questions. The potential lack of comparability between the interviews implies the possible incomparability of the

evaluations derived from the responses provided by interviewees. Consequently, the validity estimates that are based on this information may be flawed.

A third factor that could have contributed to uncontrolled variation in information in these studies was the nature of the rating scales used to evaluate the interview responses and the performance of the subjects in their jobs. In these investigations, the researchers envisaged the use of graphic rating scales for evaluating interview and job performance. Pulakos (1997) suggested that graphic scales may be anchored with verbal or numerical anchors. In the studies by Janz (1982) and Orpen (1985) graphic scales with both verbal and numerical anchors were used. Janz (1982) employed a 5-point rating scale for the evaluation of both interview and job performance while Orpen (1985) devised a 7-point graphic rating scale for the evaluation of subjects' performance on the predictor and criterion measures.

Pulakos (1997) suggested that in order to be effective, graphic rating scales should not possess less than four and more than nine rating scale points. She argued that the former does not allow for sufficient discrimination among the individuals being assessed and the latter tends to yield insignificant differentiations among ratees (Pulakos, 1997). The most significant difficulty that was noted in relation to graphic rating scales is that the points on the scale are not thoroughly defined (Pulakos, 1997). This creates a situation in which the assignment of ratings is largely based on individual raters' interpretations of what the ratings at different scale points mean.

In the studies by Janz (1982) and Orpen (1985) the use of graphic scales could have resulted in different interpretations of the performance levels associated with each scale point. This, in turn, implies that the standards applied by the different interviewers could have lacked consistency. This would necessarily have impacted on their evaluations of interviewees and on the validity estimates based on these evaluations. One way in which Janz (1982) and Orpen (1985) could have overcome this difficulty is through the use of behaviour-based rating scales such as behaviour observation scales (BOS) or behaviourally anchored rating scales (BARS), since these are oriented towards overcoming the lack of specificity and definitional problems associated with graphic scales (Pulakos, 1997). Pulakos (1997) argued that even though research has revealed that no one type of rating format is psychometrically superior to other types of rating mechanisms, it has been empirically proven that well-developed behaviour-based rating scales in conjunction with appropriate rater training can yield ratings that possess highly acceptable levels of reliability and accuracy. This implies that the use of such scales could have significantly reduced the extent of uncontrolled variation in information in these investigations.

### 5.1.2 An Exploration of the Threats to External Validity

In order to determine the extent to which the validity findings yielded in the investigations by Janz (1982) and Orpen (1985) could be generalised, these studies were compared in terms of the research procedures they employed and the validity estimates they ultimately yielded. The following factors, which could have posed as significant threats to the external validity of these investigations, were identified.

Firstly, the manner in which the interviewers were selected in these studies differed markedly. Janz (1982) obtained his group of traditional unstructured interviewers by offering a monetary incentive to undergraduate students from business courses who volunteered to participate in his study. The group of patterned behaviour description interviewers consisted of individuals who were enrolled for a directed course in personnel research. These interviewers had prior training in the technique of patterned behaviour description interviews (Janz, 1982). Orpen's (1985) interviewers were randomly selected from the employees in the personnel department of the company within which the research was conducted. None of these individuals had any exposure to interviewing prior to the study (Orpen, 1985). In both these investigations the researchers designed training programs for the patterned behaviour description interviewers and the traditional unstructured interviewers in order to familiarise them with the interviewing techniques that they were required to implement. The fact that the interviewers Janz (1982) used received training in the technique of patterned behaviour description interviewing prior to the training program to which they were exposed in the study, implies that they are theoretically not equivalent and hence, not comparable to the patterned behaviour description interviewers that Orpen (1985) used in his investigation. Due to their more extensive training, the interviewers Janz (1982) used could have rendered more effective evaluations than the interviewers Orpen (1985) used. This could have resulted in the validity findings that were ultimately yielded in these investigations being significantly different. This, in turn, could have detracted from the extent to which the findings obtained in each of these studies could be generalised to other settings or contexts, groups of people and different time periods. In the context of this criticism, the validity estimates may be regarded as a function of the type of interview used in combination with the training to which interviewers were exposed (training constitutes an aspect of the experimental arrangement). Dooley (1995) referred to threats to external validity of this nature as setting (experimental)-by-intervention interactions.

Janz's (1982) investigation was conducted purely for research purposes while Orpen's (1985) study was conducted in an actual employment situation. Thus, the contexts in which the studies were conducted differed significantly from one another. This reality could have impacted on the manner in



which the patterned behaviour description interviews, in particular, were implemented in these different contexts and on the validity findings associated with them. In this criticism, the validity estimates obtained for the interviews could be regarded as a function of the type of interview used (i.e. the independent variable) in combination with the contexts in which the studies were conducted (i.e. research context vs actual employment context). According to Dooley (1995), these specific threats to external validity, which impede the generalisability of research findings to other situations, populations and time periods, are known as setting (context)-by-intervention interactions.

A third factor, which could have impeded the external validity of these studies, is closely related to the different contexts in which they were conducted. Janz (1982) drew on a group of individuals that already occupied a position (i.e. teaching assistants) to serve as the subjects (interviewees) in his investigation. Orpen's (1985) subjects, on the other hand, consisted of actual applicants for a life assurance sales position. It is possible for this difference to render the generalisability of the findings obtained in these studies, difficult, if not impossible. In the context of this criticism the validity estimates obtained in these investigations are perceived as a function of the type of interview used in combination with the nature of the individuals that constituted the subjects (actual applicants vs individuals that already occupied a position who volunteered to participate in a research investigation). Dooley (1995) referred to these threats to external validity as selection-by-intervention interactions.

Despite all these potential threats to the external validity of the findings yielded by the investigations undertaken by Janz (1982) and Orpen (1985), a pair-wise comparison of the validity estimates associated with PBDIs (in Table 4.7 in the Results chapter) revealed that these findings were statistically consistent. This suggests that the differences in the manner in which these studies were conducted did not impact significantly on the findings. The similarities in the research design and research procedures, however, could have played a major role in attaining this consistency. Orpen's (1985) investigation was intended to be a replication of the study conducted by Janz (1982). Therefore, these studies shared many similarities in terms of their experimental arrangements:

- These studies were both conducted according to a within-subjects experimental research design strategy which controlled for the most significant threat to internal validity (i.e. time-tied extraneous variables);
- The PBDIs in both investigations were developed and scored in exactly the same way;
- The duration of the interviews in both instances was 30 minutes;
- The training programs to which the patterned behaviour description interviewers and the traditional unstructured interviewers were exposed, in these investigations, was exactly the same; and
- The researchers in both these investigations used graphic rating scales evaluate the subjects' performance in the interview and on the job.

In the light of the statistical consistency of the validity estimates yielded by these studies, it appears that they would possess external validity provided they were shown to be internally valid. This belief is grounded in the evidence that the findings are comparable and consequently do generalise to other groups of people, contexts, settings and times. Thus, it seems that the differences inherent in these investigations did not impact significantly on the findings they yielded.

## 5.2 Research on Structured Behavioural Interviews

In this investigation, three studies that yielded quantitative validity estimates for structured behavioural interviews (SBIs) were located. These studies all formed part of a research report that was published by Motowidlo et al. (1992). In these studies, the researchers were primarily interested in obtaining criterion-related validity estimates for structured behavioural interviews. In addition, however, they derived insights into construct validity, correlations between the results of the SBIs and other interviews that were conducted previously as part of an initial hiring process for some subjects in some studies, correlations between the SBIs and other performance measures such as aptitude tests, grade point averages and class rankings in college and variations in the performance of different gender and race groups in the SBIs. Like in the previous category of studies on the patterned behaviour description interview, the focus will be solely on the criterion-related validity estimates yielded in these studies and the author will explore the impact of the manner in which these studies were designed on the results that were obtained. The additional information was not considered in the present investigation since it deviated from the primary and secondary research goals.

### 5.2.1 An Exploration of the Threats To Internal Validity

In this section, some of the characteristic threats to internal validity to which correlational studies are prone, will be explored in an effort to ascertain whether or not these featured in the studies conducted by Motowidlo et al. (1992). In addition, recommendations on how these threats were or could have been appropriately managed in these investigations will be provided.

#### 5.2.1.1 Research Design Strategy: Correlational Studies

In the investigations on SBIs by Motowidlo et al. (1992) the researchers' main purpose was to ascertain the validity associated with these interviews. In order to accomplish this aim, they correlated the evaluations rendered by interviewers using the structured behavioural interviewing technique with evaluations of job performance rendered by supervisors. This implies that in the studies on SBIs, the researchers did not manipulate the variables that they were interested in investigating. In terms of Stern and Kalof's (1996) classification scheme, these investigations fit the description of correlational studies.

#### 5.2.1.2 Threats to Internal Validity in Correlational Studies

Stern and Kalof (1996) proposed that organismic variables constituted the most severe threat to the internal validity of correlational studies. The remaining threats to internal validity are exactly the same as those discussed in the context of within-subjects experimental research designs. This is illustrated in Table 4.2 in the Results chapter of this investigation.

##### (i) Organismic Variables

Stern and Kalof (1996) defined organismic variables as those characteristics that subjects introduce into the research setting. These include attributes such as sex, race, age, experience, etc. (Stern and Kalof, 1996). In the second and fourth studies on the validity associated with SBIs, Motowidlo et al. (1992) drew on individuals that were already employed in a group of organisations that participated in the research endeavours (i.e. the researchers used a concurrent validation strategy). These subjects were explicitly asked to pretend that they were applying for jobs they currently held in the SBIs to which they were exposed. In these instances, it is possible that the experience the existing employees had gained during the time in which they were employed could have impacted on their responses to the SBI questions, the evaluations rendered by interviewers on the basis of these responses and ultimately

on the validity estimates yielded. In these studies, the researchers attempted to control for the extraneous effects of this potential organismic variable in two ways.

Firstly, they included individuals who had occupied their positions for between six and 24 months in order to minimise the impact of previous experience in the position being interviewed for on the interview responses and consequently, the validity estimates. Stern and Kalof (1996) referred to this control mechanism as matching. Technically, the process of matching entails the comparison of a group of individuals who are equivalent in terms of an extraneous variable in order to rule this variable out as the source of an alternative explanation for a specific hypothesised relationship (Stern and Kalof, 1996). In the investigations that yielded concurrent validity estimates for SBIs, the researchers acknowledged the impact that using employees with previous experience in the position being interviewed for as interviewees could have had on the results. Therefore, they employed the technique of matching to ensure only the inclusion of individuals with minimal experience (more than six but less than 24 months) in their samples of interviewees.

Secondly, the researchers explicitly asked the interviewees to refrain from drawing on the experiences they had encountered in their current positions, in their responses to the questions posed to them. Instead, subjects were encouraged to draw on the experiences that they encountered before they occupied their current positions. In the literature on methods to control the threats presented by organismic variables, the author did not encounter a control technique that could be used to describe the second practice that Motowidlo et al. (1992) employed to minimise the potential effects of previous experience on the validity findings that were yielded. However, in Motowidlo et al.'s (1992) validation research, this strategy could have assisted in minimising the effect of previous experience in the position being interviewed for as an alternative explanation for the findings that were obtained.

In the third study undertaken by Motowidlo et al. (1992), a predictive validation strategy was employed. The researchers utilised actual job applicants as their subjects (interviewees) for the purpose of this investigation. Apart from the insight that the individuals in the sample were included on the basis of the whether or not their criterion data (in the form of supervisors' ratings) was available within 24 months of being hired by their respective organisations, the researchers did not provide details about the criteria they used for selecting their applicants. In addition, they did not outline the characteristics of the sample of individuals that were ultimately included in the validation study. Thus, it was not possible to comment on whether or not there were any extraneous organismic variables that could have impacted on the validity results that were yielded.

However, according to the insights derived from the work of Stern and Kalof (1996), organismic variables could have posed a threat to the internal validity of this investigation. The researchers should have employed control measures in their predictive validation study to minimise the potential for the validity results to be a function of organismic variables that could have been introduced into the study by the subjects. The author believes that if the subjects were selected using the technique of random sampling then the potential impact of the organismic variables that they introduced into the investigation would have been randomised. Stern and Kalof (1996) recommended the use of a between-subjects experimental research design with the random assignment of subjects to different manipulations of the independent variable as a mechanism to control for the threat of organismic variables in the context of internal validity. However, Motowidlo et al. (1992) were merely concerned with ascertaining the validity of the predictions associated with the use of the SBI in their third investigation (i.e. their research aim was comfortably accommodated by a correlational research design strategy). Therefore, this method for controlling the impact of organismic variables does not appear to be appropriate for their specific research goals. Motowidlo et al. (1992) did not acknowledge the possibility that the results they obtained this investigation could have been distorted by subject variables. This appears to be a flaw in their investigation. These researchers should have explored the subject variables that were likely to have served as alternative explanations for their findings and controlled for them by using the technique of matching, for example (Stern and Kalof, 1996).

#### (ii) Invalid Operational Definitions

Stern and Kalof (1996) suggested that the threat of invalid operational definitions is second in significance to the threat of organismic variables in correlational research. In the validation efforts on the structured behavioural interview, the same potential extraneous variables, which were discussed in relation to PBDIs, could have confounded the operational definitions.

In the research on SBIs it is, once again, unclear if language proficiency on the part of subjects could have constituted an alternative explanation for the findings that were obtained. However, the possibility that this could have occurred exists. A sure-fire way to eliminate language proficiency as an alternative explanation is to ensure that subjects are proficient in comprehending and expressing themselves in the language in which interviews are conducted. Motowidlo et al. (1992) did not provide sufficient information in their article for a commentary to be offered on whether or not language proficiency could have explained the results yielded in these investigations. Nevertheless, it is imperative that future employment interview researchers understand the potential threat that this poses to their validation efforts and employ measures to combat this threat.

SBIs are premised on the same principle as PBDIs, which advocates that the best predictor of future behaviour or performance is past behaviour or performance. As a result, subjects are also required to draw on their memories in order to answer the questions that are posed in SBIs. Hence, it is possible that memory inaccuracies could serve as an alternative explanation for the results obtained in these interviews. Once again, there was insufficient information provided by Motowidlo et al. (1992) to comment on whether or not this factor featured as an alternative explanation in their studies. Researchers who employ behaviour description interviews should acknowledge the potential of this threat and take appropriate measures to control for it. In the discussion of invalid operational definitions in relation to PBDIs, it was proposed that one such measure is to ask subjects to recall their most recent experiences in their responses to the behaviour description questions that are posed to them.

Memory inaccuracies could have served as alternative sources of explanation in the second and fourth studies by Motowidlo et al. (1992). In these investigations, the researchers explicitly asked subjects to draw on the experiences they encountered before they commenced work in their current positions. Although this practice served as a control for the threat of the organismic variable of previous experience in the position being interviewed for, it could have exacerbated the threat of invalid operational definitions. The fact that the subjects were directed to remember events that had occurred in their distant past implies the possibility that the responses they offered to the SBI questions reflected their ability to remember these events as opposed to their actual performance in these events. As a result, the validity estimates that were based on this information could have been flawed. This scenario illustrates the possibility for researchers to introduce new threats to internal validity in their attempts to control for other threats.

The interviewers in these studies were exposed to training programs, which were designed to equip them with the necessary skills and knowledge to fulfil their responsibilities effectively and consistently. The researchers did not state that they engaged in this practice in an attempt to avoid the threat of invalid operational definitions. However, the author believes that exposure to training programs could assist to control for this threat by ensuring that interviewers are trained to elicit and evaluate job-related information as opposed to information that they may perceive as relevant on the basis of their subjective preferences, attitudes, views, biases and prejudices. Training programs could assist in ensuring that the evaluations interviewers make are a function of the requirements of the interviewing exercise and not their individual opinions, beliefs and values.

In the same way that the training of interviewers can assist to ensure that they evaluate what they are required to and thereby, minimise the threat of invalid operational definitions in validation research, the training of performance evaluators can also assist in minimising this threat. In the research on SBIs, the researchers did not expose the individuals they drew on to render criterion evaluations, to standardised training initiatives to ensure that they would be equipped to measure job performance in an objective manner. Therefore, the possibility exists that the criterion ratings could have been distorted by subjective elements introduced by performance evaluators. This could have resulted in a situation in which the evaluations that were made were not a function of the job performance of the subjects but rather of the subjective biases and opinions of the evaluators. In this manner the dependent variable (i.e. the criterion) could have been distorted. In the light of this possible threat, it seems wise for all employment interview researchers, who embark on validation efforts, to ensure that the individuals who will render performance ratings possess the requisite skills and knowledge to engage in their functions effectively. In this manner it is hoped that that they will assess what they are supposed to assess (i.e. job performance).

### (iii) Sampling Biases

In studies two and four by Motowidlo et al. (1992) the author was unable to comment on the sampling technique that was used to obtain the interviewees, on the basis of the information provided. Therefore, apart from the insight that the subjects were already employed in the companies that participated in the research, in the jobs for which they were interviewed, the researchers did not stipulate the criteria they used to obtain their interviewees. This insight alludes to the fact that the original population was a conveniently available group of employees. However, the technique that was used to select the sample from this population is not known. Therefore, a commentary of the possible sampling biases that could have influenced the results yielded by these two investigations, was not possible.

In the third study, which employed a predictive validation strategy, a convenience sampling technique appears to have been employed to obtain the 195 subjects that ultimately participated in the research. This sample was obtained by including all the interviewees for whom criterion data (in the form of supervisors ratings) became available within 24 months of being hired. The use of this technique calls the representativeness of the sample into question and it is possible that the sample was biased. This implies that the individuals in the sample could have differed systematically from the other individuals in the population, thereby precluding the extension of the results obtained in the context of the sample to the entire target population.

Stern and Kalof (1996) proposed that the most effective way to counteract the threat of sampling biases and, in so doing, to obtain a representative sample is to use the technique of random sampling. In the predictive validation study that Motowidlo et al. (1992) embarked on, it would have been preferable for the researchers to randomly draw a sample of subjects from the original population of individuals who qualified to participate in the investigation. They could have used the power analysis technique to ascertain the number of individuals that would have constituted an acceptable sample for the purpose of their study.

In all of their validation studies, Motowidlo et al. (1992) made use of relatively large samples. They used over 150 individuals in all their investigations. In the light of these relatively large samples, it is likely that the findings obtained by these researchers were not distorted by extreme values in the sample data (Dyer, 1995). Therefore, in relation to these investigations, the sample sizes could have been amenable to the attainment of unbiased and accurate validity results.

#### (iv) Uncontrolled Variation in Information

In the realm of social scientific research one instance in which the threat of uncontrolled variation in information is a reality, is when researchers require the same information to be collected from or about their subjects but have to rely on a number of different individuals to collect this information. Stern and Kalof (1996) proposed that subjects may respond differently to different interviewers on the basis of the characteristics they introduce into the research setting such as age, race, sex, etc. In the studies by Motowidlo et al. (1992) several interviewers conducted the SBIs with the samples of subjects and engaged in evaluations of the responses they (interviewees) offered to the questions they were asked in the interview interactions. In the second and fourth concurrent validation studies the researchers used 25 and 18 interviewers respectively and in the third study the researchers used 19 interviewers. In general, the researchers provided very little information about the interviewers. Therefore, it is not possible to comment on the characteristics that these individuals could have introduced into the investigations to fuel the threat of uncontrolled variation in information.

It was noted earlier, in the discussion of the threat of uncontrolled variation in information in relation to research on the PBDI, that a less than optimal way in which researchers could control for this threat is to hold interviewers constant. In the research on SBIs this practice would have rendered the individual validation studies practically impossible given the large numbers of subjects. The notion of one interviewer having to conduct interviews with 164 subjects in the second validation study, 195 individuals in the third and 165 in the fourth appears to be grossly unrealistic. If the researchers did



attempt to hold the interviewer constant in order to eliminate the potential threat of uncontrolled variation in information, the research would probably have taken several years to complete.

In the investigations by Motowidlo et al. (1992) the researchers employed a more effective control for the potential threat of uncontrolled variation in information by holding the procedures that they used constant:

- The dimensions assessed in the SBIs were derived from a rigorous job analysis procedure known as the critical incident technique;
- The dimensions assessed in each SBI were consistent and the same base questions were posed to subjects in the same order to assess these dimensions;
- The rating scales used to evaluate the responses to interview questions were standardised, behaviourally anchored and replete with descriptive examples of ineffective, adequate and effective behaviours;
- The job performance of subjects was evaluated with consistent behaviourally anchored rating scales; and
- In each study the interviewers were exposed to standardised training programs to equip them with the necessary skills to conduct and evaluate SBIs.

It was noted, in relation to the threat of uncontrolled variation in information in the discussion on PBDIs, that despite the consistent procedures researchers employed, the possibility for this threat still existed. This stance is maintained in the context of the studies on the SBIs. Although the training programs to which the interviewers were exposed in each study were consistent and standardised, it is possible, like it was in the case of PBDIs, that the manner in which the interviewers implemented the skills and knowledge that they gleaned from these training programs differed. It was suggested earlier that this likelihood could have impacted on the evaluations that interviewers made and consequently, on the validity estimates based on these evaluations. The recommendations proposed for curbing the impact of this threat in the research on PBDIs are applicable to the investigations on the SBIs as well. The researchers could have held the interviewer constant. However, it was noted that the practical implications that this would have posed were unrealistic. Secondly, the researchers could have employed monitoring procedures (i.e. they could have video-taped or audio-taped the interviews) to ensure that the interviewers performed their responsibilities effectively and consistently.

In the validation studies on the SBI, interviewers asked interviewees exactly the same base questions. However, on the basis of the responses they received, they were allowed the latitude to probe with

appropriate follow-up questions in order to obtain a complete picture about what the interviewees did in the situations they reported, the contexts in which they displayed their reported behaviours and the outcomes of these behaviours (Motowidlo et al., 1992). This implies that each interviewee could have been asked different discretionary probes and in this way the threat of uncontrolled variation in information could have been realised. However, in SBIs this threat is not as pronounced as it is in research on PBDIs because of the use of common base questions. Nevertheless, the use of SBIs, which espouse the use of discretionary probes, does introduce the possibility that the information collected by different interviewers from different interviewees will not be consistent and hence, the comparability of this information may be rendered difficult. It was noted earlier, in the discussion of the threat of uncontrolled variation in information in relation to research on PBDIs, that the incomparability of the interviews implies the incomparability of the evaluations derived from the responses provided by interviewees. In the light of this information, the validity estimates in the studies on SBIs may have been flawed.

In the research by Motowidlo et al. (1992) on SBIs two additional factors that could have reduced the potential impact of the threat of uncontrolled variation in information were identified. Firstly, the researchers conducted pilot studies in order to finalise and revise the structured behavioural questions that were ultimately included in the interview schedules. This preliminary groundwork could have ensured that only those questions that were clear, easily understood and unambiguous were included in the schedule, thereby eliminating the possibility that existed for interviewees to interpret the questions differently. In the absence of these pilot studies, it is possible that the information derived from and about the interviewees could have been different, inconsistent and consequently, incomparable. Secondly, the researchers used behaviourally anchored rating scales to evaluate the interview performance and job performance of the interviewees. In general, behaviourally based rating scales are well defined (Pulakos, 1997). Their design promotes consistency in the manner in which they are used and ensures standardisation in the manner in which various rating scale points are interpreted by evaluators. This implies that with the use of behaviourally based scales, uncontrolled variation in the evaluations of both interview and job performance may have been minimised and possibly avoided.

### 5.2.2 An Exploration of the Threats to External Validity

The following threats to external validity could have impacted on the extent to which the results yielded by the validation efforts of SBIs could be generalised to other contexts, groups of people and time periods.

#### (i) Setting (experimental setting)-by-Intervention Interactions

In the three studies in question, it is possible that the manner in which the investigations were designed and conducted could have interacted with the independent variable (i.e. the interviews investigated) and in so doing, could have confounded the validity estimates that were obtained. In each of these studies, the researchers designed a training program to which they exposed their interviewers in order to equip them with the skills they required to implement the SBIs effectively. However, in studies two and four the researchers exposed structured behavioural interviewers to one-day training programs while in study three, interviewers received a three-day training program. The inconsistency in the extent to which interviewers were trained in these investigations could have rendered the results that they yielded inconsistent and incomparable. This implies that the possibility exists that the findings obtained in these studies are not generalisable due to differences in the training programs to which interviewers were exposed. It appears that the training programs used in studies two and four were similar in content and duration. Therefore, it is likely that the findings yielded by these investigations could be statistically consistent. This stance implies that the author anticipated the possibility that the findings of studies two and four may have been statistically significantly different from the results yielded by study three since this investigation made use of a training program that differed in content and duration from the afore-mentioned studies.

In the fourth validation effort, Motowidlo et al. (1992) made use of two independent criterion measures as their dependent variables. One of these was a behaviourally based 7-point rating scale which was used to assess interview performance in relation to pre-determined dimensions that were regarded as crucial for effective performance on the job. The second criterion measure consisted of a set of activity statements, which were rated in terms of a 5-point graphic scale. It is the author's contention that the validity estimate that was yielded using this second criterion measure could have differed significantly from those yielded in the second and third studies (in which the researchers used behaviourally anchored rating scales to obtain criterion scores) due to the difference in the type of rating scales used in the evaluation of subjects on the criterion measure.

#### (ii) Setting (context)-by Intervention Interactions

In the same way that Janz (1982) and Orpen's (1985) investigations were conducted in contexts that differed significantly from each other, the context in which studies two and four, by Motowidlo et al. (1992), were conducted differed dramatically from the context in which study three was conducted. The reason for this discrepancy is that the former investigations were conducted primarily for research

purposes while the latter study was conducted in an actual employment situation. The variability in the research contexts could also account for why the training programs, to which the interviewers in these investigations were exposed, differed in content and length. In the study that was conducted in an actual employment situation, the researchers may have deemed it more important, given the larger bottom-line implications associated with making effective hiring decisions, to train the actual interviewers more intensively than the interviewers who participated, in the second and fourth studies, merely for the purpose of research. These differences in the research contexts could have impacted on the validity estimates yielded and could have resulted in a lack of generalisability between the results obtained in studies two and four and those obtained in study three.

### (iii) Selection-by-Intervention Interactions

The use of actual job applicants in study three and the use of employees who already occupied the jobs for which the interviews they participated in were designed in studies two and four could have impacted on the extent to which the findings in the first study were consistent with the findings yielded by the latter two studies. As a result the external validity of these studies could have been called into question. A further aspect that could have impacted on external validity is the type of individuals that comprised the subject populations in each of these studies. In the second validation study the subjects consisted of management incumbents who were drawn from a range of functional areas within the participating organisations. In the third investigation, the subjects consisted of applicants for entry-level management positions across a range of functional areas and in the fourth validation effort the researchers drew on a group of marketing incumbents from a range of functional areas to serve as their subjects. The use of individuals from different occupational spheres could have impeded the generalisability of the findings yielded in these studies.

Notwithstanding the potential impact of all of the above-mentioned factors to detract from the generalisability of the validity findings yielded by the studies on SBIs, the meta-analytic technique for engaging in a pair-wise comparison of the findings suggested that the results yielded by these investigations were consistent with each other. Further, the meta-analytic technique that Rosenthal (1991) proposed to compare the findings of three or more independent studies also suggested that the findings yielded by all the investigations on SBIs are statistically consistent. This implies that the studies are comparable and that they are likely to generalise to different situations or settings and to different groups of people if they possess internal validity. This finding may be attributed to the various similarities that these investigations possessed in terms of the procedures, processes and techniques they employed.

### 5.3 Research on Situational Interviews

The largest number of studies located in this investigation yielded validity estimates for situational interviews (SIs). In addition to ascertaining criterion-related validity estimates for these interviews, the majority of these studies were also oriented towards assessing inter-rater and internal consistency reliability estimates. The three studies conducted by Latham et al. (1980), the two investigations conducted by Latham and Saari (1984) and the investigation conducted by Weekley and Gier (1987) yielded both inter-rater and internal consistency reliability results. In one of the investigations undertaken by Latham and Saari (1984) a test-retest reliability estimate was also reported. The study reported by Robertson et al. (1990) yielded only an internal consistency reliability estimate while the study conducted by Stohr-Gilmore et al. (1990) yielded only an inter-rater reliability estimate. Thus, these studies were all strongly oriented towards the estimation of reliability as well as validity.

#### 5.3.1 An Exploration of the Threats to Internal Validity

In this section the threats to internal validity, to which the studies on situational interviews were exposed, will be explored.

##### 5.3.1.1 Research Design Strategies

There were eight studies that yielded validity estimates for SIs. Of these, the three studies reported by Latham et al. (1980), the two investigations conducted by Latham and Saari and the investigations by Weekley and Gier (1987) and Robertson et al. (1990) employed a correlational research design strategy. The study that was conducted by Stohr-Gilmore et al. (1990) employed a between-subjects, non-equivalent groups, research design strategy (Stern and Kalof, 1996).

##### 5.3.1.2 Threats to Internal Validity

The studies that conformed to the correlational research design strategy are prone to the same fundamental threats to internal validity that were discussed in relation to the research on SBIs which also employed this research design strategy. The ways in which these threats manifested in the correlational studies on the situational interview will be discussed. The threats that the study, which employed a between-subjects, non-equivalent groups, research design strategy, is prone to, are different. In order to discuss the threats to internal validity to which the studies on SIs were exposed, the author will structure the discussion on the basis of the research design strategies that were used in situational interview research.

### (i) Correlational Studies

In the studies classified as correlational, the researchers did not manipulate the independent variables (i.e. the interviews). They merely sought to ascertain the relationship between the interview scores and the criterion scores in an effort to compute the validities associated with situational interviews.

#### (a) Organismic Variables

The subjects in the first study by Latham et al. (1980) were all males. In the second study by these researchers the subjects were all males and all White. In the third study reported by Latham et al. (1980) the subjects were all Black. In the first study reported by Latham and Saari (1984) the subjects were all females. In the study by Weekley and Gier (1987) the subjects constituted individuals who had worked previously in the fine jewellery department of a major department store and who applied for the position of sales associate. All these samples consisted of homogeneous groups of individuals. Their homogeneity was defined in terms of the subject variables or organismic variables that they introduced into their respective studies. These included race, gender and previous experience in a common environment.

In the studies, which used samples that comprised solely of males or females, gender could have served as an alternative explanation for the findings obtained. In the studies, which used homogeneous groups of Black or White individuals in the samples, race could have served an alternative explanation for the findings that were yielded. In the study in which the subjects were all previously employed in a major department store, it is possible that the experience these individuals had gained in their previous jobs could have constituted an extraneous variable, which impacted on the results that were obtained. In the study that Robertson et al. (1990) conducted, it is possible that the similarity of the jobs that the internal candidates occupied and the fact that they held these jobs in the same organisation, could have served as alternative explanations for the findings obtained, on the basis of the organismic variables that the subjects introduced into the study.

In the author's opinion, the alternative explanations that may be attributed to the organismic variables of gender and race could have been controlled in these investigations if the researchers relied on the technique of random sampling to obtain their subjects. The use of this technique would have provided individuals from the target population with an equal chance of being selected into the sample and in this way the effects of these organismic variables would have been randomised. However, in the first study that Latham et al. (1980) conducted, the researchers claimed to have used the technique of random sampling to obtain their sample of 49 unionised hourly sawmill workers from a population of

207 individuals. Despite the use of this technique, all the individuals included in the sample were males. This situation could have arisen if all the individuals in the target population were male. Therefore, there may have been only male subjects to choose from. In this specific instance the use of the random sampling technique did not eliminate the potential impact of the organismic variable, in question, on the findings yielded by the investigation. In situations where the entire target population is homogeneous, the attainment of a sample of homogeneous individuals is inevitable.

A second way in which the researchers could have controlled for the possible threats posed by the organismic variables of gender and race is by including these potential extraneous variables as independent variables in their hypothesis in order to assess their impact on the dependent variable in question (Stern and Kalof, 1996). In this way the researchers could have compared the validity estimates yielded for male subjects with those yielded for female subjects. Such a comparison would have ascertained whether or not the organismic variable of gender impacted on the validity findings associated with the use of situational interviews. The organismic variable of race could have been handled in a similar manner.

In the second concurrent validation study that Latham et al. (1980) conducted, the researchers reported the possibility for the interview evaluations to be affected by external factors. These factors included, inter alia, additional job knowledge that subjects may have possessed, differences in the subjects' motivational levels and the added maturity of subjects due to their status as employees rather than job applicants. In the author's opinion, these factors could equally have influenced the job performance ratings that the subjects received. The impact of the external factors that the subjects introduced into the research setting could have impacted on the validity results that were obtained. Therefore, they could have constituted sources of alternative explanations for the research findings. The researchers (Latham et al., 1980) proposed that these organismic variables could be overcome through the use of a predictive validation strategy as opposed to a concurrent validation strategy. This recommendation is based on the premise that concurrent designs utilise individuals that are already employed within organisations and have experience in the jobs for which the interviews are designed. In predictive designs, however, researchers generally use actual job applicants who have not had experience in the jobs for which they are interviewed. Therefore, the use of predictive rather than concurrent validation strategies could eliminate the organismic variable of previous experience in the actual job being interviewed for. On the basis of this insight, the results yielded by the predictive validation studies conducted by Latham et al. (1980) [Study 3], Latham and Saari (1984) [Studies 1 & 2], Weekley and Gier (1987), Robertson et al. (1990) and Stohr-Gilmore et al. (1990), were not affected by the organismic variable of previous experience in the job being selected for. However, other organismic variables could have served as alternative explanations for the findings yielded in these investigations.

In the concurrent validation studies that Motowidlo et al. (1992) reported, the researchers proposed that the effects of previous experience may have been minimised by including individuals with little experience in the sample and by asking them to draw on the experiences that they encountered before they occupied their current positions. These methods could also have been employed by the situational interview researchers to control for the extraneous organismic variable of previous experience.

#### (b) Invalid Operational Definitions

With the exception of memory inaccuracies, the factors that could confound the operationalisation of the situational interview are the same as those discussed for PBDIs and SBIs. The reason that memory inaccuracies on the part of the subjects is no longer a possible extraneous variable is due to the emphasis of the questions in situational interviews on future intentions rather than past behaviours and experiences. In the discussion on SBIs it was noted that the threat introduced by invalid operational definitions into correlational research is second in significance to the threat of organismic variables (Stern and Kalof, 1996). Therefore, these threats are important and may be addressed using the recommendations proposed in the discussion of the research on PBDIs and SBIs (with the exception, of course, of the recommendations for overcoming the threat of memory inaccuracies).

Discussions of invalid operational definitions were offered in relation to the research on PBDIs and SBIs. In these discussions, it was noted that one way in which the confounding of the independent and dependent variables could be minimised, is if interviewers and performance evaluators received standardised training programs that equipped them with the knowledge and skills that are necessary to evaluate subjects' performance in the interview and their performance on the job, consistently and objectively. In the correlational studies on the situational interview, Latham et al. (1980) reported exposing the individuals who rated subjects on the criterion measure to standardised training programs. In the author's opinion, this could have reduced the possibility for the definition of the dependent variable, which was investigated in these studies, to have been rendered invalid since the training is likely to have assisted performance evaluators to evaluate job performance and not contaminate these evaluations with their own subjective preferences and opinions. However, the threat of invalid operational definitions in relation to the dependent variable, existed in all the correlational studies that did not expose performance evaluators to standardised training programs. Of the seven correlational studies identified in the category of situational interviews, only two studies (Robertson et al., 1990; Weekley and Gier, 1987) reported the use of explicit training programs for the situational interviewers. Based on the same argument proposed above, training programs could have minimised the potential for invalid operational definitions of the independent variables in these studies. However, in the studies, which did not report the use of interviewer training programs, this threat could have prevailed.



### (c) Sampling Biases

In the first study conducted by Latham et al. (1980), the researchers used the technique of random sampling in order to obtain their sample of 49 subjects from a population of 207 unionised hourly sawmill workers. According to Stern and Kalof (1996), when the technique of random sampling is used, the sample may be regarded as representative of the population from which it was drawn. This implies that the validation results obtained in the context of this study may be extended to the population of unionised hourly sawmill workers within the organisation in which the research was conducted. These researchers overcame the potential threat of sampling biases which could have influenced the internal validity of their study had the random sampling technique not been used.

In the third validation study conducted by Latham et al. (1980) and in the first study that Latham and Saari (1984) engaged in, the use of saturation surveys was noted, to obtain the subjects. This implies that all the individuals that applied for the positions in question were interviewed, hired and evaluated in terms of specific criterion measures in order to calculate the validity estimates associated with the situational interviews that were used in these investigations. The use of this technique to obtain the subjects that participated in the validation initiatives precludes the possibility of sampling biases due to the fact that the entire populations of individuals that could possibly have been included in these studies, were in fact included.

In the second study conducted by Latham and Saari (1984) and in the study conducted by Robertson et al. (1990) the researchers used the convenience sampling technique to obtain the subjects that participated in these validation efforts. In the first study, the sample consisted of 157 individuals who were hired from a group of 349 recruits while the subjects in the second study comprised of 63 people for whom interview and criterion data were available for the purpose of validation. According to Stern and Kalof (1996), the use of unrepresentative samples in research efforts introduces the potential for sampling biases. The implication of these biases is that the results yielded by these research efforts are applicable only to the samples from which they are derived and may not be generalised to the larger populations from which these samples are drawn. In order to overcome the potential sampling biases that could have resulted through the use of the convenience sampling technique, Latham and Saari (1984) could have derived a random sample from the 157 individuals that were hired. Alternatively, they could have used the technique of purposive sampling to obtain a sample that was representative of the total population of hired employees in terms of characteristics such as gender, age, race, work experience, and other significant factors that could have impacted on the variables that were deemed important to the investigation. It is unlikely, that Robertson et al. (1990) could have used either of these sampling techniques effectively in the light of the small group of subjects that were initially

available to them. Therefore, the best that these researchers could have done was to acknowledge the potential threat that sampling biases could have posed to their findings even though these effects may be regarded as insignificant according to Stern and Kalof (1996).

The sample that Latham and Saari (1984) derived from a larger group of recruits could have also been prone to the difficulty of restriction of range, which is a phenomenon that occurs when the range of possible scores on one or both variables under investigation is limited (Heiman, 1998). Latham and Saari (1984) conducted their validation study with the group of individuals that received the highest interview scores on the basis of which they were hired. Therefore, it is possible that restriction of range could have distorted the validity findings that were obtained. Although these researchers did not attempt to control for this possible distortion, they could have done so by drawing on the statistical controls that exist for correcting research findings that are plagued by the problem of restriction of range.

The author was unable to ascertain the sampling technique that was used to obtain the sample of first-line foremen in Latham et al.'s (1980) second investigation. Therefore, it was not possible to comment on the potential biases that were introduced into this investigation by the sampling technique that the researchers employed. In general, the researchers who conducted the validation research on situational interviews used relatively small samples. As a result, it is possible that the validity estimates derived from these samples may have been distorted. In the realm of statistics, there are procedures that exist for correcting the biases introduced by small samples. Although some of the researchers (Robertson et al., 1990; Weekley and Gier, 1987) reported the use of measures to correct the validity estimates yielded in their investigations, they were not explicit about whether or not these controls included a correction for small sample size. This technique may be used in all validation research on structured employment interviews in order to eliminate the potential distortion that small sample sizes are likely to inflict on the validity estimates that are yielded. A more obvious way to control for small sample sizes is to use larger samples in validation efforts. It was noted earlier that the technique of power analysis may be used to ascertain acceptable sample sizes. However, in actual selection situations, it may not always be possible to use acceptably sized samples due to resource constraints. It is in these situations that statistical control is likely to be most beneficial.

#### (d) Uncontrolled Variation In Information

The threat of uncontrolled variation in information exists when research procedures allow for inconsistencies to arise in the information that is obtained for the purpose of research endeavours. Stern and Kalof (1996) proposed that one way in which to control this threat is to hold the research

procedures that are used constant. However, in reality, even concerted efforts to ensure that the same information is obtained from or about different people, will not necessarily eliminate the potential for the threat of uncontrolled variation in information to exist. In the correlational studies on the situational interview, the author identified the potential for the following factors to render the information that was collected from or about different individuals inconsistent. Consequently, these factors could have posed a threat to the internal validity of these investigations.

On the basis of the information that the researchers, who conducted the situational interview studies, provided it seems that more than one interviewer was used to conduct the situational interviews in each of the seven correlational studies and in the between-subjects experiment. Consequently, all these studies are likely to have been prone to the threat of uncontrolled variation in information. In the three studies conducted by Latham et al. (1980), however, the researchers used one set of individuals to conduct the situational interviews and another set of individuals to evaluate them. This practice could have served to exacerbate the already existing possibility for error and inconsistency in the interview evaluations and the validity findings that were ultimately derived from them.

If different interviewers are used to elicit information of the same nature from research subjects (interviewees) then the potential exists for this information to be inconsistent and incomparable. In order to guard against this possibility, some of the researchers who conducted investigations on situational interviews (Robertson et al., 1990; Weekley and Gier, 1987) reported that they subjected their interviewers to training in the use and evaluation of these interviews. This was done in an effort to strive for consistency, and consequently, comparability in the manner in which the interviews were conducted and evaluated so that the validation results based on these interviewer interventions would be rendered more accurate. However, it was noted in the discussion on uncontrolled variation in information, in the context of PBDIs and SBIs, that the standardised training exposure to which interviewers are exposed does not guarantee consistency in the information they obtain due to individual differences in the manner in which they apply the knowledge and skills with which they are equipped in the training programs. Despite the potential for the threat of uncontrolled variation in information to manifest even after precautionary measures have been employed, the use of standardised training efforts for interviewers is imperative since it does appear to reduce the possibility of this threat.

Latham and Saari (1984) conducted a validation study with actual job applicants who were subsequently hired into an organisation. In this investigation, the researchers did not comment explicitly on whether or not the interviewers that participated were exposed to a standardised training program in order to equip them with the skills they needed to fulfil their responsibilities. These

researchers discovered that the interviewers did not use the situational interview correctly. Instead of recording and scoring the answers to each question, the interviewers did not engage in recording at all. They used the questions and the scoring guides to render global impressions of the candidates and assigned summary scores to them. The researchers did not state that their interviewers were not trained. However, it seems that a lack of training may have contributed to this situation. In the author's opinion, global impressions imply the use of subjective interpretations and perceptions (i.e. interviewer discretion), which could have introduced inconsistency into the manner in which the interviewees' responses were evaluated. Such inconsistencies could, in turn, have impacted on the internal validity of the findings that were yielded by this investigation.

One way for researchers to minimise the impact of the threat of uncontrolled variation in information on the internal validity of research findings, is to monitor the manner in which the interviews are conducted and evaluated and the manner in which the criterion measure is used and evaluated. The author has provided some suggestions for monitoring in the discussions of uncontrolled variation in information in relation to PBDIs and SBIs. These are likely to be of tremendous assistance in the quest for consistency since researchers can exclude the information they deem to be inconsistent and in so doing, eliminate potential contributors to uncontrolled variation in information. However, these specific recommendations may not necessarily be generalisable to the monitoring of the use and evaluation of criterion measures.

However, it is possible for inconsistencies, in the manner in which criterion measures are used and evaluated, to be addressed through the training of performance evaluators. Latham et al. (1980) employed this practice in the three studies that they conducted. Their ultimate aim in doing so was to minimise the prevalence of rating errors in observing and evaluating others. Performance evaluators and interviewers can both introduce inconsistencies into research efforts as a function of their humanness. In the same way that standardised interviewer training programs go some way to reduce the impact of uncontrolled variation in information on the results that are yielded, standardised training programs for performance evaluators are likely to have the same effect.

In the context of SBIs the author commented on the potential for pilot studies on interviews to minimise the possibility for uncontrolled variation in information. The same arguments apply to the investigation by Weekley and Gier (1987). These researchers engaged in a pilot study on the situational interview they investigated. None of the other researchers reported using this practice in their validation studies on situational interviews. If they had, it is possible that some of the ambiguities, that could have influenced the responses offered to some of the situational questions used,

could have been overcome. Consequently, the potential for uncontrolled variation in information could have been minimised.

In interview research, in general, it is possible that interviewers and performance evaluators could introduce their individual preferences, biases, opinions and interpretations into the research process despite efforts to ensure consistency and standardisation. One way in which this threat could be controlled is to make use of panel interview designs and panel performance evaluator forums. In this manner, the individual and subjective issues that interviewers and performance evaluators present can be diffused and checked by other individuals who may have their own set of biases but who may also be able to offer alternate insights and perspectives into how subjects should be evaluated and why. The most important benefit of using groups of evaluators to evaluate individual subjects is that each evaluator can question and challenge others to offer sound and logical justifications for their evaluations. In the studies by Latham et al. (1980) and in the second study by Latham and Saari (1984) the researchers made use of panel interviews. Therefore, in these studies, the possibility for uncontrolled variation in information could have been reduced. The studies conducted by Robertson et al. (1990) and Weekley and Gier (1987) did not adopt a panel interview design. In these investigations the interviews were conducted on a one-to-one basis. This implies that these studies could have been more prone to the threat of uncontrolled variation in information. In the first study reported by Latham and Saari (1984), it is not clear whether a one-on-one or a panel interview design was used. It may be prudent for employment interview researchers to adopt the use of panel interview designs since this can serve as one mechanism to guard against the possibility for uncontrolled variation in information. This recommendation is pertinent not only in the context of research on situational interviews but is also applicable in relation to the research on PBDIs and SBIs.

#### (ii) Between-Subjects Experiment

In their investigation Stohr-Gilmore et al. (1990) set out to ascertain whether or not there was a positive association between the dependent variables, which were depicted by two training programs and two performance evaluations, and the independent variables, which assumed the form of two types of employment interviews (viz. one with situational questions and the other without situational questions). In order to investigate this relationship these researchers employed a between-subjects non-equivalent groups research design strategy. They exposed their first group of subjects to the type of employment interview that excluded situational questions, late in 1985. The second group of subjects was exposed to interviews that contained situational questions, early in 1987. Therefore, in this investigation, the type of interview to which researchers exposed their subjects was manipulated. Further, this manipulation occurred in the context of two independent groups of applicants who were not obtained

randomly. Hence, the research design strategy that was employed in this study is a between-subjects non-equivalent groups research design strategy as defined by Stern and Kalof (1996).

According to Stern and Kalof (1996), the most significant threats to which between-subjects, non-equivalent groups, research design strategies are exposed are less severe than the most significant threats inherent in within-subjects experiments and correlational research. These threats are time-tied variables, organismic variables and invalid operational definitions (Stern and Kalof, 1996). The threats of sampling biases and uncontrolled variation in information are also prevalent but less significant than the afore-mentioned threats, according to Stern and Kalof (1996). These writers proposed that the most significant threats in the realm of between-subjects experiments that use non-equivalent groups correspond in severity to the second most significant threats in the context of within-subjects experiments and correlational research (viz. invalid operational definitions) (Stern and Kalof, 1996). In the following discussion, the threats to internal validity that Stern and Kalof (1996) identified for research that conforms to this particular research design strategy will be outlined. This discussion will emphasise how these threats manifested in the investigation by Stohr-Gilmore et al. (1990) and will offer recommendations on how these researchers could have counteracted them.

#### (a) Organismic Variables

In the investigation by Stohr-Gilmore et al. (1990) the subjects that were drawn on for the purpose of validation consisted of two independent groups of applicants selected for the position of correctional officer. The first group of individuals was exposed to an interviewing condition that did not include the use of situational questions and was assessed late in 1985. The second group of individuals consisted of applicants for the same position who were hired with the aid of an interview that included situational questions. This second group of subjects was assessed early in 1987. According to Stern and Kalof (1996) the most significant threats that between-subjects experiments, that use non-equivalent groups, are exposed to are organismic variables. In this study the use of two independent groups of individuals implied the possibility of this threat, since each group may have introduced a unique set of organismic variables, which may have systematically influenced the results that were obtained. However, the researchers did not offer sufficiently detailed information for the author to attempt to identify the presence of such variables and to postulate about the potential impact they could have had on the validity results.

Stern and Kalof (1996) suggested that the impact of organismic variables may be minimised with the use of an experimental design, which advocates the random assignment of subjects to each treatment condition. In this study, an experimental design was employed but the random assignment of subjects

to each condition was not an option. This was due to the fact that each condition constituted a part of an actual selection program, which was used in the selection of applicants at two distinct points in time. A second technique that these writers proposed is matching (Stern and Kalof, 1996). This technique requires subjects to be matched on all the important factors that could influence the effects of the independent variables so that the results obtained, in the research efforts in which this technique is employed, may be attributed primarily to the effects of the independent variables in question. In this study, the technique of matching could have been employed to control for the possible threat of organismic variables. The researchers could have identified those subject variables that were most likely to have confounded the effects of the independent variable and could have matched subjects on these variables in order to eliminate them as sources of alternative explanations (Stern & Kalof, 1996). A further strategy that Stern and Kalof (1996) proposed is to include the subject variable that is likely to confound the effects of the independent variable in the hypothesis. However, the researchers did not use any of these techniques to control for the possibility of organismic variables in their investigation. Consequently, the results they obtained may have been distorted.

#### (b) Time-Tied Extraneous Variables

The fact that each group of subjects was exposed to a treatment condition (i.e. either the interview with situational questions or the one without situational questions) at a separate point in time implies the potential for the threat of time-tied extraneous variables to impact on the findings that were attained. However, these threats could only have influenced the results of this investigation at the point that the validity estimates yielded by each type of interview were compared. It did not have an effect on the validity estimate that was yielded for the structured interview (viz. the situational interview) that was used in the study. Therefore, no further consideration will be offered to this threat, since it did not have an impact on the specific findings with which the present investigation is concerned.

#### (c) Invalid Operational Definitions

The same arguments that were made in the context of the correlational studies on situational interviews are applicable to this particular study in relation to this threat to internal validity.

#### (d) Sampling Biases

The samples of subjects that were used in each interviewing condition consisted of conveniently available applicants who applied for the job of correctional officer in 1985, 1986 and 1987. The conscious choice that Stohr-Gilmore et al. (1990) made to include only those individuals who had been

hired necessarily implied an unrepresentative sample that was plagued with the difficulty of restriction of range. The relatively small sample sizes that were used to obtain the validity estimates in question could have led to further attenuation of the results, thereby impacting detrimentally on the internal validity of the study. The most effective manner to overcome the potential for sampling biases, according to Stern and Kalof (1996), is to use the technique of random sampling. However, especially in relation to research that is conducted in actual selection situations, it is difficult if not impossible to employ this sampling technique. The samples of subjects that are used in these contexts are, more often than not, obtained using the technique of convenience sampling (i.e. the individuals who are hired are included in the samples). There are also statistical controls available for the threats posed by the use of small samples and restriction of range. These difficulties are common in validation research on employment interviews.

#### (e) Uncontrolled Variation in Information

The interviewers were exposed to a training program in this investigation. This implies that the potential for uncontrolled variation in information could have been significantly reduced even though the possibility for this threat to materialise still existed. The existence of such a threat is informed by the variability that could have occurred as a function of the individual differences between the interviewers, in terms of how they implemented the skills and knowledge with which they were equipped during these programs.

Although the performance evaluators were not trained in the use and evaluation of the criterion measures, if they had been, the potential for uncontrolled variation in information could have been significantly reduced. For the same reasons as those that explain why the threat of uncontrolled variation in information cannot be completely overcome by the training of interviewers, the training of performance evaluators is not likely to eliminate the threat of uncontrolled variation in information.

The researchers in this study envisaged the use of group or panel interviews for conducting the interviews with the second group of individuals (i.e. the situational interviews). For the same reasons as those offered in relation to the correlational studies on situational interviews, this practice could have minimised the effects of uncontrolled variation in information.

There are other ways in which these researchers could have controlled for the possible threat of uncontrolled variation in information in relation to the validity results that were yielded for the structured employment interview in question. One such way is if researchers held the interviewer who



conducted the situational interviews with the 33 interviewees constant. Another way is if they engaged in a pilot study to eliminate potential ambiguities in the questions included in the interview schedule.

### 5.3.2 An Exploration of the Threats to External Validity

In the discussion of the threats to internal validity, the studies on the situational interview, that were located for the present investigation, were divided into those that employed a correlational research design and those that used a between-subjects non-equivalent groups experimental research design. The main reason for drawing this distinction was to explore the threats to internal validity that were inherent in these investigations of the two primary research design strategies that they employed. In the discussion of the threats to external validity, that the research on situational interviews are likely to be prone to, these divisions will not be adhered to.

In terms of the experimental arrangements that existed in these investigations, the author believes that the following factors could have detracted from the generalisability of the findings that were yielded, to other settings, populations and time periods:

#### (i) Training Programs for Interviewers

In the discussion on uncontrolled variation in information, for the set of studies on situational interviews that employed a correlational research design strategy, it was noted that not all the researchers were explicit about whether or not they exposed their interviewers to training programs. In the investigations by Latham et al. (1980) and Latham and Saari (1984), in which the researchers were not explicit about whether or not the interviewers were trained, the possibility exists that they were not. In the remaining studies by Weekley and Gier (1987), Robertson et al. (1990) and Stohr-Gilmore et al. (1990), the researchers did state that their interviewers were subjected to a standardised training program. Although it is not clear whether or not the interviewers were, in fact, trained when the researchers were not explicit on this issue, the author believes that if they were not trained then the findings yielded by these studies could differ significantly from those yielded in the studies in which the interviewers were trained. In other words, it is possible that the exposure of interviewers to training programs, or the lack thereof, could have impacted on generalisability.

In addition, in the studies in which the interviewers were trained, the researchers, in each investigation, used unique training programs, which differed from each other in terms of essential design features such as content and duration. This could also have detracted from the comparability of the findings yielded by these studies (Robertson et al., 1990; Stohr-Gilmore et al., 1990; Weekley and Gier, 1987).

## (ii) Training Programs for Performance Evaluators

In the studies conducted by Latham et al. (1980), the researchers noted that the performance evaluators were exposed to training programs. However, in the remaining studies on the situational interview this practice was not employed. As a result, it is possible that the findings yielded by the former investigations could have differed significantly from those that were yielded in studies that did not engage in the practice of training performance evaluators.

## (iii) Panel Interview Designs Versus One-to-One Interview Designs

In the three studies conducted by Latham et al. (1980), the second study by Latham and Saari (1984) and the study by Stohr-Gilmore et al. (1990), the researchers employed panel interview designs. Robertson et al. (1990) and Weekley and Gier (1987) made use of one-to-one interviews. In their first study, Latham and Saari (1984) were not explicit about whether they employed a panel or one-to-one interviewing strategy. The possibility exists for the findings obtained in the investigations that employed a panel interview design, to be significantly different from the findings yielded by investigations that used a one-to-one interview design.

It was noted that some of the studies on situational interviews were conducted in actual employment situations while others were conducted purely for the purpose of research. In the first two studies by Latham et al. (1980) and in the first study by Latham and Saari (1984), the researchers drew on existing employees (i.e. unionised hourly saw-mill workers, first-line foremen and office clerical personnel) to assess the concurrent validity associated with the situational interviews in question. These studies appear to have been conducted purely for research purposes. On the other hand, the third study by Latham et al. (1980), the second investigation by Latham and Saari (1984), and the investigations by Weekley and Gier (1987), Robertson et al. (1990) and Stohr-Gilmore et al. (1990) made use of actual job applicants to ascertain the validity estimates associated with the situational interviews in question. These investigations employed predictive validation strategies. Since some of these investigations were conducted in actual employment situations, while others appear to have been conducted solely for the sake of advancing validity research and knowledge on situational interviews, there is a possibility that the differences in the contexts in which the investigations were conducted, could limit or impede the generalisability of the findings that they yielded.

In the discussions of the threats to external validity in relation to PBDIs and SBIs, the author commented on Dooley's (1995) contention that aspects of the samples used in research investigations,

could interact with the independent variable, in question, to influence the comparability of studies that are oriented towards the examination of comparable research questions using samples that may have been drawn from significantly different populations. In the research on situational interviews, it was noted that the researchers used vastly different groups of people in their samples. Latham et al. (1980) used unionised hourly sawmill workers in Study 1, first-line foremen in Study 2 and applicants for entry-level work at a pulp-mill in Study 3. Latham and Saari (1984) drew on office clerical personnel in a regional office of a wood products company in Study 1 and applicants for entry-level utility work in a newsprint mill in Study 2. Weekley and Gier (1987) made use of applicants for the position of sales associate in their investigation. Robertson et al. (1990) drew on internal candidates in clerical and administrative jobs who were considered for selection and promotion within an organisation. Stohr-Gilmore et al. (1990) made use of the entire population of individuals who were selected for the position of correctional officer. Therefore, comparability of the results yielded by these investigations may be difficult. This difficulty may be further exacerbated by the use of applicants versus existing employees in these investigations.

Despite the possible impact of these threats on the external validity of the validation research on situational interviews, a cumulative comparison of the validity estimates, which were based on the criterion of supervisors' ratings using behaviourally-defined rating scales and which excluded corrected validity estimates, revealed that the results for all the studies were consistent with each other. This implies that the validity estimates that were yielded may be generalised to other settings, populations and time periods provided that these studies possess internal validity. However, an interesting phenomenon was observed in the pair-wise comparisons of these findings. In general, the results, obtained using this meta-analytic technique, revealed that the findings were consistent with each other irrespective of the potentially significant differences that were noted between the studies. However, in the second study conducted by Latham and Saari (1984) distinctively higher Z-scores were noted. Typically, higher Z-scores are associated with significant differences between research results. In the pair-wise comparisons, significant differences were discovered at the 5% level between the finding yielded by this investigation ( $r=0.14$ ) and the findings yielded in the first study by Latham et al. (1980) (for the estimate of 0.46) and the corrected validity estimates of 0.38 and 0.43 that were attained in the study by Robertson et al. (1990). The comparison between Latham and Saari's (1984) validity estimate of 0.14 and Latham et al.'s (1980) estimate of 0.50 yielded a statistically significant difference at the 1% level. This finding implies that there was a 1% possibility that this finding could have occurred by chance.

Latham and Saari (1984) noted that a shortcoming of this particular investigation was that the interviewers did not use the situational interview correctly. They did not record the interviewees' responses to each question or assign scores to each response (Latham and Saari, 1984). Instead, they used the questions and the scoring guide to assist them in formulating global impressions of the interviewees (Latham and Saari, 1984). The interviewers assigned summary scores to them on the basis of these overall impressions (Latham and Saari, 1984). The incorrect use of the situational interview in this investigation could have contributed to the statistically significant differences in the findings. The observation of these statistically significant differences between the afore-mentioned studies suggests that they are not comparable and that the findings yielded by Latham and Saari (1984), in their second study, do not generalise to other populations, settings and time periods.

The author contemplated the possibility that the corrected validity estimates calculated in Studies 11 and 12, which were reflected in estimates 11b, 12c and 12d in Table 4.7 in the Results chapter, could have been significantly different from the uncorrected validity estimates that were yielded in the other studies on situational interviews. However, the pair-wise comparisons of the results demonstrated that the corrected validity estimates did not differ significantly from the uncorrected estimates. This implies that the corrected and uncorrected findings are generalisable provided that the studies in question are shown to be in possession of internal validity.

#### 5.4 Conclusion

This chapter explored the potential threats to internal and external validity to which the studies on structured employment interviews, analysed in the present investigation, were prone. This exploration occurred in the context of three distinct categories of structured employment interviews (viz. patterned behaviour description interviews, structured behavioural interviews and situational interviews).

It was noted that the researchers who embarked on investigations into the PBDI took cognisance of the most significant threat to internal validity that the research design strategy (i.e. within-subjects experiment) they employed, was most prone to and employed appropriate measures (i.e. counterbalancing) to control for it. In the concurrent validation studies that Motowidlo et al. (1992) embarked on in relation to SBIs, the use of the technique of matching and a concerted effort on the part of interviewers to discourage subjects from drawing on experiences that they were exposed to in their current positions, was noted. The rationale for this was to minimise the effects of experience in the jobs being interviewed for on the validity findings yielded. Unfortunately, in relation to Motowidlo et al.'s (1992) predictive validation study, the correlational investigations that were conducted on the

situational interview and the between-subjects experiment on the situational interview, there does not appear to have been a concerted effort, on the part of researchers, to exert controls to counteract the most significant threats to which their research design strategies were prone.

In relation to some of the less significant threats to which the studies, analysed in the present investigation, were exposed (viz. invalid operational definitions, sampling biases and uncontrolled variation in information), the author did not identify the use of any overt control measures on the part of researchers. However, as a function of the investigation of structured employment interviewing techniques, the potential impact of some of these threats was inadvertently minimised. For example the use of standardised training programs for interviewers and criterion evaluators could serve to reduce the impact of invalid operational definitions and uncontrolled variation in information. The latter threat also appears to have been reduced by the use of panel interview designs and pilot studies for finalising structured employment interview schedules.

In the exploration of the threats to internal validity, the author commented on whether or not the researchers employed appropriate controls. In the instances in which they did, the techniques used to minimise the impact of specific threats to internal validity were noted. In cases where the researchers did not employ active measures to counteract the effects of the threats to which their research could have been prone, recommendations were offered on how such threats could be controlled and how these controls could be implemented in specific research endeavours. In instances where the use of controls that inadvertently could have combatted some of the threats to internal validity was perceived, these were noted and the possible implications of their use, was explored.

In relation to the threat of sampling biases two key issues were considered. The first factor that was considered was how the sample was obtained and the second factor that was considered was the size of the sample. On the basis of these crucial pieces of information and with the insights obtained on sampling in general research methodology literature, the author commented on how sampling biases could have been minimised in the research that was analysed in the present investigation.

Thus, the discussions of threats to internal validity, in relation to the three categories of structured employment interviews mentioned above, yielded insights into what the researchers did and did not do to control for specific threats and recommendations, pertaining to how specific threats could have been controlled if researchers did not employ appropriate control techniques. In some instances, the practices that researchers engaged in during their investigations alerted the author to potential ways in which specific threats to internal validity could be controlled. In these instances, the recommendations were extended to other contexts in which they were perceived to be applicable and effective.

In the discussion of external validity, all the possible threats that could have impacted on the generalisability of the findings obtained in the studies within each structured employment interview category, were contemplated. Thereafter, the results that were yielded by the meta-analytic techniques devised by Rosenthal (1991) were used for comparing the findings of independent research efforts in order to comment on whether or not they were comparable. In this manner qualitative explorations of external validity were backed up with quantitative data regarding whether or not the results were comparable. In relation to the investigations on PBDIs, the author concluded that the validity findings yielded in the studies by Janz (1982) and Orpen (1985) could be generalised to the contexts and populations used in each of these investigations. Similarly, the findings obtained in the multiple validation studies on SBIs, conducted by Motowidlo et al. (1992), could, be generalised to the unique settings that prevailed and the discrete populations that were used in these research initiatives. In relation to the external validity of the research on the situational interview, it was found that the findings obtained in the studies by Latham et al. (1980), Latham and Saari [Study 1], Weekley and Gier (1987), Robertson et al. (1990) and Stohr-Gilmore et al. (1990) could be generalised to the different populations and settings used to conduct these investigations. However, the findings of the second study that Latham and Saari (1984) conducted tended to differ significantly from those obtained in some of these investigations. Therefore, the author concluded that the generalisability of the findings obtained in this study was limited.

## Chapter 6

### Conclusions and Recommendations

#### 6.0 Introduction

In this study, the author explored the threats to internal and external validity that have plagued international research on structured employment interviews, which falls into two primary categories (viz. behaviour description interviews and situational interviews). In the former category, these threats were explored independently in relation to two sub-classes of behaviour description interviews that were identified (viz. patterned behaviour description interviews and structured behavioural interviews).

The author observed that researchers who engaged in the validation of structured employment interviews, generally, drew on two research design strategies (viz. correlational research design strategies and experimental research design strategies). The former were used in instances where the researchers were primarily concerned with determining the validity estimates associated with specific variants of structured employment interviews. Experimental research designs were typically employed when researchers were concerned with comparing the validity estimates yielded by different types of employment interviews.

In relation to the studies included for analysis in this investigation, the use of within-subjects and between-subjects experimental designs was noted. The former research design strategy was observed in the investigations by Janz (1982) and Orpen (1985). These researchers ascertained the validity estimates associated with patterned behaviour description interviews and traditional unstructured interviews in each study by using a common set of subjects to derive these estimates. Once these estimates were yielded for each separate variant of employment interview, they were compared. The use of a between-subjects experimental design was noted in the research on situational interviews, in a study by Stohr-Gilmore et al. (1990). These researchers compared the validity estimates associated with a selection process that included a situational interview and one that did not. In order to engage in this comparative analysis the researchers drew on two independent groups of interviewees who served as their subjects. Hence, the classification of this investigation as a between-subjects experiment.

## 6.1 Recommendations

On the basis of the discussion of the major threats to internal and external validity, the following recommendations were derived for South African researchers. Whenever possible and appropriate, these recommendations were tailored for specific realities in the South African context. In addition, the author offered insights into the practicalities that were likely to influence specific recommendations in actual organisational contexts as opposed to simulated selection contexts, which are typically devised and used purely for research purposes. Essentially these recommendations have been clustered into three broad categories. In the first category, commentary will be offered on the use of suitable research design strategies. The second category of recommendations pertains to the threats to internal validity that have characterised international research on structured employment interviews. In the third category of recommendations, insights will be offered on the likely implications of threats to external validity in the South African context and how they may be effectively managed.

### 6.1.1 Research Design Strategies

In the author's opinion, validation research on structured employment interviews may employ either correlational or experimental research design strategies depending on the primary aims of the researchers. If researchers are merely concerned with determining the validity estimates associated with a specific type of employment interview, then a correlational research design is appropriate. If, however, researchers are interested in comparing the validity estimates associated with different types of employment interviews in order to, for example, comment on which type may be more effective at predicting future job performance, then an experimental research design is likely to be more appropriate.

### 6.1.2 Threats to Internal Validity

In this discussion, the likely manifestations of the threats to internal validity in validation research in the South African context will be explored and suggestions will be made on how researchers could control for these threats. The threats that will be explored in this section are: organismic variables, time-tied extraneous variables, invalid operational definitions, sampling biases and uncontrolled variation in information.

According to Stern and Kalof (1996), depending on the research design strategies that researchers employ in their validation efforts, their investigations are likely to be prone to different threats to internal validity. The most significant threat that plagues the internal validity of correlational research



manifests as organismic variables (Stern and Kalof, 1996). In within-subjects experimental research designs, the most significant threat to internal validity takes the form of time-tied extraneous variables (Stern and Kalof, 1996). Between-subjects experiments that use non-equivalent groups of subjects are plagued by both organismic and time-tied extraneous variables (Stern and Kalof, 1996). However, these threats are less severe in between-subjects experiments. In terms of their severity, they are equivalent to the threat of invalid operational definitions in each of these research design strategies, according to Stern and Kalof (1996).

Every facet of South African society has been maimed by the separatist policy of apartheid. This policy created blatant inequalities among the different racial groupings that characterise South African society. Although the policy of apartheid is no longer actively implemented in the South African context, its effects on the previously disadvantaged groups are still evident today, despite the efforts that have been made, and that continue to be made, to rectify the injustices of the past. The author believes that it is likely to take several decades before the effects of apartheid are eliminated. In the meanwhile, the potential adverse effects that this separatist policy has had on all South Africans, in every facet of life, should be acknowledged and constructive measures should be employed to combat them.

#### 6.1.2.1 Organismic Variables

As a result of the gross inequalities to which entire groups of individuals have been exposed in the South African context, organismic variables are likely to pose a considerable challenge to employment interview researchers who embark on validation studies using correlational research design strategies. Some of the factors that could serve as sources of alternative explanations for the findings obtained in validation efforts on structured employment interviews in South Africa include race, unequal access to educational opportunities and unequal access to job opportunities. South African researchers should employ active measures to control for these potential extraneous variables in their validation efforts. The author believes that the following interventions are likely to be of assistance in this regard.

##### (i) Random Selection of Subjects

If subjects are selected randomly for correlational investigations, then the subject or organismic variables that they introduce into the research are likely to be randomised. In the author's opinion, random selection is one way in which researchers can attempt to avoid a situation in which homogeneous groups of subjects are used. This practice is likely to prevent a situation in which subjects have the potential to introduce organismic variables that may confound the results that are

obtained in systematic ways. The random selection of subjects is likely to be possible in validation efforts that are conducted purely for the purpose of research. The author believes that in this context, researchers are likely to have more control over how they obtain their samples of subjects.

However, when validation research is conducted in actual employment contexts, the random selection of subjects is less realistic. In actual selection situations, subjects are usually included in validation efforts on the basis of their interest in applying for the job in question. In other words, the technique of convenience sampling is used to obtain the subjects. One way in which a sample of randomly selected subjects can be obtained in an actual selection situation is if a sufficiently large group of subjects is interviewed and selected regardless of their performance in the interview and then a random sample is drawn for the purpose of validation.

This strategy for controlling for the threat of organismic variables in correlational research is likely to be unrealistic in organisations for several reasons. Firstly, the use of structured employment interviews for the purpose of selection is extremely resource-intensive in terms of time and human resources. In financial terms, the use of this tool is likely to have a significant impact on the bottom-line. Organisations are charged with the task of weighing whether or not the cost of their selection processes will be offset by the benefits that they are likely to yield. Usually if the benefits outweigh the costs, then organisations opt to implement such resource-intensive selection strategies. The use of structured employment interviews with large groups of individuals is likely to be an extremely costly venture. It is probably very unlikely that selectors in organisations will be given the mandate to use such a resource-intensive tool and then not base their selection decisions on the insights it yields if organisations do not have a vested interest in validating structured employment interviews.

Secondly, it is unlikely that organisations will allow their selectors to select all the individuals that applied for a particular position, regardless of their performance in the selection process, in order to validate a selection tool. Ineffective selection decisions incur tremendous costs. Therefore, the notion of selecting individuals even if they are not suited for the job in question just so that the selection tool to which they were exposed may be validated may not appear to be reasonable to organisations that are geared towards making profits. In actual selection situations the random selection of subjects may not be feasible. In such situations the inclusion of extraneous variables as independent variables in the hypothesis may be a more viable option.

## (ii) Inclusion of Extraneous Variables as Independent Variables in the Hypothesis

In order to assess the impact of the extraneous factors (i.e. organismic variables) introduced by subjects into research undertakings, on the findings yielded by these endeavours, Stern and Kalof (1996) proposed that such variables could be included in the research hypothesis as independent variables. In the South African context race is likely to feature as one of the most common organismic variables that could serve to confound the validity findings that are attained in correlational research. In order for researchers to control for and minimise the impact of this organismic variable, it could be included as an independent variable in validation efforts. In the South African context, it is likely that selectors will be confronted with the situation in which individuals from various different racial groupings will apply for a specific job. In order to comment on whether the correlations, between the scores obtained by these individuals in the structured employment interview and in the job performance measure, are influenced by their race, researchers can divide the subjects according to their racial groupings and conduct separate correlations for them. Thereafter, these correlations can be mathematically or statistically compared in order to determine whether or not they differed significantly from each other. If all other possible extraneous factors are held constant, then any differences observed may be attributed to the race of the subjects. In the South African context, race is intertwined with a myriad other factors and all of these have to be viewed holistically in order to interpret and understand the findings.

In a similar manner, other extraneous organismic variables, that could serve to confound the findings of validation research on structured employment interviews, may be controlled. In order to implement this control measure, however, it is imperative that researchers have a sufficient number of individuals in each of the groups that they intend to compare in order to ensure that the correlations that are attained in the context of each group are not a function of small sample sizes. In addition, they should strive to ensure that the number of individuals in each group is consistent. The author believes that this will assist in ensuring that the comparisons are a function of the extraneous variables in question and not the result of other factors such as small samples of subjects and discrepancies in the number of individuals in the groups being compared. If these conditions cannot be met in validation research, then researchers could use statistical mechanisms to control for small sample sizes and render the groups being compared equivalent in terms of their size.

## (iii) Matching

In the validation studies conducted by Motowidlo, et al. (1992), the researchers made use of the technique of matching to control for the organismic variable of previous experience in the job being

interviewed for. These studies were concurrent in nature. This implied the use of individuals who already occupied the positions, in question, as subjects in the studies. Motowidlo et al. (1992) made an active attempt to control for this extraneous organismic variable by including individuals with minimal experience in their current positions in their samples. In this manner, these researchers matched the individuals in the samples on the basis of the amount of time they had spent in their current positions. The author believes that the technique of matching is likely to be invaluable in concurrent validation efforts in order to control for the organismic variable of previous experience in the position being interviewed for. The use of this technique in the South African context is likely to be effective.

In general, correlational research that is oriented towards ascertaining the validity of structured employment interviews involves the use of one group of individuals. As a result the only matching that can occur is within this group of subjects in relation to potential factors that they could introduce into the study, which could influence the validity findings that are attained. However, in between-subjects experiments, it is possible to render the groups of subjects that are investigated more comparable using the technique of matching. By attempting to render the groups more comparable in this manner, the findings that are obtained are likely to be a function of the types of selection interviews being investigated as opposed to extraneous organismic variables.

#### (iv) The Use of Job Applicants Versus Job Incumbents

Latham et al. (1980) argued that the use of predictive strategies precluded the possibility that the scores yielded by the predictive and criterion measures in validation research, could be influenced by extraneous organismic factors such as job knowledge, different motivational levels or differences in the levels of maturity of the subjects. Their argument was based on the notion that actual job applicants do not introduce such factors, which could serve to confound the findings that are obtained, into validation research efforts (Latham et al., 1980). This argument is clearly premised on the assumption that all predictive validation studies make use of actual job applicants as opposed to individuals that are already employed in the positions being interviewed for. However, it was noted in the study by Janz (1982) that a predictive validation strategy was used with a group of teaching assistants who were already occupying the position that was interviewed for. In Janz's (1982) study, the criterion information was obtained after the interviews had been conducted. Therefore, it was predictive in nature. The author believes that the use of actual job applicants instead of existing employees is likely to minimise the impact of organismic variables. This opinion is based on the notion that when actual applicants are used, it is unlikely that they are obtained from a homogeneous group of individuals and consequently, their potential to introduce systematic organismic variables into research initiatives is minimised. Therefore, the use of predictive or concurrent validation strategies is not an issue in

relation to the threat of organismic variables. Instead the use of job applicants as opposed to the use of job incumbents is likely to be of consequence.

#### (iv) Other Techniques for Controlling the Threat of Organismic Variables in Validation Research

Motowidlo et al. (1992) made use of the following techniques to control for the threat of organismic variables in their concurrent validation studies. Firstly, they asked their subjects to pretend that they were applying for the positions in question even though these individuals were already occupying these positions. Secondly, they asked their subjects to draw on experiences that they had encountered in the jobs that they held before they commenced work in their current positions. In this manner, these researchers strived to control for the extraneous organismic variable of previous experience in the position being interviewed for (a common concern especially in concurrent validation studies). When confronted with this difficulty in their validation attempts, South African employment interview researchers could employ these controls to minimise the impact of this organismic variable.

There are essentially two main difficulties associated with the control mechanisms that were proposed, by Motowidlo et al. (1992), for minimising the threat of the organismic variable of previous experience in the position being interviewed for. Firstly, the interviewers or researchers who conduct the structured employment interviews may not inform the subjects that their responses need to conform to specific requirements (viz. pretending that they were actually applying for the job in question and drawing on experiences that they encountered before they occupied their current position). One way in which researchers could ensure that interviewers provide this information to their subjects is by engaging in video or audio monitoring procedures. However, the use of these monitoring procedures is fraught with difficulties. Subjects are often reluctant to grant interviewers and researchers consent to record their interviews in actual employment situations. In the absence of this consent, the use of monitoring procedures is not an option. This recommendation is likely to be constructive in validation research on structured employment interviews that is conducted purely for the purpose of research in a simulated employment situation.

The second difficulty associated with these control measures is the lack of certainty that subjects will take heed of interviewers' and researchers' instructions to pretend that they were actual job applicants and to draw on experiences that they encountered before they commenced work in their current positions. This difficulty is more challenging than the first. It may be overcome if researchers and interviewers have an insight into subjects' current and past employment records. In this way they could strive to verify that the subjects were adhering to the specific instructions that they were given in the concurrent validation studies. However, it is unlikely that this suggestion will enable researchers

and interviewers to effectively implement the control measures that Motowidlo et al. (1992) proposed for the minimising threat of the organismic variable of previous experience in the position being interviewed for.

Researchers who are interested in conducting validation research using between-subjects experimental research design strategies are likely to encounter similar difficulties with regard to the impact of organismic variables on their findings. The same recommendations that have been proposed to combat this threat to internal validity in correlational research, are applicable to between-subjects experiments.

#### 6.1.2.2 Time-Tied Extraneous Variables

Stern and Kalof (1996) argued that the threat of organismic variables is held constant in within-subjects experiments. However, a significant threat to internal validity, that researchers using this research design strategy are likely to encounter, is that of time-tied extraneous variables. Janz (1982) and Orpen (1985) made use of the control mechanism of counterbalancing in their studies which were designed to compare the validities associated with patterned behaviour description interviews and unstructured interviews. This strategy is likely to be effective when subjects are exposed to two different types of employment interviews, as was the case in these investigations. In such situations researchers can attempt to ensure that the timing or order of the independent variables to which subjects are exposed, does not impact on the results that they attain. The technique of counterbalancing is likely to be the most effective means to manage the threat of time-tied extraneous variables in validation studies that employ within-subjects experimental designs to estimate the validities associated with structured employment interviews.

The threat of time-tied extraneous variables also features in between-subjects experimental research designs that use non-equivalent groups of subjects. Stern and Kalof (1996) suggested that in between-subjects, non-equivalent groups, experiments, this threat is not as severe as it is in studies that employ within-subjects experimental research design strategies. However, the possibility for them to feature in the former context implies that they should be effectively controlled. In the between-subjects experiment conducted by Stohr-Gilmore et al. (1990), it was noted that the implications of time-tied extraneous variables on the validity estimates obtained for the situational interview under investigation, were minimal. The effect of this threat was only realised at the point when the validity estimates for the two types of interviews under investigation were compared. This implies that in the attainment of the validity estimate for the structured employment interview in question, the impact of this threat was non-existent. Therefore, it is unlikely that the use of between-subjects experimental research design strategies for the validation of structured employment interviews will yield significant threats to the

internal validity of the findings that are obtained for specific interview variants in terms of time-tied variables in the South African context.

However, the threat of time-tied extraneous variables is likely to be of consequence in within-subjects experimental research designs, especially if the researchers who use this strategy expose each of their subjects, not only to both the treatment conditions, but to two instances of each condition, just as Janz (1982) and Orpen (1985) did. Researchers in the South African context who contemplate the use of this strategy should take cognisance of the threat of time-tied extraneous variables to which their endeavours are likely to be exposed. In order to address this threat, constructive measures, such as the technique of counterbalancing may be employed.

### 6.1.2.3 Invalid Operational Definitions

The threat of invalid operational definitions is of equal severity to the threats of organismic variables and time-tied extraneous variables in between-subjects experiments that employ non-equivalent groups of individuals as subjects (Stern and Kalof, 1996). However, the threat of invalid operational definitions is regarded as less significant than the afore-mentioned threats in correlational and within-subjects experimental research design strategies (Stern and Kalof, 1996).

The author noted that the language proficiency of subjects could serve as a potential alternative explanation for the findings attained in validation research, since it could confound the operational definition of the independent variables under investigation. In South Africa where there are eleven official languages, this problem is likely to manifest. It is imperative for South African researchers to attempt to ensure that their subjects are comfortable understanding and expressing themselves in the language in which employment interviews are conducted. Given the diversity of individuals that are likely to apply for positions in organisations, it may be necessary to translate structured employment interviews into other languages, based on the needs of the subjects. This would necessarily have implications for the organisations, interviewers and researchers in question. These role-players would have to ensure that the interviews are appropriately translated and that they have interviewers, who are proficient in the languages in question, in order to effectively conduct the interviews and interpret the responses that they elicit from subjects. This is likely to incur extra costs for organisations in actual selection situations and for researchers in simulated selection contexts. The results that are likely to be attained will be invaluable in researchers' attempts to rule out language proficiency as a possible confounding factor of the independent variables under investigation in validation research on structured employment interviews.

The stereotypes, prejudices and biases instilled among the different racial groupings during the apartheid era in South Africa could potentially manifest in the interviewers', performance evaluators' and subjects' roles in validation research on structured employment interviews. These subjective factors, which may be introduced by human participants into the validation process, could serve to confound the definitions of the variables under investigation. In these instances, researchers may be confronted with situations in which the interview scores and/or the performance evaluation scores are a function of the subjective biases of the interview or job performance evaluators and/or the subjects. In the South African context, the threat of invalid operational definitions may be minimised by reducing the potential for subjective biases to serve as alternative explanations for the findings that are attained. This could be accomplished by ensuring that evaluators are adequately and appropriately trained in how to execute their roles optimally and that the process is closely monitored through the use of audio or video technology. In this manner researchers could strive to ensure that the information obtained is congruent with the operational definitions that they have earmarked for investigation. It is imperative, however, that should such monitoring occur, the consent of the subjects is attained. This control may only be viable (if resources are available) in validation efforts that are conducted in simulated, as opposed to actual, employment situations, since these may be perceived as less threatening by both organisations (employers) and interviewees alike, in the author's opinion.

In the validation of behaviour description interviews, memory deficits in subjects could render the operational definitions of the independent variable invalid. It is possible that the findings of these research efforts could be attributed to the ability or inability of subjects to accurately recall past situations and how they behaved in them. It is imperative for South African researchers who are interested in validating variants of behaviour description interviews to take cognisance of this possible confounding of the independent variable. In order to control for this threat, researchers could opt to use other types of structured interviews (i.e. situational interviews) in order to eliminate it entirely. In situational interviews, this threat is non-existent because the interview questions require subjects to provide accounts of future intentions in the context of hypothetical situations. Behaviour description interview questions, on the other hand, require actual accounts of past experiences. Another way in which this threat may be minimised in validation research on behaviour description interviews is by requesting that subjects to offer accounts of their most recent relevant experiences in their responses to behaviour description interview questions. If they do, in fact, take heed of this request, the threat of invalid operational definitions, which is introduced by memory deficits, as a possible alternative explanation for the findings that are attained, could be reduced.



In the South African context, previously disadvantaged groups of individuals have historically been denied access to opportunities in education and employment. These individuals may not have relevant previous work experience to draw on in behaviour description interviews. This reality may serve to confound the operational definitions that are formulated, by researchers, for the independent variables in the validation research on these interviews. In the author's view, if behaviour description interviews are conducted with applicants with no previous relevant work experience, then the information that is derived from these individuals will deviate from the type of information that these interviews were originally intended to elicit. This could confound the results that are ultimately obtained by rendering the operational definition of the independent variable invalid. This difficulty may be overcome in the South African context in two ways. Firstly, researchers could tailor their definition of the independent variable (i.e. a variant of behaviour description interviews) to include any previous experience whether or not it occurred in an employment situation. Secondly, researchers could make use of other types of structured employment interviews such as situational interviews, which do not depend on subjects' previous experiences.

In the previous chapter, it was argued that the problem of memory deficits could also plague interviewers and evaluators of job performance. This could detract from the internal validity of validation efforts on structured employment interviews, by rendering the scores that are assigned to subjects on the interview and criterion measures, a function of the ability of the evaluators to remember what the subjects said and did in these contexts. This threat can be minimised if interviewers and job performance evaluators keep accurate records or notes of the subjects' behaviour during the period in which they are evaluated. This is likely to reduce the potential for memory deficits by interviewers and job performance evaluators to feature as an alternative explanation for the results that are obtained in validation efforts.

#### 6.1.2.4 Sampling Biases

Stern and Kalof (1996) identified sampling biases and uncontrolled variation in information as the least significant threats to internal validity in the three research design strategies that were employed by international researchers to conduct validation research on structured variants of employment interviews (viz. correlational studies, within-subjects experiments and between-subjects experiments, non-equivalent groups). Sampling biases occur when the samples that are obtained are not representative of the populations from which they are drawn. Stern and Kalof (1996) proposed two key ways in which the threat of sampling biases could be minimised. Firstly, they suggested that the technique of random sampling be used. Secondly, if this is not possible, they proposed the use of the technique of purposive sampling to ensure that the sample is representative of the larger population

from which it is drawn. Stern and Kalof (1996) proposed that if samples differ systematically from the populations from which they are derived, then alternative explanations may exist for the findings that are yielded in the context of these samples.

In the author's opinion, the use of random samples of subjects in validation research in actual selection situations is inappropriate. This is due to the reality that individuals are generally selected for different phases of the selection process on the basis of selectors' perceptions of whether or not they are likely to be suited to the position in question. Therefore, random sampling does not appear to be a viable option in this context. It may be more feasible in validation research, which is conducted for the sole purpose of research in simulated selection contexts. It is likely to be more appropriate to use the technique of purposive sampling in actual selection contexts. In South Africa, there are groups of individuals that have been historically disadvantaged as a result of the legalisation of apartheid. In order to remedy the injustices of the past, organisations are compelled to institute employment equity practices and to implement these fairly. The author believes that the use of purposive sampling is likely to be invaluable in obtaining applicants that are representative of the larger population from which they are drawn in terms of a number of criteria such as race, gender, age, etc. In addition, selectors in South African organisations should endeavour to select employees on the basis of these ratios in order to ensure an equitable and representative work force.

The author noted that there are two other threats in relation to sampling that may serve to flaw the findings that are obtained in validation research on structured employment interviews. The first pertains to the use of small samples to obtain validity estimates. According to research methodology theory, the use of small samples is likely to yield contaminated validity results. In actual selection situations, however, small samples may be inevitable. In such situations, statistical measures could be employed to control for the impact of small sample sizes on the findings that are attained. A second factor, which features prominently in actual selection situations, is restriction of range. This phenomenon occurs because only those individuals with high interview scores are likely to be selected and consequently, included in the sample for validation. Researchers in the South African context could also employ statistical procedures to control for restriction of range when they are confronted with this phenomenon. In order to control for the respective problems of small sample sizes and restriction of range in more obvious ways researchers could contemplate including more subjects in the samples for validation research and selecting all the individuals instead of only those that performed well on the selection measures. However, due to the financial implications of these suggestions for organisations, the use of statistical controls for these problems usually has to suffice in validation research that is conducted in actual employment situations.

#### 6.1.2.5 Uncontrolled Variation in Information

Stern and Kalof (1996) argued that the threat of uncontrolled variation in information is usually prevalent when researchers are charged with the task of obtaining the same information from or about different people. They proposed that one way in which to control for uncontrolled variation in information is to hold the interviewers or performance evaluators constant (Stern and Kalof, 1996). However, it was noted, in the studies that were analysed in this investigation, that it was not always feasible for researchers to hold these individuals constant in validation efforts on structured variants of employment interviews. Stern and Kalof (1996) suggested that this technique would not always be appropriate and proposed that researchers hold the procedures that they use constant, in order to counteract this threat to internal validity. In structured employment interviews, interviewers strive for consistency by using standardised procedures and processes to interview their applicants and evaluate their responses. Some of the ways in which the researchers did invariably control for this threat, albeit inadvertently, was by:

- Using consistent procedures in the design, implementation and evaluation phases of the interview and criterion measures;
- Exposing the interviewers and, in some cases, the performance evaluators to standardised training programs;
- Using panel interview designs as opposed to one-on-one interview designs; and
- Conducting pilot studies to reduce ambiguity in interview questions.

The use of these procedures could assist in counteracting the threat of uncontrolled variation in information in validation research on structured employment interviews. However, they do not guarantee the elimination of this threat. The author believes that the use of monitoring procedures can assist researchers to ensure that the information on which they base their validity results is consistent and oriented towards the measurement of the same essential variables and phenomena, provided that researchers obtain consent from their subjects for the use of these procedures.

#### 6.1.3 Threats to External Validity

South African researchers, who are keen to validate structured employment interviews, should take cognisance of the threats to internal validity that they are likely to encounter depending on the research design strategies that they employ to conduct their investigations. The recommendations that have

been proposed for controlling these threats are likely to be invaluable in ensuring the internal validity of validation research. In their attempts to strive for internal validity in their validation research, South African researchers will be able to ensure the generalisability of their validity findings in the realm of structured employment interviews. In this way, South African researchers can build a valuable database of internally and externally valid validity research on structured employment interviews in order to justify their use for the purpose of rendering crucial selection decisions in compliance with South Africa's employment equity legislation.

In this study the author outlined the threats to external validity which could have hampered the generalisability of the findings obtained in validation research on patterned behaviour description interviews, structured behavioural interviews and situational interviews respectively. The threats to external validity revolved around the differences in research investigations in terms of:

- The experimental conditions;
- The context in which the research was conducted;
- The individuals that participated in the investigations; and
- The time periods across which the research was conducted.

In the international investigations reviewed in this study, differences in the experimental conditions, the research context, the research participants and the times at which the research was conducted, generally, did not appear to detract from the comparability of the findings yielded in the studies that were compared. However, only further research on external validity in the South African context, will assist researchers to determine whether or not these threats to external validity are likely to manifest in local validation research.

In order to determine whether or not these threats to external validity did, in fact, manifest in the studies that were analysed, the author used a meta-analytic technique proposed by Rosenthal (1991) to compare the validity findings yielded for structured employment interviews. This technique yielded consistency in the validity findings that were obtained for patterned behaviour description interviews, structured behavioural interviews and seven of the eight investigations on situational interviews. This finding implied that, in general, differences in the experimental conditions, the research context, the research participants and the times at which the research was conducted did not appear to have impeded the generalisability or comparability of the findings that were attained. The author concluded,

on the basis of these results that these investigations were likely to possess external validity provided that they possessed internal validity.

The comparisons of the findings that were obtained in one of the studies by Latham and Saari (1984), with the findings yielded in the other validation research on situational interviews, revealed that this validity estimate differed significantly from the other estimates. The author perceived that the most significant difference between the study by Latham and Saari (1984) and the other investigations on situational interviews was the manner in which the situational interview was implemented. It is possible that the incorrect use of the interview by the interviewers in Latham and Saari's (1984) investigation could have detracted from the generalisability of this research effort to other populations, settings and time-periods provided that the conditions for internal validity were met.

Using the meta-analytic technique for comparing the findings yielded by international validation initiatives on structured employment interviews, it was concluded that differences in experimental conditions, research contexts, research participants and time periods did not appear to influence the comparability of the validity results. This is based on the assumption that the studies that were compared possessed internal validity. However, it is possible that validation research in the South African context will yield different results. In order to comment on the external validity of validation research on structured employment interviews, it is imperative that multiple investigations are engaged in. These investigations may only be compared if they are oriented towards assessing the same essential variables. Using these criteria as guidelines, South African researchers could explore the external validity of the findings they obtain in their validation research on structured employment interviews in order to ascertain whether or not these are generalisable to other experimental conditions, research contexts, research participants and time periods.

## 6.2 Conclusion

In the light of South Africa's employment equity legislation, organisations are compelled to use measurement instruments that are psychometrically sound in terms of validity, reliability and adverse impact, in order to render key organisational decisions. If South African organisations wish to continue using employment interviews to render such decisions, then they will have to adopt the use of structured variants of these interviews, which have been shown, in international research efforts, to possess higher validity and reliability than their unstructured counterparts. Presently, validation research on employment interviews, in general, and structured employment interviews, in particular, is virtually non-existent in the South African context. However, this state of affairs will have to alter in

order for South African organisations to comply with the requirements of employment equity legislation. The author believes that validation research on structured employment interviews will have to become a reality in the very near future in order for South African organisations to prove that their instruments are valid and reliable. In anticipation of this reality, the author believes that the insights that have been provided in this study on the threats to internal and external validity of international research efforts will assist South African researchers, in particular, to engage in productive validation research on structured employment interviews in the South African context.

In this investigation, the author engaged in a qualitative analysis of secondary research. Although the exploration of the threats to the internal and external validity of the studies included for analysis was attempted in a systematic manner, it is possible that subjective errors of judgement and perception could have occurred. This difficulty may be overcome by future researchers in the context of primary validation research on structured employment interviews. Future researchers should include the mechanisms that they employ to control for potential threats to internal validity as independent variables in studies that are characterised by experimental research design strategies. In this manner, the implications of these controls may be investigated. This information is likely to be invaluable to other validation researchers who can draw on these insights and improve their subsequent research.

In this study, the threats to internal and external validity that could have manifested in international research initiatives were explored independently. This constituted a further limitation in this study. In the exploration of the potential threats to internal validity the author did not comment on whether or not the studies in question actually possessed the quality of internal validity. The use of this approach implied that it was not possible to offer a definitive comment on whether or not the studies were generalisable to other contexts, populations and time-periods. However, a meta-analytic technique was used to compare the findings obtained in international research on structured employment interviews. On the basis of these comparisons a commentary was offered on whether or not the potential threats to external validity were, indeed, actualised in the multiple investigations on patterned behaviour description interviews, structured behavioural interviews and situational interviews. The author commented, on the basis of these findings, on whether the studies in question would have possessed external validity had their internal validity been established. Future researchers can build on the present study by commenting on the internal validity of the studies that were analysed. This information can then be used to comment on their external validity.

In order for research efforts to be sound, it is imperative that they possess validity and reliability. In the present investigation, the author only considered the validity of the studies in question and not their reliability. This constitutes a further limitation of the study. Future researchers could consider the

investigations that were analysed in the present investigation, in terms of their reliability in order to offer a more holistic picture of their adequacy as research initiatives.

## **Bibliography**

Arvey, R.D. and Campion, J.E. (1982). The Employment Interview: A Summary and Review of Recent Research. Personnel Psychology, 35, pp. 281-322.

Berman, J.A. (1997). Competence-Based Employment Interviewing. Westport, Connecticut: Quorum Books.

Black, T.R. (1993). Evaluating Social Science Research: An Introduction. London: Sage Publications.

Burnett, J.R., Fan, C., Motowidlo, S.J. and Degroot, T. (1998). Interview Notes and Validity. Personnel Psychology, 51, pp. 375-396.

Burnett, J.R. and Motowidlo, S.J. (1998). Relations Between Different Sources of Information in the Structured Selection Interview. Personnel Psychology, 51, pp. 963-983.

Campion, M.A., Palmer, D.K. and Campion, J.E. (1997). A Review of Structure in the Selection Interview. Personnel Psychology, 50, pp. 655-702.

Campion, M.A., Pursell, E.D. and Brown, B.K. (1988). Structured Interviewing: Raising the Psychometric Properties of the Employment Interview. Personnel Psychology, 41, pp. 25-42.

Cascio, W.F. (1991). Applied Psychology in Personnel Management (Fourth Edition). Englewood Cliffs, New Jersey: Prentice-Hall International.

Conway, J.M., Goodman, D.F. and Jako, R.A. (1995). A Meta-Analysis of Interrater and Internal Consistency Reliability of Selection Interviews. Journal of Applied Psychology, 80(5), pp. 565-579.

Dipboye, R.L. (1992). Selection Interviews: Process Perspectives. Cincinnati, Ohio: South-Western Publishing Company.

Dipboye, R.L. (1997). Structured Selection Interviews: Why Do They Work? Why Are They Underutilized? In N. Anderson and P. Herriot (Editors), International Handbook of Selection and Assessment (pp. 455-473). Chichester: John Wiley and Sons.



Dooley, D. (1995). Social Research Methods (Third Edition). Englewood Cliffs, New Jersey: Prentice-Hall.

Dreher, G.F., Ash, R.A. and Hancock, P. (1988). The Role of the Traditional Research Design in Underestimating the Validity of the Employment Interview. Personnel Psychology, 41, pp. 315-327.

Dyer, C. (1995). Beginning Research in Psychology. Oxford: Blackwell Publishers.

Eder, R.W. and Ferris, G.R. (Editors) (1989). The Employment Interview: Theory, Research, and Practice. London: Sage Publications.

Eder, R.W., Kacmar, K.M. and Ferris, G.R. (1989). Employment Interview Research: History and Synthesis. In R.W. Eder and G.R. Ferris (Editors), The Employment Interview: Theory, Research, and Practice (pp. 17-31). London: Sage Publications.

Employment Equity Act, No. 55. (1998). Government Gazette, 400(19370), pp. 1-54.

French, W.L. and Bell, C.H., Jr. (1995). Organization Development: Behavioral Science Interventions for Organization Improvement (Fifth Edition). Englewood Cliffs, New Jersey: Prentice-Hall International.

Gabris, G.T. and Rock, S.M. (1991). Situational Interviews and Job Performance: The Result in One Public Agency. Public Personnel Management, 20(4), pp. 469-483.

Gatewood, R.D. and Field, H.S. (1994). Human Resource Selection (Third Edition). Fort Worth, Texas: The Dryden Press/Harcourt Brace College Publishers.

Gehrlein, T.M., Dipboye, R.L. and Shahani, C. (1993). Nontraditional Validity Calculations and Differential Interviewer Experience: Implications for Selection Interviews. Educational and Psychological Measurement, 52, pp. 457-469.

Gerber, P.D., Nel, P.S. and van Dyk, P.S. (1995). Human Resources Management (Third Edition). Halfway House, Johannesburg: International Thomson Publishing (Southern Africa).

Hakel, M.D. (1971). Similarity of Post-Interview Trait Rating Intercorrelations as a Contributor to Interrater Agreement in a Structured Employment Interview. Journal of Applied Psychology, 55(5), pp. 443-448.

Handy, C. (1993). Understanding Organizations (Fourth Edition). London: Penguin Books.

Handy, C. (1994). The Empty Raincoat: Making Sense of the Future. Sydney: Arrow Business Books.

Harris, M.M. (1989). Reconsidering the Employment Interview: A Review of Recent Literature and Suggestions for Future Research. Personnel Psychology, 42, pp. 691-726.

Heiman, G.W. (1998). Understanding Research Methods and Statistics: An Integrated Introduction for Psychology. New York: Houghton Mifflin Company.

Heneman, H.G., Schwab, D.P., Huett, D.L. and Ford, J.J. (1975). Interviewer Validity as a Function of Interview Structure, Biographical Data, and Interviewee Order. Journal of Applied Psychology, 60(6), pp. 748-753.

Huffcutt, A.I. and Arthur, W. Jr. (1994). Hunter and Hunter (1984) Revisited: Interview Validity for Entry-Level Jobs. Journal of Applied Psychology, 79(2), pp. 184-190.

Hunter, J.E. and Hunter, R.F. (1984). Validity and Utility of Alternative Predictors of Job Performance. Psychological Bulletin, 96(1), pp. 72-98.

Hunter, J.E. and Schmidt, F.L. (1990). Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. London: Sage Publications.

Janz, T. (1982). Initial Comparisons of Patterned Behavior Description Interviews Versus Unstructured Interviews. Journal of Applied Psychology, 67(5), pp. 577-580.

Janz, T. (1989). The Patterned Behaviour Description Interview: The Best Prophet of the Future Is the Past. In R.W. Eder and G.R. Ferris (Editors), The Employment Interview: Theory, Research, and Practice (pp. 158-168). London: Sage Publications.

- Janz, T., Hellervik, L. and Gilmore, D.C. (1986). Behavior Description Interviewing: New, Accurate, Cost Effective. Englewood Cliffs, New Jersey: Prentice-Hall.
- Jelf, G.S. (1999). A Narrative Review of Post-1989 Employment Interview Research. Journal of Business and Psychology, 14(1), pp. 25-58.
- Latham, G.P. (1989). The Reliability, Validity, and Practicality of the Situational Interview. In R.W. Eder and G.R. Ferris (Editors), The Employment Interview: Theory, Research, and Practice (pp. 169-182). London: Sage Publications.
- Latham, G.P. and Saari, L.M. (1984). Do People Do What They Say? Further Studies on the Situational Interview. Journal of Applied Psychology, 69(4), pp. 569-573.
- Latham, G.P., Saari, L.M., Pursell, E.D. and Campion, M.A. (1980). The Situational Interview. Journal of Applied Psychology, 65(4), pp. 422-427.
- Lewis, C. (1992). Employee Selection (Second Edition). Cheltenham, England: Stanley Thornes Publishers.
- Locke, E.A. (1966). The Relationship of Intentions to Level of Performance. Journal of Applied Psychology, 50(1), pp. 60-66.
- Locke, E.A. and Latham, G.P. (1990). A Theory of Goal Setting and Task Performance. Englewood Cliffs, New Jersey: Prentice-Hall.
- London, M. (1995). Self and Interpersonal Insight: How People Gain Understanding of Themselves and Others in Organizations. Oxford: Oxford University Press.
- Maas, J.B. (1965). Patterned Scaled Expectation Interview: Reliability Studies on a New Technique. Journal of Applied Psychology, 49(6), pp. 431-433.
- Mayfield, E.C. (1964). The Selection Interview – A Re-evaluation of Published Research. Personnel Psychology, 17, pp. 239-260.

- McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. and Maurer, S.D. (1994). The Validity of Employment Interviews: A Comprehensive Review and Meta-Analysis. Journal of Applied Psychology, 79(4), pp. 599-616.
- Moffatt, G.W.B. (1969). The Selection Interview: A Review. Personnel Practice Bulletin, 25(1), pp. 15-23.
- Motowidlo, S.J. and Burnett, J.R. (1995). Aural and Visual Sources of Validity in Structured Employment Interviews. Organizational Behavior and Human Decision Processes, 61(3), pp. 239-249.
- Motowidlo, S.J., Carter, G.W., Dunnette, M.V., Tippins, N., Werner, S., Burnett, J.R. and Vaughan, M.J. (1992). Studies of the Structured Behavioral Interview. Journal of Applied Psychology, 77(5), pp. 571-587.
- Muchinsky, P.M. (1993). Psychology Applied to Work: An Introduction to Industrial and Organizational Psychology (Fourth Edition). Pacific Grove, California: Brooks/Cole Publishing Company.
- Muchinsky, P.M., Kriek, H.J. and Schreuder, A.M.G. (1998). Personnel Psychology. Johannesburg: International Thomson Publishing (Southern Africa).
- Orpen, C. (1985). Patterned Behavior Description Interviews Versus Unstructured Interviews: A Comparative Validity Study. Journal of Applied Psychology, 70(4), pp. 774-776.
- Pulakos, E.D. (1997). Ratings of Job Performance. In D.L. Whetzel and G.R. Wheaton (Editors), Applied Measurement Methods in Industrial Psychology (pp. 291-317). Palo Alto, California: Davies-Black Publishing.
- Reilly, R.R. and Chao, G.T. (1982). Validity and Fairness of Some Alternative Employee Selection Procedures. Personnel Psychology, 35, pp. 1-62.
- Robertson, I.T., Gratton, L. and Rout, U. (1990). The Validity of Situational Interviews for Administrative Jobs. Journal of Organisational Behaviour, 11, pp. 69-76.
- Rosenthal, R. (1991). Meta-Analytic Procedures for Social Research (Revised Edition). London: Sage Publications.

Rosenthal, R. and Rosnow, R.L. (1984). Essentials of Behavioral Research: Methods and Data Analysis. New York: McGraw-Hill Book Company.

Ryan, A.M., McFarland, L., Baron, H. and Page, R. (1999). An International Look at Selection Practices: Nation and Culture as Explanations for Variability in Practice. Personnel Psychology, 52, pp. 359-391.

Rynes, S.L. (1989). The Employment Interview as a Recruitment Device. In R.W. Eder and G.R. Ferris (Editors), The Employment Interview: Theory, Research, and Practice (pp. 127-141). London: Sage Publications.

Sarantakos, S. (1998). Social Research (Second Edition). London: Macmillan Press Ltd.

Schein, E.H. (1980). Organizational Psychology (Third Edition). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Schmitt, N. (1976). Social and Situational Determinants of Interview Decisions: Implications for the Employment Interview. Personnel Psychology, 29, pp. 79-101.

Schmidt, F.L. and Hunter, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. Psychological Bulletin, 124(2), pp. 262-274.

Schultz, D.P. and Schultz, S.E. (1986). Psychology and Industry Today: An Introduction to Industrial and Organizational Psychology (Fourth Edition). New York: Macmillan Publishing Company.

Schwab, D.P. and Heneman, H.G. (1969). Relationship Between Interview Structure and Interviewer Reliability in an Employment Situation. Journal of Applied Psychology, 53(3), pp. 214-217.

Shackleton, V. and Newell, S. (1997). International Assessment and Selection. In N. Anderson and P. Herriot (Editors), International Handbook of Selection and Assessment (pp. 81-95). Chichester: John Wiley and Sons.

Shaughnessy, J.J. and Zechmeister, E.B. (1997). Research Methods in Psychology (Fourth Edition). New York: The McGraw-Hill Companies, Inc.

Society for Industrial Psychology, (1998). Guidelines for the Validation and Use of Assessment Procedures for the Workplace. Auckland Park: Society for Industrial Psychology.

Stern, P.C. and Kalof, L. (1996). Evaluating Social Science Research (Second Edition). Oxford: Oxford University Press.

Stohr-Gilmore, M.K., Stohr-Gilmore, M.W. and 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(2), pp. 1-18.

Taylor, P. and O'Driscoll, M. (1995). Structured Employment Interviewing. Hampshire, England: Gower Publishing Limited.

Terpstra, D.E., Mohamed, A.A. and Kethley, R.B. (1999). An Analysis of Federal Court Cases Involving Nine Selection Devices. International Journal of Selection and Assessment, 7(1), pp. 26-34.

Ulrich, L. and Trumbo, D. (1965). The Selection Interview Since 1949. Psychological Bulletin, 63(2), pp. 100-116.

Wagner, R. (1949). The Employment Interview: A Critical Summary. Personnel Psychology, 2, pp. 17-46.

Weekley, J.A. and Gier, J.A. (1987). Reliability and Validity of the Situational Interview for a Sales Position. Journal of Applied Psychology, 72(3), pp. 484-487.

Whetzel, D.L. and McDaniel, M.A. (1997). Employment Interviews. In D.L. Whetzel and G.R. Wheaton (Editors), Applied Measurement Methods in Industrial Psychology (pp. 185-205). Palo Alto, California: Davies-Black Publishing.

Whetzel, D.L. and Oppler, S.H. (1997). Validation of Selection Instruments. In D.L. Whetzel and G.R. Wheaton (Editors), Applied Measurement Methods in Industrial Psychology (pp. 355-384). Palo Alto, California: Davies-Black Publishing.

Whetzel, D.L. and Wheaton, G.R. (Editors) (1997). Applied Measurement Methods in Industrial Psychology. Palo Alto, California: Davies-Black Publishing.

Wiesner, W.H. and Cronshaw, S.F. (1988). A Meta-Analytic Investigation of the Impact of Interview Format and Degree of Structure on the Validity of the Employment Interview. Journal of Occupational Psychology, 61, pp. 275-290.

Wood, R. and Payne, T. (1998). Competency Based Recruitment and Selection: A Practical Guide. Chichester: John Wiley and Sons.

Wright, O.R. (1969). Summary of Research on the Selection Interview Since 1964. Personnel Psychology, 22, pp. 391-413.

Wright, P.M., Lichtenfels, P.A. and Pursell, E.D. (1989). The Structure Interview: Additional Studies and a Meta-Analysis. Journal of Occupational Psychology, 62, pp. 191-199.

Yonge, K.A. (1956). The Value of the Interview: An Orientation and a Pilot Study. Journal of Applied Psychology, 40(1), pp. 25-31.

Zedeck, S., Tziner, A. and Middlestadt, S.E. (1983). Interviewer Validity and Reliability: An Individual Analysis Approach. Personnel Psychology, 36, pp. 355-370.