Exploring and Monitoring Interviewers' Cognitive Processes of Decision Making Using Think Aloud Method

Shiyun Zhao

Design, manufacturing & Engineering Management

University of Strathclyde, Glasgow

February 24, 2023

It never will rain roses.

When we want to have more roses we must plant trees.

- George Eliot

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

Employment interview has been one of the most commonly used selection tools across periods, regions, and industries. Aiming to select and hire talents for organisations, the performance of interviewers as the decision maker of employment interview has received wide interests from both practitioners and researchers. Regardless of a persistent call for research into the process of interviewers' decision making (IDM), most existent studies base their conclusions on product analysis that examines the mathematical correlation between interviewers' final decisions and various factors at a group level. In particular, a gap has been identified through a systematic literature review that interviewers' cognitive processes (ICP), which plays a significant role in interviewers' decision making, is under-investigated. Such a gap has not only limited the understanding of IDM but also constrained the possibility to improve it. To address the gap, this research inaugurally introduces a psychological method termed "think aloud method (TAM)" to explore ICP when completing an interview decision-making task. Specifically, the research first tested and confirmed the feasibility of applying TAM to examining ICP and developed an experiment procedure where TAM was effectively embedded in interview scenario. Then, rich data of ICP was collected from 29 participants with diversified backgrounds, all of whom were capable of working on the interview task while thinking aloud. Through a multi-level analysis of the data collected, (i) four types of information and 17 categories of cognitive actions following three information-processing strategies were recognised in ICP, (ii) four major characteristics of ICP were identified and quantified to enrich the understanding of the panorama, (iii) three cognitive dimensions were constructed to measure the corresponding aspects of

interviewers' judging behaviour that were believed to affect their decision quality, (iv) direct evidence of seven typical decision patterns and bias were identified from ICP, and (v) associations between various contextual factors (i.e., gender, national region of interview experience, experience level as interviewer, rating scale type) and ICP were tested, where the impact of these factors on ICP were found except for interviewers' gender. The key findings of this research as well as their niche within the broader context of IDM research are discussed in detail with the trustworthiness of the research justified. This research uncovering the black box of interviewers' cognitive process is expected to contribute both to the academic knowledge of IDM and the real-world practice of employment interview in various aspects, including promoting the understanding of IDM process at a cognitive level, monitoring the ongoing process of IDM rather than relying on the decision outcomes, guiding the improvement of interview design, and facilitating the development of interviewer training courses as well as decision aids. Potential limitations as a result of sample size and fixed task scenario are also discussed with the opportunities of future research highlighted.

Acknowledgement

Foremost, I would like to express my deepest gratitude to my supervisor, Dr. Andy T.C. Wong, whose professional guidance and patient help have carried me throughout all the stages of this research. My sincere thanks also go to Prof. Yi Qin, my secondary supervisor, and the colleagues of DMEM, who have continuously supported me with their thoughtful feedback.

I owe my greatest appreciation to all the contributors to this research, whose generosity, kindness, and valuable outputs have made the work possible. A special word of gratitude is owed to Thomas O'Connor for his insightful advice on the topic and cordial concern for the research progress; to Dr. Youssef Shaalan, Dr. He Sun, and Dr. Jinkai Zhang for their self-giving assistance in the development of this research while they were fighting own ways through PhD programmes; and to my dearest friends, Chenkai Ma, Helin Zhao, and Miaomiao Chang, for all the comforts and courage they have offered me in those tough time.

I am greatly indebted to my parents for their sincere, eternal, and unconditional love. They signified their whole-hearted support when I proposed to go for a PhD adventure, and they have been my strongest backing throughout these years. In spite of the tremendous costs – in every sense - they afforded, mom and dad have never required anything from me except for health and happiness.

I also want to thank Zhenyu Fang, my caring, loving, and supportive husband (in the near future). I was extremely fortunate to meet him in the summer of 2015, and

we shared all those laughs and tears, discoveries and failures, encounters and partings in the following seven years studying abroad together.

By the end of this acknowledgement, I want to take the opportunity to express my remembrance of Grandpa who had left right before I finished the research. “She will come back as soon as she completes her thesis”, he said.

Contents

List of Tables

List of Figures

Acronyms/Abbreviations

AJG	Academic Journal Guide
AI	Artificial intelligence
BARS	Behavioural-anchored rating scale
BBL	Behavioural-based level
CCL	Criteria-compliance level
EBL	Evidence-based level
HRM	Human resource management
IM	Impression management
ICP	Interviewers' cognitive processes
IDM	Interviewers' decision making
LTM	Long-term memory
STM	Short-term memory
SLR	Systematic literature review
TAM	Think aloud method

Chapter 1

Introduction

Employment interviews remain one of the most commonly used tools to select and hire talent (Harris, 1989; Macan, 2009; Levashina et al., 2014; Wanberg et al., 2020; Basch et al., 2021; Chauhan, 2022), where the interviewer as the decision maker is often identified as a crucial role for ensuring reliable and valid decision outcomes (Fox & Spector, 2000; Purkiss et al., 2006; Kausel et al., 2016; Buijsrogge et al., 2021). In order to study and improve interviewers' decision making (IDM), a great deal of studies have accumulated in the past ten decades. However, the academic attention has focused on the properties of interviewers' decisions (Maurer, 2002; Sacco et al., 2003; Agerström & Rooth, 2011; Kroll & Ziegler, 2016) rather than the ongoing process through which interviewers reach their final decisions. More importantly, while decision making is a cognitive process composed of a logical sequence of information-processing activities taking place in the decision maker's mind (Payne, 1994, Wang & Ruhe, 2007; Lunenburg, 2010; Mihaylov, 2019), there is a severe lack of investigation into interviewers' cognitive processes (ICP). Therefore, this research aims to contribute to both the understanding and the monitoring of IDM from a procedural perspective by exploring ICP during decision making.

This chapter provides an introduction to the overall research by sequentially discussing the background, niche, construct, and output of the research. First, **Section**

1.1 briefs the broader context of this research in the sense that there has been a call for investigation into the process of IDM, encouraged by which massive empirical studies have been carried out. However, the majority of these studies examine the mathematical associations between the decisions made by interviewers and various factors, and only a limited number of studies explore how IDM process unfolds. Thus, **Section 1.2** goes further into the studies focusing on the ongoing process of IDM with a particular emphasis on the features of thematic focus and methodologies adopted, from which a gap is identified in IDM research and a series of research questions to cover the gap are raised in **Section 1.3** with rationale explained. Then in **Section 1.4**, the methodology of data collection and analysis adopted in this research to seek for the answers to the research questions is described together with expected outcomes, followed by a demonstration of the significance and potential limitations of this research in **Section 1.5**. Finally, **Section 1.6** presents the structural outline of this dissertation in terms of a brief summary of both the purpose and contents of each chapter.

1.1 Background of the Research

Motivated by the great need to select and hire talents for organisations through employment interviews, vast numbers of studies are carried out aiming to study and improve IDM from various perspectives, which can be dated back to the research carried out by Scott (1915). In particular, there has been a long-standing call for process research to explore how the IDM process unfolds (Lewis, 1980; Harris, 1989; Tullar, 1989; Anderson, 1992; Graves, 1993; Posthuma et al., 2002; Macan, 2009; Levashina et al., 2014) so as to promote the understanding and the improvement of IDM.

The earliest research is constructed in a "how-to-do-it" manner based on experience and common sense without providing empirical evidence (Webster & Anderson, 1964; Mayfield, 1964; Ulrich & Trumbo, 1965). Later in 1950s, a series of empirical studies on IDM as integrated and reported by Webster and Anderson (1964) are regarded as beginning step toward a better understanding of the process of IDM (Mayfield,

1964) that not only arouse the academic interests in the process of IDM but also set an example of how empirical investigation can be conducted into it. Consequently, interviewers' decisions are investigated and compared in the majority of the subsequent studies to examine how a wide range of factors may affect the decision properties, such as reliability, validity and applicants' reactions (Zedeck et al., 1983; Taylor & Sniezek, 1984; McDaniel et al., 1994; Conway et al., 1995; Moscoso, 2000; Huffcutt et al., 2013; Woods et al., 2020). The corresponding studies are discussed by a series of reviews throughout the past decades (Schmitt, 1976; Arvey & Campion, 1982; Harris, 1989; McDaniel et al., 1994; Judge et al., 2000; Moscoso, 2000; Posthuma et al., 2002; Macan, 2009; Levashina et al., 2014).

In contrast, other studies look into the process of IDM more straightforwardly by paying attention either to interviewers' observable behaviour during decision making, such as questioning (Macan & Dipboye, 1990; Dougherty et al., 1994; Silvester & Anderson, 2003; Highhouse et al., 2019) and notetaking (Macan & Dipboye, 1994; Burnett et al., 1998; Biesanz et al., 1999; Middendorf & Macan, 2002), or to their inner thinking processes, like attributional style (i.e., how interviewers interpret the clues obtained from applicants; Gifford et al., 1985; Highhouse & Bottrill, 1995; Chapman & Webster, 2001; Reinhard et al., 2013; Culbertson et al., 2016), decision path (i.e., how interviewers' decisions are progressively formed with various intermediates involved; Kristof-Brown, 2000; Roulin et al., 2014; Kith et al., 2022) cue weighting (i.e., how interviewers weigh different cues when making decisions; Graves & Karren, 1992; Burnett & Motowidlo, 1998; DeGroot & Gooty, 2009). Another heated topic is the detection of a broad range of bias in interviewers' decisions, of which typical examples include primacy and recency effect (i.e., information received either earlier or later in the interview process is weighted more significantly by interviewers when making decisions; Lunenburg, 2010; Strawn & Thorsteinson, 2015), halo effect (i.e., one favourable characteristic of the applicant can significantly raise the evaluations on the other features of the applicant; Parsons & Liden, 1984; Dougherty et al., 1986; Zysberg & Nevo, 2004) and contrast effect (i.e., interviewers' evaluation of an applicant can be influenced by the quality of the preceding applicant; Schuh, 1978; Cesare et al., 1988).

The existent studies not only expand the knowledge of IDM but also facilitate the improvement of real-world interviews through various approaches, such as structuring the components of interview procedure (Campion et al., 1997; Chapman & Zweig, 2005; McCarthy et al., 2010; Huffcutt et al., 2013; Kausel et al., 2016) and developing decision aids like different types of rating scales (Fay & Latham, 1982; Green et al., 1993; Klehe et al., 2008; Melchers et al., 2011; Cambon & Steiner, 2015). However, there are still a lot of aspects of IDM that remain to be explored.

1.2 State of the Problem

In spite of the rich findings regarding IDM process, it can be perceived to a certain degree that most studies rely their conclusions basically on controlled trials to examine the mathematical correlations between the decisions made by interviewers and the factors investigated (e.g., cognitive, social, situational), whereas the information uncovered about the ongoing process of IDM process is quite limited. However, a review of process research on IDM is absent and even the boundary of the target research is not clear due to a lack of definition. Therefore, this research provides a definition of process research on IDM, based on which a systematic literature review is conducted to clarify the current state of studies that explore the ongoing facts during the process of IDM.

The review findings suggest that, while IDM is a typical cognitive task where a series of information-processing activities taking place in interviewers' minds, only a mere number of studies investigate interviewers' cognitive process (ICP) during decision making. Specifically, as a vital element of human decision making (Newell & Simon, 1972; Ungson et al., 1981; Corner at al., 1994; Leonard et al., 1999; Azuma et al., 2006; Wang & Ruhe, 2007), the investigation into ICP may uncover in-depth details of IDM, for example, the information interviewers search and use, the evaluation criteria applied, and the way interviewers compare alternatives, which can assist the improvement of IDM. Nevertheless, a severe lack of corresponding research is identified from

four perspectives.

First, the research exploring the ongoing facts during IDM only accounts for small proportion compared to the broad interests in IDM, and the overall number of empirical studies investigating ICP is even smaller. Secondly, among the limited empirical studies on ICP, most conclusions are derived from interpretations of indirect evidence, such as the mathematical associations among different elements involved in IDM and the segmented judgments made by interviewers that contribute to their final decisions, whereas less attention is paid to interviewers' actual thinking process. For example, the findings concerning the impact of interviewers' initial impression about an applicant on the hiring decision are usually generated by comparing the numerical scores or ranks that interviewers give to the applicant at different stages of interview (Barrick et al., 2012; Swider et al., 2016; Carnes et al., 2019). Thirdly, even for those studies aiming to directly examine interviewers' thoughts during decision making, their data collection is completed basically through retrospective methods (e.g., self-report, survey) that ask interviewers to recall own thinking processes of decision making after completing an interview task (Graves & Karren, 1992; Hebl & Kleck, 2002; Podsakoff et al., 2011; Wilhelmy et al., 2016). This technique has been criticised as being vulnerable to subjects' self-interpretation, memory loss, and memory distortion, which may lead to deviated understanding of the subject's real cognitive process (Nisbett & Wilson, 1977; Ericsson & Simon, 1984; Russo et al., 1989; Neuman & Schwarz, 1998; Nielsen et al., 2002). Fourthly, most of the existent studies on ICP concentrate on particular aspects of the decision-making task (e.g., information gathering, attributing, cue weighting, specific bias), whereas an empirical study exploring the overall prospect as well as the characteristics of ICP is absent.

Therefore, a research gap is identified in IDM research that there is a severe lack of empirical studies on ICP during decision making especially based on the data that can effectively represent interviewers' real thinking processes. Such a gap not only limits the understanding of IDM, but it also prevents stakeholders from knowing and monitoring interviewers' performance during interview processes at a more in-depth

level. For instance, the evaluation of IDM has long been limited to the psychometric properties (e.g., reliability, validity, and applicants' reactions) of the decisions made, whereas little attention is paid to interviewers' performance during decision-making process. Particularly, it is unclear whether interviewers implement the framed interview process as expected (Van der Zee et al., 2002; Lievens & De Paepe, 2004; Chen et al., 2008; Highhouse, 2008; Nolan & Highhouse, 2014), like whether they follow the given criteria or individual preferences when evaluating an applicant and whether they make judgements based on particular evidence or their own intuitions.

1.3 Research Objectives and Questions

Without a deeper understanding of ICP that is essential but under-investigated in IDM, especially based on the data that can effectively represent interviewers' real thinking processes, further improvements to interview processes and outcomes are limited. Hence, this research aims to explore ICP during decision making and, more importantly, how new knowledge derived from ICP can facilitate the understanding and monitoring of IDM process. Correspondingly, four sets of research questions (RQs) are developed to address the knowledge gap.

Above all, since a subject's cognitive process during decision making consists of a series of information-processing activities taking place in the subject's minds (Payne, 1994; Wang & Ruhe, 2007; Lunenburg, 2010; Mihaylov, 2019), the most primary interest of the investigations into human cognitive processes lies in its constitution, such as the information processed by the decision maker and how the information is processed (Ericsson & Simon, 1984; Martin & Klimoski, 1990; Barber & Roehling, 1993). Correspondingly, the first set of questions brought up in this research intend to find out

***RQ1a**: What information do interviewers pay attention to during decision making (i.e., information type)?*

***RQ1b**: How do interviewers process the information received (i.e., cognitive action)?*

While the first set of research questions focus on generally identifying the possible elements occurring in ICP, the second sets of research questions concern more about the characteristics of ICP, with which the ICP can be described, quantified and compared among both individuals and different contexts. For instance, the decision time of interviewers is examined to investigate how factors like applicant quality, interviewer experience level, and question consistency when evaluating different applicants affects the time interviewers take to make decisions (Tullar et al., 1979; Frieder et al., 2016). Considering the possibility that observed decision time may be affected by irrelevant factors (e.g., mental distractions) especially when the data collection process is not strictly monitored and controlled, this research intends to define a characteristic of ICP that measures the level of cognitive efforts made by interviewers during decision making. Specifically, the first characteristic of ICP defined aims to measure

***RQ2a**: How cognitively active is an interviewer during decision making when evaluating an applicant (i.e., cognitive activeness)?*

In addition, due to the individual differences exist in the backgrounds of interviewers (e.g., experience, knowledge, preference), each interviewer may not adopt all the categories of elements identified when seeking the answers to the first set of research questions. Therefore, the second characteristic of ICP to discuss in this research is

***RQ2b**: How cognitively diversified is an interviewer during decision making when evaluating an applicant (i.e., cognitive diversity)?*

Following existent studies on cognitive processes (e.g., Kivetz & Simonson, 2000; Cumming et al., 2002; Jaspers et al., 2004; Barkaoui, 2011), this research also examines the frequency at which interviewers pay attention to particular types of information and adopt certain cognitive actions to process information:

***RQ2c**: At what frequency does an interviewer use different categories of cognitive*

actions to process information when evaluating an applicant (i.e., frequency of cognitive action)?

***RQ2d**: At what frequency does an interviewer focus on different types of information when evaluating an applicant (i.e., evidential focus)?*

The third set of research questions is raised in order to monitor interviewers' performance during decision making at a cognitive level, particularly their judging behaviour. Since the primary responsibility of interviewers is to assess whether an applicant meets the requirements of the target job position, plenty of research concentrates on how IDM can be improved, among which structured interview is probably the most impactful one and has been widely utilised in real-world practice (Campion et al., 1997; Chapman & Zweig, 2005; McCarthy et al., 2010; Huffcutt et al., 2013; Kausel et al., 2016; Heimann et al., 2021). Specifically, by structuring the components of interview process though providing evaluation criteria and anchored rating scales (Campion et al., 1997, Levashina et al., 2014; Debnath et al., 2015), interviewers' judging behaviour is expected to be "structured" as well so that bias caused by factors like individual preferences, discrimination and decision patterns can be minimised. Regardless of the benefits of structured interview, however, it is found that not all practitioners are willing to follow a designed interview procedure (Chen et al., 2008; Highhouse, 2008; Nolan & Highhouse, 2014). Therefore, this research attempts to develop an indicator at a cognitive level to measure:

***RQ3a**: To what extent does an interviewer follow the structured criteria given when evaluating an applicant (i.e., criteria-compliance level)?*

A common concern about human decision making is whether a decision is rationally made based on evidence or derived from the decision makers' intuition (Sinclair & Ashkanasy, 2005; Dane & Pratt, 2007; Salas et al., 2010). Correspondingly, studies on IDM also concerns whether interviewers' judgments are affected by intuitive information like initial impression about an applicant (Barrick et al., 2012; Swider et al., 2016; Florea et al., 2018; Buijsrogge et al., 2021), and interviewers are often instructed

to take notes which is believed encouraging interviewers to justify own judgements with evidence (Campion et al., 1997). However, interviewers' justification of own judgments and decisions are almost invisible in real-world interviews because they are less likely to express themselves during interview process. While some studies observe and investigate interviewers' notetaking and treat it as a reflection of how interviewers defend own judgments (Macan & Dipboye, 1994; Burnett et al., 1998; Middendorf & Macan, 2002), the fragmented notes are incomplete compared to interviewers' real thinking process. In contrast, this research exploring ICP during decision making enables a direct examination of interviewers' judging behaviour to find out:

***RQ3b**: To what extent does an interviewer tend to justify own judgements (i.e., evidence-based level)?*

Another major structured component of interview is the type of interview questions, among which past behaviour questions that ask applicants to describe what they have done in past jobs (Motowidlo et al., 1992; Green et al., 1993; Pulakos & Schmitt, 1995, Taylor & Small, 2002). Past behaviour questions become one of the most widely utilised question types because it is believed that the best predictor of future behaviour is past behaviour (Barrick et al., 2000; Huffcutt et al., 2001; Moscoso & Salgado, 2001; Bangerter et al., 2014). Correspondingly, interviewers using past behaviour questions are required to base their judgements on the behavioural clues of applicants past work performance as extracted from their responses to interview questions. It is less likely to figure out what information is used by interviewers to support own judgments without a close look at interviewers' thinking processes during decision making, thus this research investigating ICP provides an opportunity to examine:

***RQ3c**: To what extent does an interviewer make judgements based on the applicant's behavioural clues as required by past behaviour questions (i.e., behavioural-based level)?*

In addition to the three dimensions constructed to measure interviewers' judging behaviour, this research also attempts to detect cognitive evidence from ICP that ver-

ifies the existence of typical decision patterns and bias recognised in IDM (e.g., cue weighting, detection of deception, halo effect, contrast effect, similar-to-me effect). In particular, the studies on these decision patterns and bias rely significantly on comparing the decisions made by interviewers in controlled trials while direct evidence is almost absent. To facilitate the discussion and monitoring of the decision patterns and bias in IDM, ICP will be explored in detail to find out

***RQ3d**: Can any evidence be identified from ICP supporting the existence of typical decision patterns and bias?*

Lastly, it is noticed in many studies that the decisions made by interviewers may be affected by various contextual factors, such as interviewers' gender (Parsons & Liden, 1984; Raza & Carpenter, 1987; Chapman & Rowe, 2001), experience level as interviewer (Hess, 2013; Roulin et al., 2015; Frieder et al., 2016), and type of rating scale used during interview process (Fay & Latham, 1982; Maurer, 2002; Klehe et al., 2008; Melchers et al., 2011). Notably, corresponding conclusions are basically based on the examination of mathematical associations between the decisions made and the factors investigated without looking into ICP, thus little is known about how such influences take place by affecting interviewers' thinking processes during interviews. Hence, with the knowledge of ICP generated in the previous research questions, the impact of contextual factors on IDM in terms of the characteristics of ICP during decision making and the three cognitive dimensions of judging behaviour constructed will be examined. Specifically, in addition to the three factors mentioned (i.e., gender, experience level, rating scale type), the national region of previous interview experience, although having not been introduced into employment interview research yet, is taken into the analysis as well considering the possible impact caused by different regional cultures (Sanchez-Burks et al., 2006; Manroop et al., 2013) and employment laws (Williamson et al., 1997; Morgeson et al., 2008) on IDM. Hence, the last research question queries about

***RQ4**: Does (a) gender, (b) national region of interview experience, (c) experience*

level as interviewer, and (d) rating scale type lead to significant difference in the characteristics of ICP and the cognitive dimensions of interviewers' judging behaviour?

1.4 Methodology and Expected Outcomes

A psychological method called think aloud method (TAM) is applied to collect data of ICP, which is introduced by Watson (1920) and Duncker and Lees (1945) and standardised by Ericsson and Simon (1984). TAM requires a subject to keep talking aloud own concurrent thoughts while working on the task of interest and the verbal protocols are explicitly recorded for analysis, which are useful to translate the ongoing thinking process taking place in the subject's short-term memory into the most direct evidence of the subject's cognitive process (Van Someren et al., 1994). The technique is regarded as superior to other approaches in human cognition research, like introspection (Ericsson & Crutcher, 1991; Tordesillas & Chaiken, 1999), retrospection (Kuusela & Paul, 2000; Gero & Tang, 2001; Van Gog et al., 2005), questioning and prompting (Xun & Land, 2004; Ge et al., 2005) and dialogue observation (Cox et al., 1999; Chi et al., 2008), because the data collected with TAM provides the most direct evidence of a subject's cognitive process free from the negative influences caused by disturbance of process, memory errors, and the subject's own interpretations (Davison et al., 1997; Nielsen et al., 2002; Heerkens & Van Der Heijden, 2005; Bannert & Mengelkamp, 2008; Eccles & Arsal, 2017).

TAM is widely used in various domains to explore subjects' cognitive processes when working on the task of interest, such as clinical decision making (Funkesson et al., 2007; Forsberg et al., 2014), consumers' decision making (Dhar & Simonson, 1992; Reicks et al., 2003), design (Lloyd et al., 1995; Perry & Krippendorff, 2013), utility testing (Roberts & Fels, 2006; Olmsted-Hawala et al., 2010), language learning (Bernardini, 2001; Li, 2004) and essay rating (Wolfe et al., 1998; Cumming et al., 2002; Han, 2017). The technique has also been utilised in recruitment research, for example, to study how applicants make decisions about whether or not to apply for a job (Barber

& Roehling, 1993), but it has not yet been adopted to explore ICP when making hiring decisions.

Due to the requirements of TAM on both the features of the task investigated and the qualification of subjects, an experiment procedure is developed and tested through a pilot study first. After confirming the feasibility and validity of embedding TAM in interview scenarios, the experiment procedure is then carried out in for data collection, where 29 participants with diversified backgrounds are involved as interviewers. The verbal protocols that represent interviewers' actual thinking processes when making hiring decisions are recorded and processed following a standard procedure (Fonteyn et al., 1993; Van Someren et al., 1994; Charters, 2003; Eccles and Arsal, 2017), from which the elements of ICP like information type and cognitive actions are identified and defined (RQ1a-b). Afterwards, the elements identified are investigated to develop quantitative measurements for both the four ICP characteristics proposed (RQ2a-d) and the three dimensions of interviewers' judging behaviour (RQ3a-c). Then the protocols are scrutinised to detect cognitive evidence for a wide range of decision patterns and bias commonly recognised in IDM (RQ3d). As the final step of data analysis, the association between the four contextual factors (i.e., gender, national region of interview experience, experience level, rating scale type) and the ICP characteristics as well as interviewers' judging behaviour are examined to investigate the potential impact of these factors on IDM at a cognitive level (RQ4).

1.5 Significance and Potential Limitations

The findings generated in this research can both contribute to the academic knowledge of IDM and be used to facilitate the real-world practice from several perspectives. The contributions to academic knowledge lie in four aspects. First, this research becomes the first to provide an overall prospect of interviewers' thinking processes during decision making with the information types and cognitive actions involved in ICP unveiled. Secondly, the findings regarding the characteristics of ICP, the heatedly

discussed dimensions of interviewers' judging behaviour, and the cognitive evidence detected for interviewers' typical decision patterns and bias not only enrich the knowledge of IDM themselves but can also be introduced as variables into future research on IDM. Meanwhile, the associations between several contextual factors and IDM are examined which, if identified, can unveil the mechanism to a certain degree of how a certain factor affects interviewers' decisions by altering the ICP involved. Last but not the least, as a pioneering work that introduces TAM to the investigation into IDM, this research not only confirms its feasibility but also proposes a detailed procedure of experiment development and validation that can be transferred to different interview contexts of interest (e.g., different industries, job positions, interview types).

On the part of practical implications, the findings generated in this research can be expected to assist the improvement of IDM in real-world practice from two major perspectives. On one hand, the knowledge of ICP can promote the development of decision aids based on artificial intelligence that can assist IDM and thus enhance the efficiency and fairness of recruitment. On the other hand, the findings can provide guidance for the improvement of interview settings as well as the design of training courses for interviewers.

Apart from the rich findings and contributions, two limitations of this research are pointed out as well. One is the small sample size due to the nature of TAM which aims for in-depth investigation into human cognitive process of individuals rather than large groups of people. Hence, while the sample size in this research is sufficient for the qualitative exploration of ICP, it may constrain the generalisation of the statistical findings concerning the associations between the contextual factors and ICP. The other concern is about the single source of data (i.e., verbal protocols) and the fixed interview scenario adopted in the experiment (e.g., the job position, questions and responses, and evaluation criteria). While the data analysis is conducted at a level that is less vulnerable to these differences in experiment design, the trustworthiness of the findings can be further enhanced if tested in different contexts or double checked through other methods (e.g., survey, behavioural observations).

1.6 Structural Outline of the Dissertation

The dissertation unfolds sequentially throughout five chapters. **Chapter 1** has introduced the broader context of the whole research and the problem identified, followed by the objectives of this research and the questions to solve. Then the overall methodology and expected outcomes have been stated. The significance and the potential limitations of the research have also been discussed.

Chapter 2 of literature review will first present the background of the academic interests in IDM process, especially those uncovering the ongoing facts during IDM process. Afterwards, a systematic literature review will be conducted to clarify the current state of process research on IDM, from which a gap will be identified.

To cover the gap identified in the previous chapter, **Chapter 3** will justify the adoption of a qualitative research approach (i.e., TAM) from three perspective. First, the history and mechanism of TAM will be introduced, followed by a comparison among the technique and other common methods used in human cognition research. Then, the feasibility of applying TAM to explore ICP will be discussed, based on which an experiment procedure will be developed, tested and revised. The research design in terms of data collection and data analysis will be reported in detail with rationale explained, and the trustworthiness of the research will be argued.

In **Chapter 4**, the key findings generated through data collection and analysis procedure introduced in Chapter 3 will be presented in accordance with the research questions proposed, of which a more in-depth discussion will be provided in **Chapter 5** within a broader context of the topic. In addition, the contributions to knowledge and practical implications of the research will be discussed as well in Chapter 5 with the limitations and the opportunities of future research identified.

Chapter 2

Literature Review

Driven by the need to improve the quality of interviewers' decisions, a great deal of studies exploring IDM have been carried out throughout the last ten decades. In particular, there has been a long-standing call for more studies on how interviewers' decisions are made from a processive perspective (i.e., process research on IDM) rather than focus solely on the psychometric properties of interviewers' final decisions. In spite of the fruitful studies accumulated, a comprehensive and targeted review is yet to be carried out so as to clarify the current status and future prospects of process research on IDM.

This chapter of literate review consists of four sections. **Section 2.1** briefs the background of the academic interests in IDM and the evolution of process research that uncovers the ongoing facts during the IDM process. It points out in particular that an overall image of the process research on IDM is lacking regarding both the contributions and the features of existent studies. As a response to this situation, a systematic literature review (SLR) is conducted in **Section 2.2** that presents the status of process research on IDM from multiple aspects, including thematic focus and key outcomes, background information, as well as methodological features. Based on these review findings, a significant gap is recognised and stated in **Section 2.3** that, while decision making is a typical cognitive task where a subject's thinking process (i.e., cognitive

process) plays a vital role, there is a severe lack of investigation into interviewers' cognitive process (ICP) that lead to their final decisions. Finally in **Section 2.4**, a summary of the whole chapter is presented.

2.1 Background and Overview

2.1.1 Employment Interview and Interviewer as the Decision Maker

Employment interview has been one of the most commonly used selection tools for decades (Ulrich & Trumbo, 1965; Harris, 1989; Buckley et al., 2000; Wanberg et al., 2020; Chauhan, 2022). It is not only a preferred selection method among organisational decision makers, like supervisors and human resource practitioners, but it has also been perceived by applicants as an inevitable part of personnel selection or even fundamental process of job search success (e.g., Lievens et al., 2003, 2005; Topor et al., 2007; Anderson et al., 2010).

The prevalence of employment interview in real-world practice has also impelled great academic interests in the last 100 years. However, it seems that such popularity has made employment interview a tacit thing among scholars and they may not consistently use the exact term "employment interview" but adopt several substitutes to refer to the same thing, like "interview" "job interview" "recruitment interview" and "selection interview". The definition of employment interview, if not absent, also varies across different studies. Several example definitions of employment interview are listed in Table 2.1. It can be seen that some of the definitions emphasise the nature of employment interview as a social interaction between two parties and thus state not only the activities of interviewers but also that of applicants during interview process (Dipboye, 2005; Macan, 2009). Among the definitions of this kind, the most widely accepted one is proposed by Dipboye (2005), who defines it as a two-way interaction between interviewer and applicant where the both parties exchange information and make decisions. A corresponding procedural model is developed by Dipboye as well to

Table 2.1: Example definitions of employment interview

Literature	Definition of Employment Interview
McDaniel et al. (1994)	… a selection procedure designed to predict future job performance on the basis of applicants' oral responses to oral inquiries
Judge et al. (2000)	… a selection tool where interviewers gather and evaluate information about applicants.
Huffcutt & Youngcourt (2002)	… a face-to-face interaction conducted to determine the qualifications of a given individual for a particular open position.
Posthuma et al. (2002)	Interviews are designed, in part, to gather information about job applicants so judgments about future work performance can be made.
Dipboye (2005)	… a two-way interaction between interviewer and applicant where the both parties exchange information and make decision.
Macan (2009)	The interview is a method of assessment … during which interviewers assess applicant traits or constructs. The employment interview is a social interaction where the interviewer and applicant exchange and process the information gathered from each other.
Levashina et al. (2014)	… a personally interactive process of one or more people asking questions orally to another person and evaluating the answers for the purpose of determining the qualifications of that person in order to make employment decisions.

illustrate the core interview process embedded in multiple levels of environments (see Figure 2.1).

On the contrary, other definitions of employment interview focus more on the role of interviewers as the decision maker during this process (Judge et al., 2000; Huffcutt & Youngcourt, 2007; Posthuma et al., 2002; Levashina et al., 2014). According to its definition in psychological and cognitive science, decision making is a cognitive process composed of a logical sequence of information-processing activities taking place in the decision maker's mind (Payne, 1994, Wang & Ruhe, 2007; Lunenburg, 2010; Mihaylov, 2019). Regardless of the literal discrepancies in these definitions of employment interview, several terms can be repeatedly recognised pointing out these information-

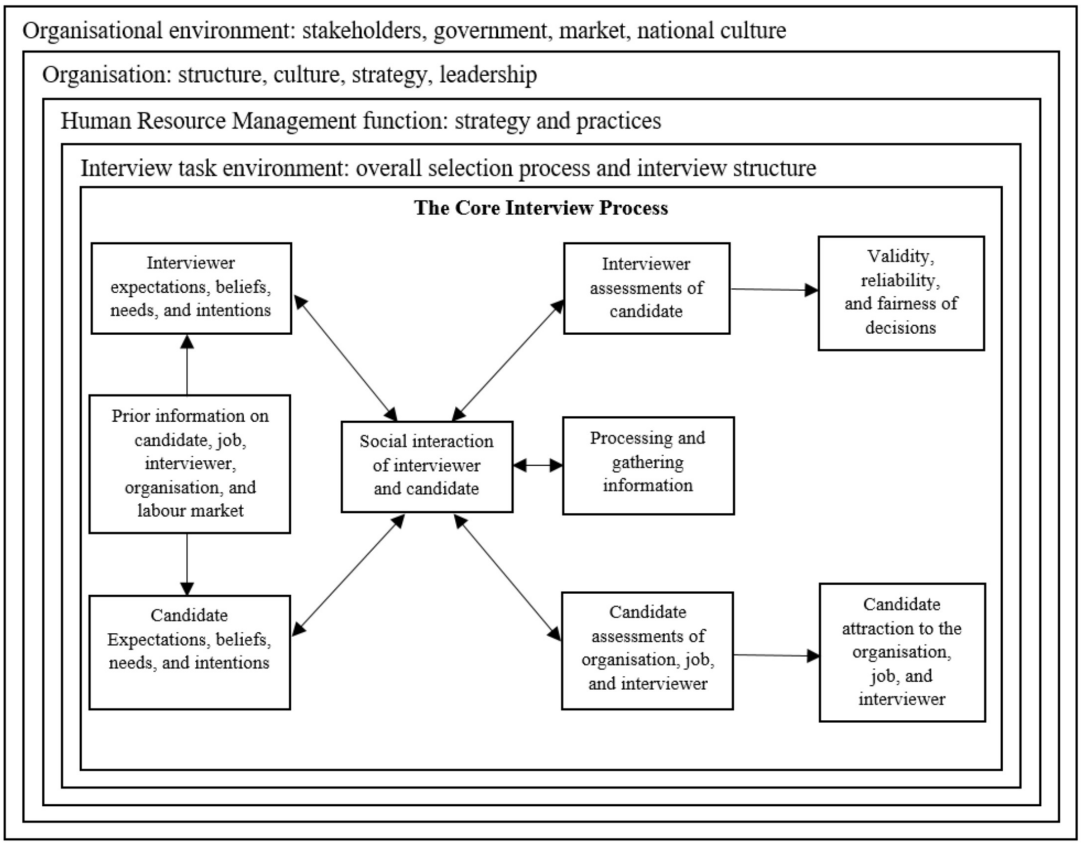


Figure 2.1: Employment interview as a two-way interaction (Dipboye, 2005)

processing activities embedded in interviewers' task, such as gathering information (from the applicant), assessing (the applicant's qualification), and making (hiring) decisions. A more integrated narrative of interviewers' performance as a decision maker is provided by Dipboye (1992) from an information-processing perspective that

> *. . . as the interview unfolds, an interviewer selectively attends to information that attracts attention. The interviewer then attempts to explain what is observed by attributing applicant behaviours to underlying traits. This information is then encoded and stored in memory. At some later point, the interviewer must retrieve the information from memory and translate it into a judgment.*

2.1.2 Evolution of Research on IDM

Motivated by the need to select and hire talents for organisations, a great deal of academic efforts and attempts have been made to improve interviewers' decisions. The empirical research on IDM could date back to the study conducted by Scott (1915) where six interviewers interviewed and rated 36 applicants for sales positions. Scott compared the ratings given by different interviewers while little relationship was found. Similar findings were also generated in a series of studies later on (Hollingworth, 1922; Snow, 1924; Corey, 1933), which aroused doubts in the validity and reliability of employment interview as well as efforts to improve it. However, most of the earliest discussions either took a "how-to-do-it" formula based on experience and common sense that provided guidance with little empirical evidence or focused solely on the outcomes of employment interview without investigating the process (Webster & Anderson, 1964; Mayfield, 1964; Ulrich & Trumbo, 1965).

Later in 1950s, a series of empirical studies on IDM (Springbett, 1958; Anderson, 1960; Sydiaha, 1959, 1961, 1962; Rowe, 1963) were undertaken at McGill University and reported by Webster and Anderson (1964). Rather than simply compare the interview outcomes, these studies became the initial attempts to uncover the procedure of how interviewer reached their decisions from the information gathered. Such work was regarded as a beginning step toward a better understanding of the process of IDM (Mayfield, 1964) that affected the subsequent research on the topic from two perspectives. On one hand, academic interests in the process of IDM were stimulated and scholars started to realise that past research focusing the outcome of employment interview failed to explain the mechanisms of IDM and provided limited guidance of improvement (Schmitt, 1976; Lewis, 1980; Arvey & Campion, 1982; Tullar, 1989). Correspondingly, investigations were encouraged to explore what led to the discrepancies in interviewers' final decisions. On the other hand, Webster and Anderson also set an example of how such kind of empirical studies could be conducted, following which massive studies on IDM were brought up to identify the factors that affected interviewers' decisions, such

as applicants' individual differences (e.g., demographics, physical appearance, disability, obesity, and stigma), interviewers' individual differences (e.g., gender, experience, personality, stereotype), social factors (e.g., applicants' verbal and nonverbal cues, impression management tactics, interviewer-applicant similarity), and situational factors (e.g., interview structure, technology). The fruitful findings could be found in massive reviews and meta-analyses (see for example, Wright et al., 1989; Posthuma et al., 2002; Macan, 2009; Levashina et al., 2014).

Despite of the fruitful findings regarding the influential factors in interviewers' decisions, increasingly more scholars noticed that the research focus was pretty much on those superficial and observable factors and the conclusions were mainly based on simple bivariate relationships. In particular, while IDM was "a complex multifaceted process with underlying psychological determinants" (Posthuma et al., 2002), little was known about the psychological mechanisms of how these factors led to different discrepancies in interviewers' decisions (Arvey, 1979; Hogarth & Einhorn, 1992; Dipboye, 2005; Buijsrogge et al., 2016). Thus, research was urged to promote the understanding of the real and actual process of IDM (i.e., process research), especially from a cognitive perspective.

Recalling that the IDM process involves a sequence of information-processing activities where interviewers act as an active seeker, receiver, and processor of information (Dipboye, 1982; Dreher & Sackett, 1983), massive empirical studies were conducted to observe and analyse interviewers' behaviour when performing these activities. Some information-processing behaviours of interviewers are observable, like questioning (Dougherty et al., 1994; Silvester & Anderson, 2003; Highhouse et al., 2019) and notetaking (Burnett et al., 1998; Biesanz et al., 1999; Middendorf & Macan, 2002). In contrast, investigations were also conducted interviewers' decision-making "behaviours" taking place inside of their minds. For instance, interviewers' attributional styles were examined regarding how they interpreted the clues obtained from applicants, such as applicants' past and interview performance (Tucker & Rowe, 1979; Gifford et al., 1985; Chapman & Webster, 2001) and their impression management tactics (Schmid Mast

et al., 2011; Roulin et al., 2015; Culbertson et al., 2017). Meanwhile, the majority of process research on IDM focused on how interviewers' final decisions were made from various perspectives, for instance, how interviewers weighed different cues (Dougherty et al., 1986; Motowidlo & Burnett, 1995; DeGroot & Gooty, 2009) and what factors (e.g., dimensional evaluations, perceived person-organisation fit, interviewers' liking) mediated the formation of the final decisions (Rynes & Gerhart, 1990; Cable & Judge, 1997; Hebl & Kleck, 2002; Barrick et al., 2010; Madera & Hebl, 2012).

Another heated topic was the recognition of a wide range of bias existing in IDM. For example, confirmatory bias was found in interviewers' questioning strategy that interviewers tended to seek for information that confirm their initial impression of the applicant (McDonald & Hakel, 1985; Binning et al., 1988; Florea et al., 2018); primacy and recency effect was noticed that information received either earlier or later in the interview process was weighted more significantly by interviewers when making decisions (Farr, 1973; Lunenburg, 2010; Strawn & Thorsteinson, 2015); halo effect was identified in interviewers' decisions in the sense that one favourable characteristic of the applicant could significantly raise the evaluations on the other features of the applicant (Hakel, 1971; Parsons & Liden, 1984; Dougherty et al., 1986; Zysberg & Nevo, 2004); and it was also testified that interviewers' evaluation of an applicant can be influenced by the quality of the preceding applicant, which was termed as contrast effect (Schuh, 1978; Cesare et al., 1988). There were also a few other aspects of IDM covered, like interviewers' confidence in own decision (Young & Kacmar, 1998; Dipboye & Jackson, 1999; Kausel et al., 2016; Carnes et al., 2019), awareness of own decision-making process (Valenzi & Andrews, 1973; Zedeck & Kafry, 1977; Stumpf & London, 1981; Graves & Karren, 1992), and memory accuracy after completing the interview process (Johns, 1975; Baron, 1986, 1987; Middendorf & Macan, 2002). However, the knowledge contributed to these topics was quite limited with only a small number of studies found.

All these studies uncover the process of IDM to a certain degree, whereas a review that comprehensively reports the contributions and features of the existent studies

is absent. Particularly, while plenty of impactful reviews on IDM can be identified (Schmitt, 1976; Arvey & Campion, 1982; McDaniel et al., 1994; Campion et al., 1997; Posthuma et al., 2002 Levashina et al., 2014), most of them are not constrained to process research and only a limited number of reviews are found concentrating on specific perspectives of IDM process. For instance, Huffcutt (2011) reviews the empirical studies on the constructs evaluated by interviewers and develops a model that specifies three main sources from which interview those constructs are derived (i.e., job-related interview content, interviewee performance, personal and demographic characteristics). As a matter of fact, even a clear definition of process research is still absent. In order to present the current status of process research on IDM and to provide insights into the opportunities of future research, it is necessary to set the boundary of process research on IDM and conduct a comprehensive and targeted review.

2.2 Systematic Review of Process Research on IDM

A SLR was conducted in a comprehensive, transparent, and replicable way to obtain a critical view of process research on IDM. This section first inclusively defines the boundary of process research on IDM, based on which a thorough article searching and screening procedure is carried out. The articles selected are then critically reviewed and analysed from three perspectives, namely the thematic, background, and methodological features, which not only present a multi-dimensional picture of the research on IDM process but also indicate the possible directions for future research.

2.2.1 Definition of Process Research

The idea that decision making is a sequence of mental activities taking place between the presentation of a stimulus and the execution of a response stimulates the development of "process research" into decision making (Crozier & Ranyard, 2002) as well as corresponding "process approaches" (e.g., eye-movements tracking, information

board technique, and think aloud protocols) with which researchers follow and draw conclusions about the psychological process during decision making (Svenson, 1996). Similarly in IDM research, Arvey and Campion (1982) call for process research into the causes of differences in interviewers' evaluations and suggest that "little is known about why differential evaluations are made and what goes on in the interview to influence the evaluations ... more research concerning the perceptual processes which might account for the phenomenon observed." However, in spite of the long-lasting interest in IDM process and the abundant studies accumulated, a clear definition of the so-called "process research" on IDM is yet to be proposed, without which it is less likely to carry out a SLR on it.

In order to check the features of research that uncovers the IDM process, the author scans massive empirical studies for their research focus and methodological characteristics. It is noticed that both the types of data collected and the ways in which variables are correlated to interviewers' decisions can be labelled as either processive or non-processive regarding whether it uncovers (at least part of) the actual facts of interview process. As for the types of data collected, variables straightforwardly representing interviewers' ongoing performance within interview scenarios are defined as **processive variables** in this research, such interviewers' questioning, note taking, cue weighting, and decision time. On the contrary, variables that can hardly be interpreted alone to uncover the process of interview are defined as **non-processive variables**, such as individual differences (e.g., age, gender, physical appearance, experience, personality) and interview settings (e.g., question types, criteria of evaluation, because interviewers may not strictly follow the designed interview settings when carrying out the interview process).

The studies on IDM also differ in the way they establish correlations between certain factors and interviewers' decisions. As previously mentioned, a lot of studies derive their conclusions from simple bivariate correlations between variables and the final decisions while telling little about how such correlations occur along the IDM process. In this case, the correlation development is defined as **non-processive correlating**.

Instead, some articles elaborate the impact a factor causes on interviewers' decisions in a more processive way (i.e., **processive correlating**), for instance, by introducing intermediate variables to the models that illustrate interviewers' decision path or referring to psychological theories (e.g., attributional theory, information-processing theory) to explain how such correlations are formed.

In order to focus the scope of review on the studies that aim to uncover the process of IDM, the SLR in this research defines process research as the empirical studies that either contain at least one processive variable or involve processive correlating when explaining how interviewers' decisions are formed. Such definition of process research serves as the boundary of article selection in the following SLR.

2.2.2 Review Methodology

The SLR is carried out following the three phases show in Figure 2.2 to identify and review the target studies (i.e., process research on IDM), including (i) journal selection, (ii) a two-step article selection, and (iii)) information collection and research synthesis.

Phase 1: Journal Selection

Most literature reviews of the research on employment interview cover the studies either on a specific topic (Campion et al., 1997; Huffcutt, 2011; Levashina et al., 2014) or within a relatively short time span as a continuation of previous reviews (Harris, 1989; Judge et al., 2000; Macan, 2009), which allow them to dive directly into article searching on databases (e.g., ABI/Inform, EBSCO, Emerald, ScienceDirect, Scopus and Taylor & Francis) with particular criteria. This SLR as the first comprehensive review of the existent process research on IDM, however, starts with journal selection due to the multiple meanings of "interview". Specifically, since the criterion of "process research" cover a broad range of topics and cannot be effectively transferred to any keyword for a purposeful article searching, this review finds it necessary to inclusively locate "interview"-related articles first and then carry out a screen procedure for the

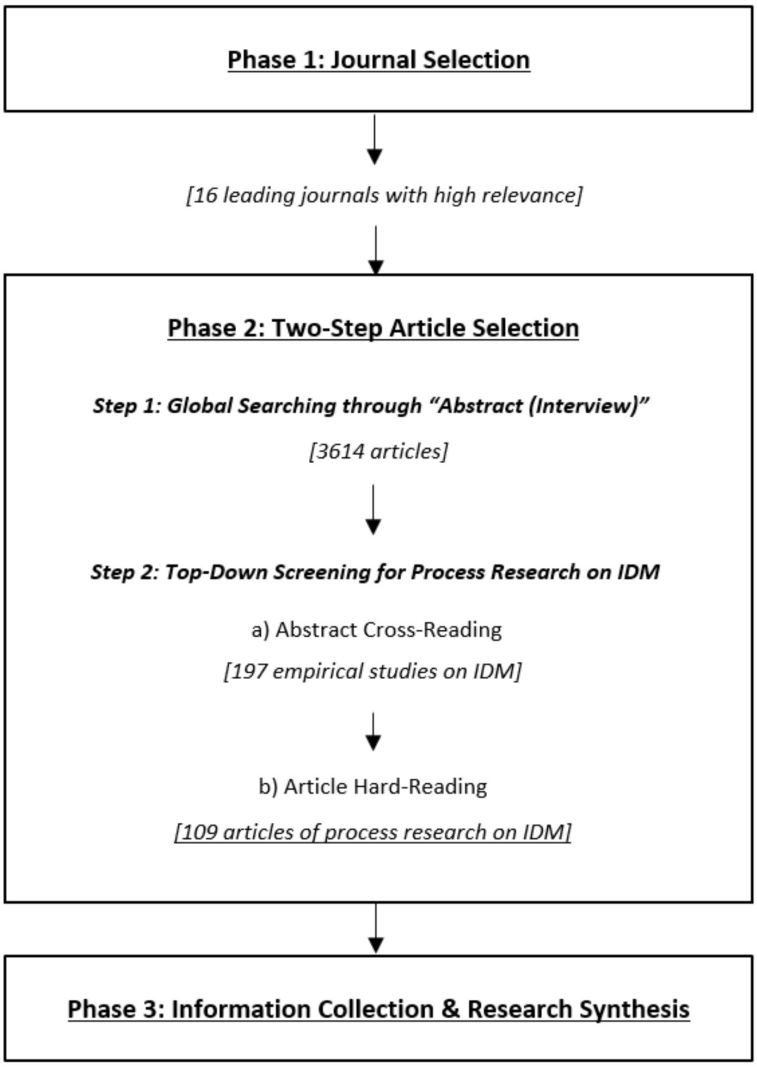


Figure 2.2: Overall procedure of the systematic literature review

target studies. Meanwhile, the term "interview" may refer to not only a personnel selection device applied in organisations but also a communication device widely applied in a variety of scenarios or as an approach of data collection (Fontana & Frey, 2005; Alshenqeeti, 2014). As a result, article searching with the keyword "interview" on any database can lead to massive irrelevant results across various disciplines. Therefore, it is necessary to check the relevance of journals to the subject of employment interview as well as their qualities.

Academic Journal Guide (AJG) rankings, a widely accepted list of journals according to their qualities within different fields (Walker et al., 2019; Bryce et al., 2020), is adopted as the benchmark of journal selection. First, the journal information of a great deal of studies on IDM is collected and then located in AJG rankings, from which it is noticed that these journals mainly belong to two fields as categorised in AJG rankings, namely *Human Resource Management and Employment Studies* and *Organisational Psychology*. Secondly, the scores of the journals listed within these two fields are checked as an indicator of journal quality and only those rated no less than 3 are included, which results in 37 journals in the field of *Human Resource Management and Employment Studies* and 40 journals in *Organisational Psychology*. Noticing that not all these 77 journals are target publishers of studies on employment interview, the relevance of each journal is examined through keyword searching for "interview" in abstract with a quick scan of the articles generated. According to the results, only four journals of human resource management (HRM) and twelve journals of organisational psychology are found relevant where over a thousand studies are identified, while the rest turn out to be irrelevant and are screened out. All these 16 journals are peer-reviewed and written in English, which are selected as the base of the following-up article selection.

Phase 2: A Two-Step Article Selection

The article selection consists of two steps, including a global searching with the function "Abstract(Interview)" and a top-down screening for process research on IDM.

1) Global searching with the function "Abstract (Interview)"

As the first step of article selection, a global searching for interview-related studies is executed through keyword searching of "interview" within the field of abstract. While it could have been much more effort-saving if a conventional searching and screening procedure based on ideal keywords could be applied, the general term "interview" instead of more specific ones (e.g., employment interview, recruitment interview, selec-

tion interview) is searched only in abstracts due to several observable facts that can significantly affect the outcomes of article searching.

First of all, "interview" can be used in other personnel-related stages as well, like employee exit interview (Garretson and Teel, 1982; Kulik et al., 2012) and performance appraisal interview (Sandlund et al., 2011; Pälli & Lehtinen, 2014). Secondly, a considerable number of similar terms can be recognised across different studies as a substitute for "employment interview", such as "recruitment interview" (Ferris et al., 2002; Dipboye, 2017), "selection interview" (Campion et al., 1997; Chapman & Zweig, 2005), or simply "interview" (Phillips & Dipboye, 1989; Hebl & Kleck, 2002). Thus, it is highly likely that a few articles will be omitted if more specific terms are utilised in keyword searching. Thirdly, it is noticed that many studies on employment interview may not necessarily mention the term "interview" in other commonly searched fields, like titles and keywords, while the abstract turns out to be the bottom line.

Taking all these concerns into account, the abstract rather than any other part of an article is selected to carry out the keyword searching, and attempts are also made to double check whether the choice of different "interview"-related terms causes a significant impact on the outcomes of article searching. To be specific, all the substitute terms of "employment interview" identified are tried out by searching each of them within the field of abstract. The results indicate that searching with "interview" leads to a much wider range of articles and covers all the outcomes generated by any of the alternative terms. Thus, it seems not only reasonable but also necessary to adopt this most general keyword to ensure that there is no significant omission of studies on experiment interview (i.e., inclusiveness). Overall, the article selection starts with a global searching for the research on employment interview by executing the function "Abstract(Interview)" in each of the sixteen journals, which results in 3614 articles.

2) Top-down screening for process research on IDM

The global article selection in the previous step enhances the inclusiveness of the

searching results, whereas massive irrelevant articles still await to be detected and removed from the pool of articles. Therefore, a top-down screening procedure needs to be carried out to purify the results for process research on IDM, which is achieved sequentially through a) abstract cross-reading to find out the studies on IDM and b) hard reading of full texts to identify process research.

To begin with, cross-reading is carried out for the abstract of each article to check if the article is about individual interviewers' decision making, during which four selection criteria are applied. First, only the articles about employment interview are included, while those studies on appraisal interview, exit interview, or utilising interview as a method for data collection are screened out. Secondly, the articles where interview is used to indicate other stages of selection procedure (e.g., pre-interview screening) are excluded. Thirdly, only the articles about IDM are selected, whereas the investigations into interviewers' attitudes toward interview settings (Nolan & Highhouse, 2014; Tsai et al., 2016) or other perspectives of interviewers' performance are filtered out, such as their interaction with applicants (Stewart et al., 2008; Wilhelmy et al., 2021). Fourthly, since this research concentrates on individual interviewers' performance during decision making, the articles concern about group decision making among panel interviewers are not considered within this review. Only 414 out of the 3614 articles resulted in the previous phase of global searching are related to employment interview, among which 163 articles are less related to IDM, 44 articles are not empirical studies on IDM (e.g., non-empirical research, reviews, meta-analyses), and 10 articles are about group decision making. Therefore, abstract cross-reading results in the inclusion of 197 articles in total.

Afterwards, these 197 articles are further scrutinised through hard reading to identify process research on IDM, where the rule of selection is derived from the definition of process research: It has to be an empirical study that either contains at least one processive variable or involves processive correlating when explaining how interviewers' decisions are formed. During this process, it is noticed that some of 197 articles identified are based on the test of simple bivariate correlations and are thus screened

out, whereas only 109 articles published in 12 journals (one HRM journal and eleven organisational psychology journals) are included in the systematic review. Particularly, due to the absence of literature review of process research on IDM while this research intends to find out the evolution of the academic focus in different research streams concerning IDM, the timeframe is not applied as a criterion of article selection and all the articles identified that meet the definition of process research on IDM are included and reviewed.

3) Information collection and research synthesis

Within this phase, the 109 articles in terms of process research on IDM are examined in detail to capture their descriptive and thematic content. In particular, an information-extraction sheet is established to collect the information of each article from three perspectives, namely

- *Thematic information*, including research stream, research aim, and conclusion(s)
- *Background information*, including journal of publication, year of publication, number of citations, and geographic origin
- *Methodological information*, including data collection method, participant type, source of evidence, and theory/model applied

2.2.3 Review Findings

2.2.3.1 Thematic Features

When examining what has been researched by the articles identified as process research on IDM, four main research streams are generated and defined. The identification of the first three research streams, namely **(S1) information gathering**, **(S2) attributing** and **(S3) IDM pattern**, basically reflects the major activities of IDM

as defined (Judge et al., 2000; Huffcutt & Youngcourt, 2007; Posthuma et al., 2002; Levashina et al., 2014), whereas **(S4) other IDM performance** is a collection of interviewers' performance during decision making that account for a relatively small part of academic interests. In addition, subcategories are defined for both information gathering and IDM pattern, but not for attributing and other IDM performance due to the small numbers of articles found. The description and literature examples of each research (sub-)stream are presented in a classification scheme (see Table 2.2). As an indicator of academic interest, the number of articles is summarised for each (sub-)stream as well.

S1: Information Gathering

Since information forms the basis of decision making, interviewers' decisions are likely to vary when different information are considered and interviewers' (S1) information gathering has drawn the attention from scholars in about one fifth ($21.10\%, N = 23$) of the process research on IDM performance. Specifically, the studies within this research stream investigate how interviewers gather information from applicants during interview and how it can be affected by a wide range of factors (e.g., organisational context, interviewer characteristics, training, applicant background and qualification). With a further inspection of the research focus, the articles can be classified into two groups, including **(S1a) questioning**, which looks into the characteristics and strategies of interviewers' questioning, and **(S1b) information collected** that discusses the features of the information collected by interviewers.

S1a: Questioning

As can be seen in Table 2.2, most articles of S1 present an interest in the characteristics of interviewers' questioning (18 out of 23), of which ten discuss the type of questions asked by interviewers (e.g., behavioural, credential, novel, brainteaser), seven studies comment on various features of interviewers' questioning (e.g., amount, topic coverage, relevance, overall time length), while the rest one study covers the both sub-streams.

Table 2.2: Classification scheme of research streams

Research Stream	Description	No. of Articles (%)	Literature Example
S1: Information Gathering	How interviewers gather information from applicants	**23 (21.10%)**	
a) Questioning	• The characteristics and strategies of interviewers' questioning	*18 (16.51%)*	Janz (1982) Dougherty et al. (1994) Silvester & Anderson (2003) Speer et al. (2022)
b) Information Collected	• The characteristics of the information collected by interviewers	*5 (4.59%)*	Macan & Dipboye (1994) Middendorf & Macan (2002) Nikolaou (2011)
S2: Attributing	How interviewers interpret the information gathered and perceive the applicant	**11 (10.09%)**	Gifford et al. (1985) Chapman & Webster (2001) Roulin et al. (2015)
S3: IDM Pattern	How interviewers score applicants and make hiring decisions	**78 (71.56%)**	
a) Rating Strategy	• How interviewers weigh different cues when making decisions	*34 (31.19%)*	Kinicki et al. (1990) DeGroot & Gooty (2009) Swider et al. (2016) Buijsrogge et al. (2021)
b) Decision Path	• How interviewers' decisions are progressively formed with various intermediates involved	*48 (44.04%)*	Cable & Judge (1997) Kristof-Brown (2000) Madera & Hebl (2012) Florea et al. (2019)
S4: Other IDM Performance	Other aspects of IDM performance not discussed above, like • task perception • decision time • memory • awareness of own decision-making process	**14 (12.84%)**	Baron (1986, 1987) Graves & Karren (1992) Kausel et al. (2016) Buijsrogge et al. (2021)

Note. 1. 17 articles are listed under multiple research (sub-)streams; 2. The percentages are calculated by dividing the number of articles covering each research (sub-)stream by the total number of the articles reviewed ($N = 109$).

As for the type of questions asked by interviewers, the influences of multiple factors on interviewers' choice of questions were tested. Two studies compared the types of questions asked by interviewers when interview settings differed (i.e., interview type,

interview medium). In particular, Janz (1982) investigated the questions asked by interviewers in either patterned behaviour description interview or unstructured interviews. It was observed that interviewers in patterned behaviour description interviews asked more credential questions, while the latter involved significantly more behavioural content. The other study conducted by Silvester and Anderson (2003) explored the potential influence of technology on interviewers' preference for more closed interview questions, while they found no significant tendency for interviewers to ask more closed questions in telephone interviews than in face-to-face interviews. One study was found paying particular attention to the impact of interviewers' personality on their choices of question types (Highhouse et al., 2019), stating that narcissism, sadism and a lack of perspective-taking could explain why some interviewers tend to use brainteasers in an interview. Taking a different perspective, Speer et al. (2022) recognised that interviewers' question preference was related to their perceived familiarity with and job-relevance of the questions.

Apart from these four studies, all the rest six articles on interviewers' question types discuss over a heated topic of interviewers' questioning strategy, namely confirmatory behaviour that describes how interviewers' questioning is affected by their first impressions of the applicant. As the conceptual springboard of confirmatory behaviour stemming from social psychology, Snyder and Swann (1978) found that people tended to gather information that supported (i.e., confirmed) their beliefs or hypotheses about another person. The phenomenon had then been widely discussed in a line of subsequent research across various social situations, among which employment interviewers' performance attracted considerable attention. Mixed results were generated in the sense that confirmatory behaviour was recognised in some of the studies (Binning et al., 1988; Dougherty et al., 1994; Florea et al., 2019) but not in the others (Sackett, 1982; McDonald & Hakel, 1985; Macan & Dipboye, 1988). Nevertheless, it should be noted that these studies vary greatly in research focus and design in spite of their common interest in interviewers' confirmatory behaviour, which could be a possible explanation of the inconsistence in observations.

In addition to the types of questions asked by interviewers, several studies ($N = 8$) are also found investigating the differences in the features of interviewers' questioning, such as topic coverage, quality of question, informational order, and time length of questioning. Specifically, three studies examined the topic coverage of the questions asked by interviewers. Keenan and Wedderburn (1980) found that the questions asked by interviewers covered more topics concerned with future job and knowledge of company than applicants' present or past academic performance, which might because, as interpreted by the authors, "interviewers preferred topic areas where they had an advantage over applicants". In the study conducted by Taylor and Sniezek (1984), differences in interviewers' topic coverage were also reported by applicants that the interview questions dealt more with non-academic issues of university life and extracurricular activities. They also found that individual interviewers showed little agreement on topic coverage and did not cover the topics they regarded as important. Besides, Stevens (1998) noticed that the topic coverage of interviewers' questions could be affected by their perceived screening-recruiting priorities and training received.

Two studies were carried out by Speer and his colleagues where the quality of interview questions was judged by expert interviewers as either effectiveness or superiority (Speer et al., 2019, 2020). The results indicated that interviewers' ability to identify effective or superior interview questions was strongly related to their general mental ability and social intelligence. Other features of questioning receiving scholars' attention include interviewers' choice of information order and the length of time they spend on questioning. Johns (1975) found that interviewers who chose the informational order tended to seek for the most important information early in the sequence and desired to terminate information gathering sooner. Macan and Dipboye (1990), who examined the potential impact of interviewers' pre-interview impressions of applicants on the length of time they spent on questioning, recognised no significant correlation. Besides, Biesanz et al. (1999) checked whether interviewers' notetaking and expectations of applicants would affect the type (i.e., whether the questions were novel or not, the proportion of positive, negative, and neutral questions) and number of the questions asked, while no correlation was verified.

S1b: Information Collected

Different from questioning, the other sub-stream of information gathering, i.e., information collected (S1b), concentrates more on what kind of information about applicants is collected (rather than sought for or used) by interviewers during interview process. Among the five studies recognised in this sub-stream, two examine the type of information collected and the other three discuss the features of information, like topic coverage and accuracy.

Regarding type of information, Burnett et al. (1998) checked the notes taken by interviewers, from which five types were identified and defined, including behavioural, dispositional, contextual, procedural, and judgmental notes. In the other study, Middendorf and Macan (2002) borrowed these concepts and took a close look at how differences in the notes taken by interviewers were related to their notetaking style (i.e., key-point notetaker, conventional notetaker) and applicants' interview performance (i.e., high vs. moderate). Typically, they found interviewers' notetaking style closely related to applicants' performance to the extent that more behavioural notes were recorded for applicants with better interview performance, while more dispositional notes were taken for applicants with moderate interview performance. They also noticed that more contextual notes were recorded by interviewers who were key-points notetakers, while conventional notetakers took down more procedural and judgmental notes.

When it comes to the features of information collected, one out of the three studies evaluated the information accuracy and the other two checked the trait coverage. In the research constructed by Macan and Dipboye (1994), notetaking was found enhancing the accuracy of interviewers' information recognition. Later by examining the information gathered by interviewers, van Dam (2003) comprehensively tracked the traits of applicant measured by interviewers and found that applicants' communication skills were assessed by most of the interviewers (93.1%), followed by teamwork skills (79.4%), adaptability (79.4%), and organizational skills (76.3%). In contrary, Nikolaou (2011) focused exclusively on interviewers' perceptions of applicants' personality and

the results of the study showed that interviewers covered all five personality dimensions in their description of applicants while with a preference for extraversion and agreeableness.

Summary of S1

A moderate number of articles ($N = 23$) are recognised investigating how interviewers differ in their information gathering behaviour, which may significantly influence their information processing and decision making during the following interview stages. In spite of the rich findings generated, three typical suggestions for future research are also provided. One is to explore the effects of a wider range of factors on interviewers' information gathering, particularly the environmental conditions like the situation of labour market, fair employment policies, and hiring quotas (Binning et al., 1988). Secondly, the repeated recognition of confirmatory behaviour leads to the query whether there is a correct or optimal level of it where interviewers can make accurate predictions about applicants based on the first impressions (Dougherty et al., 1994). Scholars like McDonald and Hakel (1985) strongly recommend more investigations into this direction, where the possibilities included examining the amount of information interviewer can effectively handle and uncovering interviewers' cognitive operations when forming impressions about applicants. Besides, regardless of the common sense that information gathering plays a significant role in decision making, it is still far from clear how interviewers' information gathering affects their decision making.

S2: Attributing

As suggested in social psychology, attributions serve as important predictors of a subject's social judgments about others, which have dominated organizational research (Martinko, 1995; De Faria & Yoder, 1997; Chapman & Webster, 2001). In the circumstances like employment interview where these judgments of other individuals greatly affect the evaluations made and actions to take, attributional processes turn out to be extremely salient. Surprisingly, only one tenth ($N = 11$) of the articles found are within this research stream, among which three studies discuss interviewers' character-

istics and ability of attributing and the other eight pay attention to interviewers' causal interpretations of applicant cues (e.g., past performance, interview performance).

Only one article commented on a single characteristic, i.e., depth, of interviewers' attributing. In this very study carried out by De Faria and Yoder (1997), interviewers were found make more in-depth attributions about the applicants with positive work history than those whose work history was negative and, as suggested by the authors, might no longer be under the consideration for the target position. As for the two studies on interviewers' attributing ability, one tested interviewers' inferential accuracy (Rothstein & Jackson; 1980), which measured how correctly interviewers could "judge an applicant's pertinent characteristics and identify behavioural exemplars as part of a pattern of behavioural consistencies" based on the limited information gathered during interview. The results indicated a high accuracy in interviewers' attributions which not only reflected the brief information they extracted from the interview but matched the empirical covariation of these behaviours as well. The other piece of work by Highhouse and Bottrill (1995) tested to what extent interviewers could identify misinformation about applicant behaviour when make attributions and interviewers turned out to be quite accepting of the misinformation and showed great confidence in the authenticity of the information.

The major function of attributions is causal analysis (Lord & Smith, 1983), based on which interviewers identify personal qualities of an applicant, determine the applicant's responsibility for a past work outcome, and make predictions about the applicant's future performance (De Faria & Yoder, 1997). Eight studies are found focusing on interviewers' causal interpretations of particular types of applicant cues, among which seven studies set an eye on applicants' concurrent interview performance and the rest one looks into applicants' past performance outcomes. In particular, Tucker & Rowe (1979) investigated whether interviewers' expectancy formed at an earlier stage would influence their causal interpretation of an applicant's past performance outcomes. They found that interviewers holding an unfavourable expectancy tended to give less credit to the applicant for past successes whereas more personal responsibility for past failures.

When it comes to applicants' interview performance, Gifford et al. (1985) investigated interviewers' perceptions of applicants' social skill and found their judgments were mainly based on three nonverbal cues of applicants, namely dress, time talked, and gesture rate. In another study carried out by Chapman and Webster (2001), the potential effect of media richness on interviewers' attributing style was examined and it was recognised that interviewers who perceived lower interview media richness were more likely to make external attributions for applicants' performance during the interview process, which then led to more favourable ratings as well.

Apart from the three studies above, scholars also paid a particular attention to how interviewers' perceived applicants' impression management tactics during interview process. Applicants' impression management (IM) is derived from their attempts to create a particular image of themselves so as to affect interviewers' perceptions of and consequently decisions made about them (Stevens & Kristof, 1995; Barrick et al., 2009; Kristof-Brown et al., 2002). According to the results of massive studies, IM tactics are applied by nearly every single applicant (Turnley & Bolino, 2001; Ellis et al., 2002; Levashina & Campion, 2007), whereas the type of tactics varies in the sense that applicants may adopt either honest IM that faithfully reports their abilities and experience or deceptive IM that embellishes their self-presentation to meet the job requirements (Gilmore & Ferris, 1989; Levashina & Campion, 2007). Giving that organisations may run the risk of hiring wrong persons if applicants are given higher scores and recruited due to their IM tactics, especially the deceptive ones, and that honesty and trustworthiness also become key dimensions to consider about when assessing applicants, it is vital for interviewers to accurately make judgments about applicants' IM tactics (Schmid Mast et al., 2011; Roulin et al., 2015).

Five articles were found digging into this topic. It was noticed that interviewers often failed to detect applicants' IM, where their detecting performance regarding those deceptive tactics was even worse than when faced with the honest ones (Roulin et al., 2015; Culbertson et al., 2016). Various factors were verified to be related to interviewers' detection of applicant deception. For instance, it was found that a warning

about deception's presence (Giordano et al., 2011), correct beliefs about indicators of deception (Reinhard et al., 2013) and careful scrutiny of interview content (Culbertson et al., 2016) could improve interviewers' detection accuracy. At the same time, several studies also examined whether interviewers with different levels of experience performed differently in deception detection but the results were mixed. While Reinhard et al. (2013) and Roulin et al. (2015) found no significant influence caused by interviewers' experience level on detection accuracy, experienced interviewers in the study of Mast et al. (2011) performed above average while student samples did not. Other interesting findings included that the type of interview questions could affect interviewers' detection of applicant IM (Roulin et al., 2015) and reviewers, who observed and reviewed the information contained in interview process, outperformed the interviewers who took part in the interactive process during interview in the detection of deceptive information (Giordano et al., 2011).

Summary of S2

While attributing plays a fundamental part of human decision making, only a limited number of studies ($N = 11$) are found investigating interviewers' attributing, among which nearly a half concentrate on interviewers' ability to detect deceptive IM tactics of applicants. There has been a call for further research on both the nature of interviewers' attributing, like the implicit criteria they apply in attributing (Rothstein and Jackson, 1980), and the factors affecting interviewers' attributing style, such as individual differences (Roulin et al., 2015) and interview context (Chapman & Webster, 2001).

S3: IDM Pattern

Giving the significant role of employment interview and the continuous academic interest in improving interviewers' decision quality, it is not surprising to find that the majority of the studies ($N = 78, 71.56\%$) attempt to uncover how interviewers score applicants and make hiring decisions. These articles can be categorised into two groups according to their research aims, namely **(S3a) rating strategy** and **(S3b)**

decision path. While the former takes a more specific perspective of interviewers' decision pattern and explores how interviewers weigh different cues of applicants in decision making ($N = 34, 31.19\%$), the latter group, adopting a more mathematical approach, investigates how interviewers' decisions are progressively formed with various intermediates involved ($N = 48, 44.04\%$).

S3a: Rating Strategy

Even if decision makers gather exactly the same information, their decisions can still vary a lot of different weights are assigned to the same piece of information. Plenty of studies are recognised investigating interviewers' rating strategies ($N = 33$) concerning about the levels of significance that interviewers attached to different cues when making decisions (i.e., cue weighting). Some of these studies ($N = 9$) were more neutral in tone and concentrated on the difference in interviewers' dimension coverage (Valenzi & Andrews, 1973; Mayfield et al., 1980; Zedeck et al., 1983; Dougherty et al., 1986; Kinicki et al., 1990; Graves & Karren, 1992) and consideration of applicants' verbal and non-verbal cues (Motowidlo & Burnett, 1995; Burnett & Motowidlo, 1998; DeGroot & Gooty, 2009) in their ratings. Nevertheless, a larger body of research ($N = 25$) dug into the typical bias commonly recognised in human decision makers like interviewers, which are listed in Table 2.3 with both definitions and the number of articles found.

Notably, order effect seemed to draw the greatest attention from scholars, which suggested that the timing and order of information occurrence would be given different levels of weights by interviewers when making decisions (Bolster & Springbett, 1961; Farr, 1973; Lunenburg, 2010; Strawn & Thorsteinson, 2015). Such effect was twofold, since interviewers might attach higher significance either to the information presented early in the process (i.e., primacy effect) or to that came in later stages (i.e., recency effect). Among the 14 studies on order effect, the majority of them ($N = 12$) discussed about a specific form of primacy effect, namely the impact of first impression on interviewers' hiring decisions. Recalling the existence of confirmatory behaviour in interviewers' questioning as discussed S1, it might not be surprising to notice the positive mathematical correlation between interviewers' first impression of the applicant

Table 2.3: Typical bias in interviewers' decision making

Bias	Definition	No. of Articles	Literature Example
Order Effect	The timing and order of information occurrence affect the levels of weights given by interviewers when making decisions	14	Barrick et al. (2012) Strawn & Thorsteinson (2015) Buijsrogge et al. (2021)
Halo Effect	A positive trait or characteristic of an applicant can influence interviewers' judgements on other unrelated factors of the applicant in a favourable way	5	Dougherty et al. (1986) Zysberg & Nevo (2004)
Information Favourability	The positive-negative nature of information affects the levels of weights given by interviewers when making decisions	3	Farr (1973) Constantin (1976)
Stereotype of Ideal Applicant	Interviewers' personal beliefs of what an ideal applicant should be like	2	Mayfield & Carlson (1966) Hakel (1971)
Contrast Effect	Interviewers' ratings given to the current applicant are usually affected by the preceding applicant's quality	2	Heneman et al. (1975) Cesare et al. (1988)
Similar-to-Me Effect	Interviewers tend to rate applicants with similar biographical backgrounds, attitudes or personalities to themselves more favourably	1	Anderson & Shackleton (1990)

and their final decisions (McDonald & Hakel, 1985; Phillips & Dipboye, 1989; Macan & Dipboye, 1990, 1994; Barrick et al., 2012; Swider et al., 2016; Carnes et al., 2019). Several studies conducted an even more in-depth investigation and examined how such first impression affected interviewers' rating process. The most significant finding might be the recognition of belief-adjustment (Strawn & Thorsteinson, 2015) or anchoring-and-adjustment (Buijsrogge et al., 2021) behaviour, a decision-making mechanism which meant that interviewers started from an initial impression of the applicant and then adjusted it until a final decision is made.

While primacy effect is often referred to as a type of bias, a few scholars have raised doubts that if interviewers' first impression about an applicant can be correct.

Typical evidence includes interviewers' performance in a task of applicant personality assessment that interviewers can accurately measure the personality characteristics of applicants solely based on the applicants' resumes (Cole et al., 2003). What's more, it should also be noted that, while first impression commonly refers to the pre-interview impression that interviewers form about an applicant based on the background information available, its definition can vary across the studies focusing on interviewers' final decisions. According to the stages at which the first impression is measured, the term may refer to either the pre-interview impression (Herriot & Rothwell, 1983; Phillips & Dipboye, 1989; Macan & Dipboye, 1990, 1994; Florea et al., 2019) or interviewers' perception and evaluation of the applicant formed throughout the rapport building stage, the first few minutes of an employment interview where the interviewer aims to reduce applicant nervousness and establish temporary relationship (i.e., rapport) with the applicant (Chapman & Zweig, 2005; Barrick et al., 2012; Swider et al., 2016; Buijsrogge et al., 2021).

In addition to the order of information, the positive-negative nature of information was verified as a key factor related to interviewers' cue weighting as well, which is termed as information favourability. In particular, interviewers were often found overly relying their decisions on negative information while placing insufficient weight on the opposite (Farr, 1973) even if the information was of low relevancy (Constantin, 1976). Research was also recognised investigating the existence of halo effect (i.e., base the ratings of a set of dimensions on a single general impression of the applicant; Hakel, 1971; Heneman et al., 1975; Parsons & Liden, 1984; Dougherty et al., 1986; Zysberg & Nevo, 2004), stereotypes of ideal applicants (Mayfield & Carlson, 1966; Hakel, 1971), perceived similarity with the applicant (i.e., similar-to-me effect; Anderson & Shackleton, 1990), and the quality of preceding applicants (i.e., contrast effect; Heneman et al., 1975; Cesare et al., 1988).

Research was also called for that brought well-developed decision-making theories and models into the investigation of interviewers' rating strategy, whereas only one study was found. Specifically, Alliger et al. (1993) introduced fuzzy set theory into

interviewers' decision-making process and noticed that, rather than explicitly determine whether an applicant was "in or out" regarding certain criteria, interviewers used fuzzy natural language distinctions when making judgments.

S3b: Decision Path

Although the number of studies is still going up that adopt univariate analysis to explore how interviewers' decisions can be affected by a certain factor (which are not included in this review since they are not process research by definition), increasingly more scholars are devoted to establishing the possible paths of interviewers' decision making ($N = 48, 44.04\%$). Their efforts are undoubtedly of great importance, especially considering that interviewers are less likely to make a sudden hiring decision but usually form an overall perception and evaluation of an applicant progressively. The path illustrating interviewers' decision making basically involves three categories of variables, including antecedents, intermediates, and the outcome (see Figure 2.3 for an example). Hence, the studies within this sub-stream are quite similar regarding the form of research output, which demonstrates how the effect of various factors (i.e., antecedents, such as applicant demographic characteristics, appearance, interview performance) on interviewers' decision is mediated by certain processive variables (i.e., intermediates, such as interviewers' perceptions of applicant characteristics, dimensional ratings, personal liking, multiple cognitive factors). The review of this sub-stream concentrates on the research distribution across the range of intermediate types, while the particular findings regarding the mathematical correlations between these factors and interviewers' decisions will not be discussed in detail. Two main categories of intermediates were recognised, namely interviewers' social perceptions of applicant and their cognitive factors.

Social perceptions of applicant. When it comes to employment interview where interviewers meet and interact with applicants, it is natural that interviewers form social perceptions of applicants that may later on play a part in their decision making. Thirty-seven articles were identified taking interviewers' social perceptions of applicant into account when investigating their decision-making path. The social perceptions

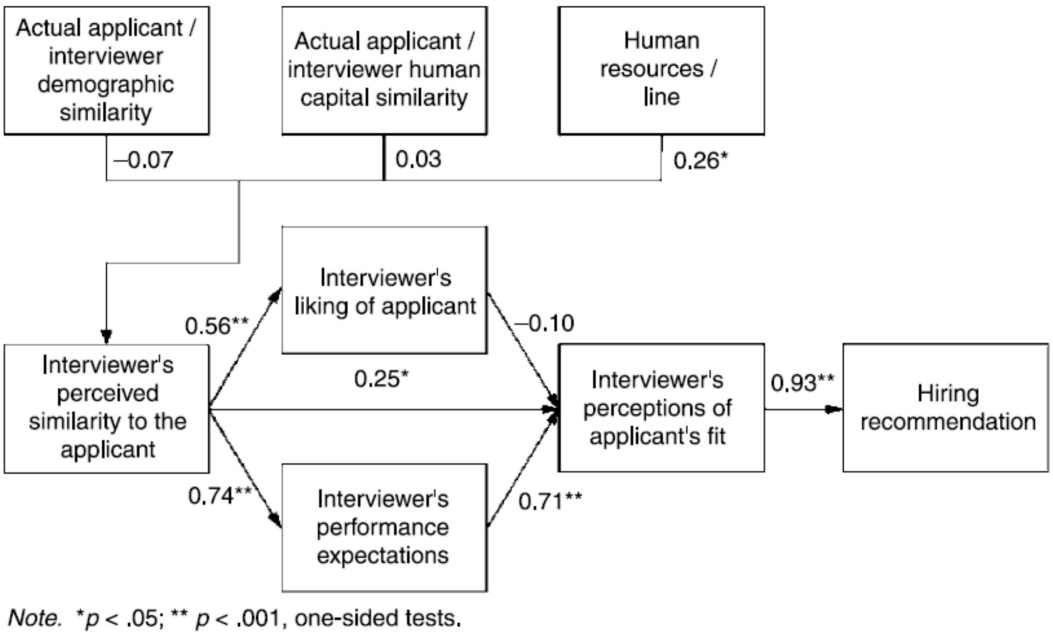


Figure 2.3: Example of interviewers' decision path (Garcia et al., 2008)

discussed in these studies can be classified into four categories, including interviewers' perceptions of an applicant's individual characteristics, interview performance, overall qualification, as well as interviewers' personal reactions to the applicant.

The first and most common type of intermediates involved in IDM path analysis were applicant individual characteristics as perceived by interviewers ($N = 25$). Among these studies, 15 out of 25 studies asked interviewers to rate applicants on a wide range of personal traits, including their competence (Howard & Ferris, 1996; Barrick et al., 2010; Thompson et al., 2015; Amaral et al., 2019), intelligence (Keenan, 1977; Baron, 1983; Raza & Carpenter, 1987), and various personality-related dimensions (Anderson & Shackleton, 1990; Rynes & Gerhart, 1990; Van Dam, 2003; DeGroot & Gooty, 2009; Roulin et al., 2014; Amaral et al., 2019), all of which were then brought into the decision path as intermediates. Another type of interviewers' perception of applicant characteristics that drew a wide attention from scholars ($N = 8$) was interviewer-applicant similarity as subjectively recognised and measured by interviewer themselves (Gallois et al., 1992; Graves & Powell, 1995; Howard & Ferris, 1996; Fox & Spector, 2000; Chen

et al., 2008). Besides, there were also some studies looking into how applicants' physical attractiveness (aron, 1983; Raza & Carpenter, 1987; Rynes & Gerhart, 1990; Tsai et al., 2012), accent (Purkiss et al., 2006; Deprez-Sims & Morris, 2013), feminity and masculinity (Kith et al., 2022) were perceived by interviewers and thus affected their hiring decision.

As for perceived interview performance of applicant, seven articles were found which examined how interviewers' decision path could be affected by their perception and interpretation of applicant interview performance. Except for one study that introduce the concept of "interview quality" as generally evaluated by interviewers into the discussion (Graves & Powell, 1996), other scholars investigated more specific aspects of applicant interview performance, including nonverbal behaviours (Parsons & Liden, 1984; Tsai et al., 2012), multiple impression management tactics (e.g., self-promotion, deceptive ingratiation, image creation and protection, Roulin et al., 2014), feature of language used (Moore et al., 2017), acknowledgement own physical disability (Hebl & Skorinko, 2005), anxiety (Feiler & Powell, 2015).

Perceived overall qualification of applicant were frequently adopted to demonstrate the formation of interviewers' decisions as well ($N = 13$), while the focus was pretty much on two dimensions, namely fit and employability. Fit, particularly person-organisation fit (i.e., P-O fit) and person-job (i.e., P-J fit), was an indicator recommended by scholars and consultants to measure whether an applicant fitted with the organisation or the target job from various perspectives, such as the strategies, culture, norms, and values (Gerstein & Reisman, 1983; Herbert & Deresky, 1987; Kerr, 1982; Leontiades, 1982; Olian & Rynes, 1984; Tichy, Fombrun, & Devanna, 1982; Rynes and Gerhart, 1990; Hedge & Teachout, 1992; O'Reilly, Caldwell, & Mirable, 1992). Seven out of the thirteen articles concerned about whether interviewers' perceptions of applicant fit with the organisation (i.e., P-O fit) or the target job (i.e., P-J fit) played a role in interviewers' decision making. Particularly, it seemed that early studies tended to focus on a single aspect, i.e., either on P-O fit (Rynes & Gerhart, 1990; Adkins et al., 1994; Cable & Judge, 1997) or P-J fit (Baron, 1983), whereas the latest research took

both aspects into analysis (Kristof-Brown, 2000; García et al., 2008; Chen & Lin, 2014). Another common dimension that partially illustrated interviewers' decision path was employability, for which four articles were identified (Rynes & Gerhart, 1990; Adkins et al., 1994; Hayes & Macan, 1997) while more recent research was lacking. Apart from fit and employability, applicant skill level (Raza & Carpenter, 1987) and a general concept of qualification (Graves & Powel, 1996) were also measured and introduced into interviewers' decision path model.

The fourth category of interviewers' social perception of applicant, i.e., interviewers' affective reactions towards applicant, turned out to be another promising angle to uncover their decision path, of which fourteen studies were found. An interesting finding was that the majority of the discussion ($N = 13$) concentrated on interviewers' positive reactions, such as liking or likability investigated in seven studies (Raza & Carpenter, 1987; Anderson & Shackleton, 1990; DeGroot & Motowidlo, 1990; Fox & Spector, 2000; Barrick et al., 2010), interpersonal attraction (Graves & Powell, 1995; Deprez-Sims & Morris, 2013), and affect (Howard & Ferris, 1996). The study conducted by Chen et al. (2010), however, specified the distinction between positive affectivity and negative affectivity brought them both into the path analysis.

Interviewer cognitive factor. Apart from interviewers' perceptions of applicant, various cognitive factors of themselves were also studied as sources of intermediates in their decision making ($N = 13$). or instance, interviewers' perceptions and understanding of the interview task, such as their perceived severity of negative concerns (Tsai et al., 2012), sex-role stereotyping (Gallois et al., 1992; Tsai et al., 2012), and accountability for the outcomes (Florea et al., 2019), were found mediating the decision path. It was also found that interviewers' reasoning style played an essential role in the formation of hiring decision, particularly their interpretations of applicants' interview performance (Baron,1986; Ramsay et al., 1997; Chapman & Webster, 2001; Hebl & Kleck, 2002), dispositions (De Kock, 2015), past performance outcomes (Tucker & Rowe, 1979). Other interviewer cognitive factors introduced into the path model of interviewer decision making included their expectations of the applicant (Biesanz et

al., 1999), attention (Madera & Hebl, 2012), confidence in own decision (Carnes et al., 2019) and mood (Baron,1987).

Summary of S3

Account for the largest proportion of process research on interviewers' decision making ($N = 78, 71.56\%$), the studies within the third research stream (S3: IDM Pattern) unveil interviewers' decision-making process to a certain extent by either investigating particular tendencies of interviewers' decision making or examining how a wide range of factors may interact with each other and collectively influence interviewers' hiring decisions. Future research is encouraged to bring more possible factors into analysis, such as interviewer fatigue (Cable & Judge, 1997), job status (i.e., high vs. low; Graves & Powell, 1995) and interviewers' personality (Purkiss et al., 2006). The rich findings regarding interviewers' decision bias also indicate the necessity of enhancing their judgment effectiveness, which may benefit from ascertained criteria (e.g., dimensionality of fit; Schwab, 1980; Rynes & Gerhart, 1990; Adkins et al., 1994; Howard & Ferris, 1996) and so-called "valid judgements" first (Kristof-Brown, 2000), followed by well-developed training courses that are currently of significant lack (Howard & Ferris, 1996; Kennedy, 1994; Purkiss et al., 2006). Furthermore, studies that introduce decision-making theories and models to promote the understanding of interviewers' performance is quite limited despite of the long-lasting call for such research.

S4: Other IDM Performance

A handful of studies ($N = 14, 12.84\%$) investigating several aspects of IDM performance other than the topics discussed above are clustered into the fourth research stream, i.e., **(S4) other IDM performance**. This group of research cover a wide range of dimensions regarding IDM performance, including their procedure compliance and (visual) attention as they carry forward the interview process, as well as their decision time, justification behaviour, and overconfidence when making decisions. Performance like interviewers' awareness of own DM process and memory accuracy are also investigated.

Procedure compliance. One of the studies concerned about the application of the highly-appraised structured interview in the sense that interviewers might not carry out the process as required (Di Milia & Gorodecki, 1997). The results confirmed such concern and showed that the situation was caused by various factors, like lack of role clarity, different interpretations of job specification and inconsistent use of the rating system, and inexperienced interviewers.

(Visual) Attention. Two studies were recognised looking into interviewers' (visual) attention during interview process. Specifically, Madera and Hebl (2012) conducted experiments to investigate interviewers' discrimination against facially stigmatized applicants and tracked their visual attentions. They introduced the concept of self-regulation, which referred to "the exertion of control over the self and occurs when a person attempts to change the way he or she would otherwise think, feel, or behave" (Muraven & Baumeister, 2000, p. 248) and was regarded as a limited cognitive resource, into employment interview scenario. They proposed that interviewers' self-regulation might be depleted if their attentions were put on applicants' stigmas rather than on interview content, and the research outcomes supported their propositions.

After stalling for nearly a decade, the research on this topic was then carried on by Buijsrogge et al. (2021), who again examined interviewers' visual attention when interviewing stigmatised applicants (defined as attention-grabbing effect) but split the discussion within different interview stages (i.e., rapport building stage, the subsequent interview stage). The results showed that interviewers' attention-grabbing effect of stigma was initially high in the rapport building and decreased in the subsequent interview stage.

Decision time. Stimulated by a persistently held belief that interviewers tended to make quick decisions within the first few minutes of the interview process (Arvey & Campion, 1982; Buckley & Eder, 1988; Huffcutt, 2010; Judge et al., 2000; Springbett, 1958), two studies were recognised investigating the difference in interviewers' decision time while the factors of concern differed. In one of the studies, Tullar et al. (1979) looked into the impact of applicant quality and interviewers' expectation of interview

length on the time they spent on decision making. It was found that interviewers required more time to make decisions both when assessing applicants of high quality and when they expected the interview to last longer. Whereas in the other piece of work completed by Frieder et al. (2016), a model was developed and tested to demonstrate how interviewers' decision time was affected by factors like interviewer experience, interviewing efficacy, question consistency, applicant order, and whether interviewers engaged in more rapport building with the applicant.

Decision justification. Only one study was found exploring how interviewers justified their decisions. To be specific, Crissy and Regan (1951) examined the evidence provided by interviewers in their interview reports and noticed that the rejected group as a whole received significantly more negative evidence. The study also revealed that more evidence was presented for the same pool of less qualified applicants if they were in the accepted group than when they were in the rejected group. Conversely, more evidence was provided in support of the "outstanding" applicants in the rejected group than for the same applicants in the accepted group.

(Over)confidence in own decision. Confidence in one's own judgments has been regarded as a vital construct when studying and measuring a subject's cognitive performance on decision-making tasks (Harvey, 1997; Klayman et al., 1999, Ratcliff & Starns, 2013), which, however, has been grossly neglected in the process research on IDM (Kausel et al., 2016). Two studies were found noticing interviewers' overconfidence in their decisions. On one hand, Buijsrogge et al. (2016) examined how an applicant's stigma affected interviewers' overconfidence in their biased judgments related to the stigma. The results indicated that interviewers reported overconfidence in their negatively biased judgments toward stigmatised applicant and the level of such overconfidence were mediated by interviewers' professional performance perceived by applicants. On the other hand, Kausel et al. (2016) compared the overconfidence of interviewers either with access to (unstructured) interview information or presented with test scores only, and the former groups if unterviewers exhibited a higher level of overconfidence than the latter.

Awareness of own decision-making process. Some scholars also expressed their curiosity about whether interviewers were aware of own decision-making process, particularly the rating process regarding the criteria they used and the weights they gave to different cues. As suggested by Graves and Karren (1992), interviewers who are more aware of own decision-making process can control own performance during this process more consciously as well as to retrospect and analyse own past decision-making performance for future improvements. Two papers were found looking into this topic, one of which was the study conducted by Valenzi and Andrews (1973), revealing serious discrepancies between interviewers' intended cue weights and the actual cue weights applied. The other contribution to this topic came from Graves and Karren (1992) who regarded the awareness of cue weighting as an indicator of interviewer effectiveness. According to their findings, the importance attached by effective interviewers to the selection criteria during decision making, when compared with those are less effective, were more likely to mirror their self-reported criteria importance.

Memory. Compared to the previous six topics, interviewers' memory turned out to have drawn the most attention from scholars ($N = 6$) and was found to be susceptible to various factors. One of the factors, according to the two articles also mentioned under the topic of interviewers' attention (Madera & Hebl, 2012; Buijsrogge et al., 2021), was the applicant's stigma. Specifically, it was found that interviewers recalled less information about the interview if the applicant was stigmatised (Madera & Hebl, 2012) and that applicant stigma had a significant moderate effect on interviewer memory accuracy in rapport building but not in the sequent interview stage (Buijsrogge et al., 2021). Other instructive findings included that (i) interviewers who chose the informational order were faster to recall the information encountered early in the interview (Johns, 1975), (ii) notetaking could increase interviewers' memory accuracy (Middendorf & Macan, 2002), (iii) applicants' self-presentation tactics (e.g., positive nonverbal cues, the use of perfume) would affect interviewers' memory of the information presented by the applicants (Baron, 1986), and (iv) interviewers recalled more information presented by the applicant that was consistent with their current mood (Baron, 1987).

Summary of S4

In spite of the limited quantity ($N = 14, 12.84\%$), the studies covered in this supplementary stream contribute to the knowledge of IDM from diversified perspectives. Apart from the rich components themselves brought into the discussion, additional insights into the relationships between these components and the outcomes as well as corresponding mechanisms for improvements are expected (Graves & Karren, 1992; Ellis et al., 2013; Buijsrogge et al., 2016; Kausel et al., 2016). Meanwhile, future research should also investigate how IDM performances in terms of these aspects can be affected by multiple individual, social, and situational factors (Frieder et al., 2016).

2.2.3.2 Background Features

The background features of process research on IDM are reported from three perspectives, namely year of publication, journal background, and geographic origin of the study.

1) Year of publication

Tracking the decade-wise evolution of process research on IDM (see Table 2.4), significant increases can be observed in the number of articles in both 1970s and 1980s, followed by a heated discussion lasting for the next few decades, except for 2000s. Specifically, process research in the earliest two decades was quite rare and focused on interviewers' rating strategy (Bolster & Springbett, 1961; Mayfield & Carlson, 1966) and justification behaviour (Crissy & Regan, 1951). Then, the quantity of articles went up rapidly in 1970s with the emergence of two new research streams, namely information gathering and attributing. A limited part of academic attention was kept on these two streams during the following years, where a subtle discrepancy could be recognised in the two sub-streams of information gathering in the sense that the studies on (S1b) information gathered only occurred during 1991-2020 with a small number. In contrary, (S3) IDM pattern dominated the research direction at a relatively stable level, whereas

Table 2.4: Time series of identified studies by research stream

Research Stream	1951-1960	1961-1970	1971-1980	1981-1990	1991-2000	2001-2010	2011-2022	2011-2022
S1: Information Gathering			**2**	**7**	**5**	**3**	**5**	**1**
a) Questioning			*2*	*7*	*3*	*1*	*4*	*1*
b) Information Collected					*2*	*2*	*1*	
S2: Attributing			**2**	**1**	**2**	**1**	**5**	
S3: IDM Pattern		**2**	**10**	**15**	**19**	**13**	**17**	**2**
a) Rating Strategy		*2*	*8*	*10*	*5*	*2*	*6*	*1*
b) Decision Path			*2*	*7*	*13*	*12*	*13*	*1*
S4: Other IDM Performance	**1**		**3**	**2**	**2**	**1**	**4**	**1**
Total	**1**	**2**	**13**	**21**	**25**	**15**	**29**	**3**

Note. 17 articles are listed under multiple research streams.

the tendencies of focus on the two sub-streams differed. As indicated by the statistics, articles contributing to the knowledge of interviewers' rating strategy kept growing from 1960s, reached its peak ($N = 10$) in 1980s, and then decreased in the following decades. However, the academic interest in interviewers' decision path increased continuously to a maximum ($N = 13$) in 1990s and remained steady throughout the next few decades.

2) Journal background

All the 109 articles reviewed are published in the leading academic journals as indicated by AJG rankings, including two articles from two different HRM journals and the rest 108 studies derive from 12 organisational psychological journals (see Table 2.5). Concerning each individual journal, *Journal of Applied Psychology* turns out to be the major contributor and provides over 27.52% ($N = 30$) of the articles, followed by other four journals with similar quantities of identified papers, namely *International Journal of Selection and Assessment* ($N = 17, 15.60\%$), *Personnel Psychology* ($N =$

15, 13.76%), *Journal of Applied Social Psychology* (N = 13, 11.93%), and *Journal of Occupational and Organizational Psychology* ($N = 13, 11.93\%$). These five journals have contributed 80.91% of the sample literature, while the contribution from the rest journals is quite limited. Nevertheless, the numbers of citations of these journals do not strictly match their ranks in terms of articles quantities. Particularly, while *Journal of Applied Psychology* again takes the first place ($N = 5664, 39.88\%$) and *Personnel Psychology* achieves 3812 (26.84%) citations with its 15 articles as well, *Journal of Applied Social Psychology* only generates 1362 (9.59%) citations and the number of citations regarding *International Journal of Selection and Assessment* with the second most articles is even much smaller ($N = 576, 4.06\%$).

3) Geographic origin

Regarding the geographic origin of the literature, only about half of the articles ($N = 56, 51.38\%$) clarify the national regions where the studies are carried out (see Table 2.6). Among these articles, a clear divide can be observed between the contribution of America and that of the remaining regions regarding the leading numbers of not only articles ($N = 36, 33.03\%$) but also citations ($N = 5885, 41.23\%$), followed by UK and Canada with a similar level of contribution in terms of both article quantities ($N = 5, 4.59\%$) and citations numbers (424 vs. 479, 2.99% vs. 3.37%). Other regions, like China, Germany, Switzerland, Australia and Greece, provide a quite limited contribution to the research topic.

2.2.3.3 Methodological Features

In addition to the thematic and background features, the characteristics of the methodologies adopted by these empirical studies on IDM process are also checked from three aspects, including method of data collection, participant type, and source of evidence (see Table 2.7).

Table 2.5: Number of articles identified in each journal

Journal	**Field**	**No. of Articles (%)**	**No. of Citations (%)**
Journal of Applied Psychology	OP	30 (27.52%)	5664 (39.88%)
International Journal of Selection and Assessment	OP	17 (15.60%)	576 (4.06%)
Personnel Psychology	OP	15 (13.76%)	3812 (26.84%)
Journal of Applied Social Psychology	OP	13 (11.93%)	1362 (9.59%)
Journal of Occupational and Organizational Psychology	OP	13 (11.93%)	1001 (7.05%)
Organizational Behavior and Human Decision Processes	OP	5 (4.59%)	690 (4.86%)
Journal of Business and Psychology	OP	4 (3.67%)	273 (1.92%)
Human Performance	OP	3 (2.75%)	46 (0.32%)
Journal of Organizational Behavior	OP	2 (1.83%)	530 (3.73%)
Applied Psychology	OP	2 (1.83%)	53 (0.37%)
European Journal of Work and Organizational Psychology	OP	2 (1.83%)	20 (0.14%)
Journal of Managerial Psychology	OP	1 (0.92%)	142 (1.00%)
The International Journal of Human Resource Management	HRM	1 (0.92%)	16 (0.11%)
Personnel Review	HRM	1 (0.92%)	16 (0.11%)
Total		**109**	**14201**

Note. OP refers to Organisational Psychology and *HRM* refers to Human Resource Management.

1) Data collection method

Four main types of data collection methods have been adopted by these 109 articles, including

- *Survey*, where data are collected from people with past experience as either interviewers or applicants through methods like questionnaire
- *Real-interview example*, where data are collected during the process of real-world interview

Table 2.6: Number of articles and citations by geographic origins

Geographic Origin	No. of Articles (%)	No. of Citations (%)
USA	36 (33.03%)	5885 (41.23%)
UK	5 (4.59%)	424 (2.99%)
Canada	5 (4.59%)	479 (3.37%)
China	4 (3.67%)	111 (1.56%)
Germany	2 (1.83%)	72 (0.51%)
Switzerland	2 (1.83%)	280 (1.97%)
Australia	1 (0.92%)	21 (0.15%)
Greece	1 (0.92%)	16 (0.11%)
Netherland	1 (0.92%)	86 (0.61%)
NA	53 (48.62%)	6806 (47.93%)
Total	**109**	**14201**

Note. 1. *NA* means that the geographic origin of the article is not clear; 2. One article consisting of studies conducted in both USA and Canada is listed under the both regions.

- *Real-interview retrospection*, where data are collected through participants' retrospection soon after the completion of a real interview

- *Mock interview*, where data are collected by conducting scenario experiments that simulate a real interview

Apart from the six articles using survey for data collection, all the other 103 studies rely on the evidence derived from a recently conducted interview process – either a real interview or a simulated one. Particularly, mock interview seems to be the most prevalent method of data collection among the articles reviewed and has been utilised in about 65.14% of them ($N = 71$). We have also noticed that mock interview can take various forms in terms of how the "applicant" is presented to the participant, which can be grouped into three categories:

Table 2.7: Classification of contextual perspectives by research stream

Contextual Perspective	S1	S2	S3	S4	Total (%)
Data Collection Method					
1) Survey	5	0	1	0	6 (5.50%)
2) Real-Interview Example	3	0	6	0	8 (7.34%)
3) Real-Interview Retrospection	3	0	19	4	24 (22.02%)
4) Mock Interview	12	11	52	10	71 (65.14%)
Live	*4*	*0*	*18*	*5*	*21 (29.58%)*
Videotaped/Audiotaped	*3*	*9*	*23*	*3*	*34 (47.89%)*
Paper Person	*5*	*3*	*12*	*3*	*18 (25.35%)*
Participant Type *(Mock Interview)*					
1) Experienced	2	2	18	5	24 (33.80%)
2) Novice	10	7	26	3	37 (52.11%)
3) Mixed	0	2	8	2	10 (14.08%)
Source of Evidence					
1) Segmented Judgments	6	10	57	2	69 (63.30%)
2) Interviewers' Processive Performance	13	1	6	7	21 (19.27%)
3) Mixed	4	0	15	5	19 (17.43%)

Note. 1. 17 articles are listed under multiple research streams; 2. The percentage of each subcategory of *Data Collection – Mock Interview* refers to the proportion of articles based on mock interview that adopt certain types of mock interview (i.e., live, videotaped/audiotape, or "paper person"); 3. The percentage in *Participant Type* refers to the proportion of articles based on mock interview that adopt certain types of participants (i.e., experienced, novice, or mixed).

- *Face-to-face interview*, where an actor plays the role of applicant and gives the same response to all the interviewers

- *Videotaped/audiotaped interview*, where interviewers are presented with prepared videos/records of interview process

- *"Paper person"*, where "interviewers" are presented with applicant information, responses to interview questions, or interview transcripts in written form

About 47.89% ($N = 34$) of the studies based on mock interview simulate an interview process with pre-recorded videos or tapes, followed by 29.58% ($N = 21$) using face-to-face interview and a similar number of studies presenting "paper person" to interviewers ($N = 18, 25.35\%$). As for the 32 articles based on real-world interview, only eight of them gather information from the exact interview process, while the rest 24 studies collect data through post-interview retrospection.

2) Participant type

It has drawn early concern from scholars whether the experience level of participants as interviewers affects the reliability of conclusions because college students (i.e., novice interviewer) have been commonly used as interviewer samples. The results are quite mixed in the sense that quite a few scholars have found no significant differences (Dipboye et al., 1975; Arvey & Campion, 1982), while some other researchers have suggested the importance of subject type (Barr & Hitt, 1986; Gordon et al., 1986; Harris, 1989). The findings of this review also show that novice interviewers have been widely adopted for data collection (articles based on real interview are not included for comparison since the participants are all experienced). As can be seen in Table 2.7, novice interviewers are involved in 66.19% ($N = 47$) of the studies, including 37 solely rely on novice samples and 10 using mixed types of participants, whereas about 33.80% ($N = 24$) studies use experienced interviewers only. Meanwhile, it is necessary to highlight the fact that none of the articles claiming a use of experienced interviewer have clarified the definition of "experienced" and a common standard is that the participant has past experience interviewing applicants or is currently responsible for recruiting.

3) Source of evidence

Apart from the data collection method, the source of evidence used in the studies reviewed also varies. To summarise, the sources of evidence can be categorised into two types, including

- *Segmented judgements* (SJ) in terms of scores or levels of indicators given by interviewers when evaluating an applicant (e.g., ratings of competencies, hiring intentions at different stages of interview)

- *Interviewers' processive performance* (IPP), like interviewers' visible behaviours (e.g., the content of notes taken by interviewers, visual attention, and decision time) as well as their thinking process during interview process

As can be seen in Table 2.7, the numbers of articles adopting these two types of evidence source are quite unbalanced. Only a limited number of articles ($N = 40, 36.7\%$) adopt the source of evidence that straightforwardly presented interviewers' decision-making performance during the process (i.e., IPP). In contrast, the products of interview process (i.e., SJ) in the terms of the segmented judgments made by interviewers about an applicant turn out to be the most preferred source of evidence when explaining the interview process. Specifically, SJ is recognised in 80.73% ($N = 88$) of the articles, among which only 19 studies also consider about IPP while the other 69 studies rely their conclusions solely on SJ. When looking at the source of evidence used in each research stream, particular preferences can be recognised as well in the sense that IPP is utilised more frequently than SJ both in (S1) information gathering (six studies with SJ, thirteen studies with IPP) and (S4) other IDM performance (two studies with SJ, seven studies with IPP), whereas SJ plays a dominating role in both (S2) attributing (ten studies with SJ, one studies with IPP) and (S3) IDM pattern (57 studies with SJ, six studies with IPP).

2.3 Gap Identification

While it has been emphasised for decades that process research is needed to provide a better understanding of how interviewers' decision-making process unfolds until they render a final decision, a great lack of corresponding studies can be perceived from three perspectives as indicated by the SLR. First, the number of process research on IDM is

quite limited. As can be recognised in the procedure of article selection and screening, only 109 articles identified on IDM are empirical research with a processive nature that involve either processive variable(s) or processive correlating when generating a conclusion. Secondly, the academic interests in different research streams are quite unbalanced. To be specific, an overwhelming majority of process research on IDM ($N = 78, 70.91\%$) focus on interviewers' decision patterns that investigate how different sources of information and factors play a part in interviewers' decisions. In contrast, only a small part of the articles identified ($N = 31, 29.09\%$) investigate other aspects of IDM, among which even fewer studies concern with interviewers' actual thinking process during decision making. Thirdly, in spite of the scholars' research aim to uncover the ongoing facts during IDM process, most of the studies rely heavily on the products of IDM rather than interviewers' decision-making performance during the process. Specifically, the conclusions of most studies ($N = 69, 63.30\%$) are based solely on the segmented judgments made by interviewers about the applicant or the statistics derived from various tests taken by interviewers, whereas only a relatively small number of studies ($N = 40, 36.70\%$) take interviewers' processive performance into the analysis of IDM process.

It is worth noting that a confined number of studies are recognised trying to investigate interviewers' thinking process based more direct evidence rather than correlation analysis or observations of interviewers' external behaviours. The sources of evidence adopted by these studies are mainly interviewers' self-reports or surveys carried out after the completion of interview process (Graves & Karren, 1992; Podsakoff et al., 2011; Kleinmann et al., 2016). While contributing to the knowledge of ICP during decision making to a certain degree, two significant limitations can be perceived in these studies which are derived from the features of data collection. On one hand, both self-reports and surveys in these studies are designed to collect information about quite specific aspects of interviewers' thinking process, like their interpretations and memories of the clues or awareness of own decision-making processes (see for example, Tullar et al., 1979; Binning et al., 1988; Hebl & Kleck, 2002; Podsakoff et al., 2011). Apart from the knowledge contributed by these studies to the understanding of ICP, a relatively

complete image of ICP when they work through the decision-making task is absent.

On the other hand, the data collection procedure based on interviewers' self-reports and surveys is usually conducted after the interview process, thus the data collected have a retrospective nature that may hamper the data validity. In particular, two main drawbacks exist in these retrospective methods when used to collect data about a subject's cognitive process, namely memory loss and memory distortion, which negatively affect the extent to which the data represent the subject's real thinking process (Nisbett & Wilson, 1977; Russo et al., 1989; Ericsson & Simon, 1993; Harte et al., 1994; Neuman & Schwarz, 1998; Nielsen et al., 2002). For example, the subject may not remember all the details of what happens during the task completion, especially when the task itself is complicated or completed a long time ago. Another possibility is post hoc rationalising where the subject reports own cognitive process as a coherent and rational trace while that may not be the case when it takes place. In some studies of IDM process, taking the experiment conducted by Tullar et al. (1979) as an example, the records of the interview process are presented to help interviewers with retrospection, whereas the real impact of such practice on the outcomes is unclear.

Hence, in addition to the overall short of process research on IDM, a research gap in IDM research can be identified that there is a great lack of empirical studies on ICP during decision making especially based on the data that can effectively represent interviewers' real thinking processes. To be specific, while IDM is a typical cognitive task where a series of information-processing activities taking place in interviewers' minds, only a limited amount of studies concern with interviewers' cognitive process (i.e., ICP) and most of the outcomes are basically supported by indirect evidence, like deduction from correlation analysis and observation of interviewers' external behaviours. Meanwhile, most of the existent studies on ICP concentrate on particular aspects of the decision-making task (e.g., information gathering, attributing, cue weighting, specific bias), whereas an empirical study exploring the overall prospect as well as the characteristics of ICP is absent. In addition, the evaluation of IDM has long been limited to the psychometric properties (e.g., reliability, validity, applicants' reactions) of the decisions

made, whereas measurements of interviewers' performance during decision-making process is absent (e.g., compliance with interview structure), which may also due to a lack of knowledge regarding ICP. Such a gap not only restrains the understanding of IDM but also conceals the opportunities to improve it.

2.4 Summary

Employment interview has been one of the most commonly used personnel selection tools across regions, industries and organisations, and it has drawn wide attention from both practitioners and scholars for over one hundred years. Driven by the need to improve the quality of interviewers' decisions, a great deal of empirical studies exploring IDM have accumulated throughout the period and there has been a long-standing call for process research to find out the actual facts that take place during IDM.

This chapter first briefs the background of academic interests in IDM as well as the evolution of process research on IDM. In spite of the fruitful results, it is noticed that a review that clarifies the current status of process research on IDM and points out the opportunities of future research is absent. Moreover, even the definition of the so-called process research is vacant. As a response to this situation, a systematic literature review is conducted for process research on IDM. First, as one of the most key criteria of article selection, the boundary of process research on IDM is discussed and defined as those empirical studies involving either processive variables or processive correlating to uncover the ongoing facts during IDM process. Afterwards, a transparent and replicable procedure of article searching and screening is carried out based on the definition. The articles identified are then critically reviewed from three perspectives, including the thematic, background, and methodological features.

According to the findings of the SLR, the number of process research on IDM is quite limited and the focuses of the articles identified are pretty much on IDM patterns, whereas less attention is paid to other aspects of IDM performance. Meanwhile,

an unbalanced reliance of existent research on the products of IDM (e.g., segmented judgments made by interviewers) is also noticed through the SLR, while only a handful of studies derive their conclusions from interviewers' processive performance during decision making (e.g., information gathering, attributing process). In particular, while IDM is a typical cognitive task where a series of information-processing activities taking place in interviewers' minds, only a mere number of studies investigate ICP and most of the outcomes are basically supported by indirect evidence. Even for those studies directly examining interviewers' thoughts, their data collection is based on retrospective methods (e.g., self-report, survey) which are vulnerable to memory distortion and memory loss. Therefore, a research gap in IDM is identified from the SLR that there is a significant lack of knowledge regarding ICP that leads to the decisions, especially based on the data that can effectively represent interviewers' real thinking process.

Chapter 3

Research Design, Data Collection and Analysis

Without a deeper understanding of ICP during decision making, especially based on the data that can effectively represent interviewers' real thinking processes, further improvements to interview processes and outcomes are limited. Hence, this research proposes to apply a qualitative research approach called think aloud method (TAM) to explore ICP. TAM is regarded as a more direct and superior way than other approaches to uncover a subject's thoughts when working on a task (Ericsson & Simon, 1984; Van Someren et al., 1994; Charters, 2003). The technique has been widely utilised in many fields, but it has not yet been used within an interview scenario. Therefore, this chapter of research design starts with an introduction of the history and mechanism of TAM in **Section 3.1**, where a comparison between TAM and other common methods used in human cognition research is presented as well. Considering the specific requirements of TAM on the choice of both task and participant, the feasibility of applying TAM to explore ICP is discussed in **Section 3.2**, based on which an experiment procedure is developed, tested and revised through a pilot study. The overall research design of data collection and data analysis are separately described with rationale explained in **Section 3.3** and **Section 3.4**, and the trustworthiness of the research is argued in

Section 3.5. Finally in **Section 3.6**, a summary of the present chapter is provided.

3.1 Think Aloud Method: A Tool Uncovering Human Cognitive Process

This section introduces the psychological method, i.e., TAM, which is widely used in multiple domains to explore subjects' cognitive processes. In particular, the emergence and mechanism of TAM in human cognition research is explained, followed by a comparison between TAM and other typical approaches utilised in the investigation into human cognitive process. Afterwards, the particular requirements of the technique on the choice of task and participant are discussed.

3.1.1 Emergence and Mechanism of TAM

Psychologists has long been interested in exploring human cognitive process when performing different tasks, such as decision making (Chapman & Elstein, 2000; Stiegler & Tung, 2014), designing (Chan, 1990; Kelley, 2008), and educating (Fiedler & Beier, 2014; Miller & Dumford, 2016). Early research was based on the subject's introspection, a research technique that collects data by involving trained observers to carefully and objectively observe and analyse their own thoughts (Payne et al., 1978; Ericsson & Crutcher, 1991; Harte et al., 1994; Boren & Ramey, 2000; Nielsen et al., 2002). Nevertheless, two main defects are noticed when utilising introspection to study a subject's cognitive process. On one hand, the data collected with this method is vulnerable to the observer's own interpretation even if the observer is highly trained. Consequently, the data can deviate significantly from the real cognitive process that took place. On the other hand, the data collected through introspection is only accessible to the subject performing the thinking process and it is impossible to replicate the empirical study, thus constraining scientific discussions over the cognitive process investigated.

The limitations of introspection incurred suspicious and led to the establishment of behaviourism by John B. Watson in the 1930s (Watson, 1913, 1920; also see for a review, Ericsson & Crutcher, 1991). Criticising introspection for its lack of objectivity and reliability, behaviourists focused exclusively on the observable behaviour of subjects. They believed that a subject's behaviour was shaped by environmental stimuli through the conditioning when the subject interacted with the environment, whereas those theorizing about processes that could not be directly observed from the outside the body should be banned. Behaviourism as a research technique had its advantages in psychology in the sense that it allowed researchers to collect and quantify data without relying on observers' interpretation, whereas this methodology could hardly tap into subjects' cognitive processes.

Later in 1940s, psychologists like Watson (1920) and Duncker and Lees (1945) who were curious about human cognitive processes while suspicious of introspection brought a methodological advancement into the field by proposing a new methodology called TAM. As a conceptually straightforward method for the process data of human cognition (Payne, 1994), TAM requires a subject to verbally report all the thoughts that come to mind (i.e., think aloud) when working on the task of interest, and the concurrent verbal reports are recorded and investigated to understand the subject's cognitive process. For example, as early as the study conducted by De Groot (1946), he used TAM to investigate expert chess players' thinking processes. Specifically, De Groot identified a great pool of specific concepts and principles used by the players and noticed that their cognitive processes were actually a progressive refinement of a plan with these concepts and principles. Another typical study was carried out by Newell and Simon's (1972) who combined think aloud protocols (i.e., verbal reports collected through TAM) with computer models of problem-solving processes and constructed a detailed model of participants' cognitive processes when solving the well-known cryptarithmetic problem (i.e., to find out the ten digits which could substitute the ten letters in this equation to make it a correct arithmetic sum, DONALD + GERALD = ROBERT).

In spite of the inspiring findings generated through this emerging methodology,

however, a few sceptical notions existed concerning about the validity of data. For example, it appeared that the way scholars gathered and analysed data with TAM varied tremendously and many research publications hardly provided any detail of the procedure, thus it was suggested that the information provided by verbal reports was informal and required verification by other data (Deffner, 1990; Tamler, 1998; Boren & Ramey, 2000). Scholars also worried that whether TAM as a secondary task would alter the subject's performance and the corresponding cognitive process, because the verbalising procedure will cost at least some of the subject's cognitive resources (Crutcher, 1994; Payne, 1994). As the interest in human cognitive processes grew even faster, a variety of discussions on the validity and the practice of TAM were accumulated, among which the work of Ericsson and Simon (1984) was believed to be the best source which not only defended the validity of TAM but also shaped and uniformed how TAM was applied in the subsequent cognitive process studies to a great extent. Based on substantial theoretical and empirical evidence, they provided a detailed discussion over the concerns and mechanism of using verbal protocols as data and stressed several issues that determined the data validity, especially the choice of task and how the data collection procedure should be carried out.

According to the human memory system proposed by Atkinson and Shiffrin (1968) to illustrate the mental processes of a subject when working on a task, human memory system contains three major components (see Figure 3.1 for a simplified model), namely sensory buffer(s), short-term memory (STM, also called working memory), and long-term memory (LTM). When a subject interacts with the external environment, information is received by various sensory buffers (e.g., eyes, nose) of the subject and transformed into an internal form (i.e., perception). The information is then processed by the subject's STM, where a small amount of information can be held for a short of time in an active and readily available state. A large amount of information flowing into STM is then filed and stored in LTM for an extended period (i.e., storage) and can be directly retrieved to STM (i.e., retrieval) when needed.

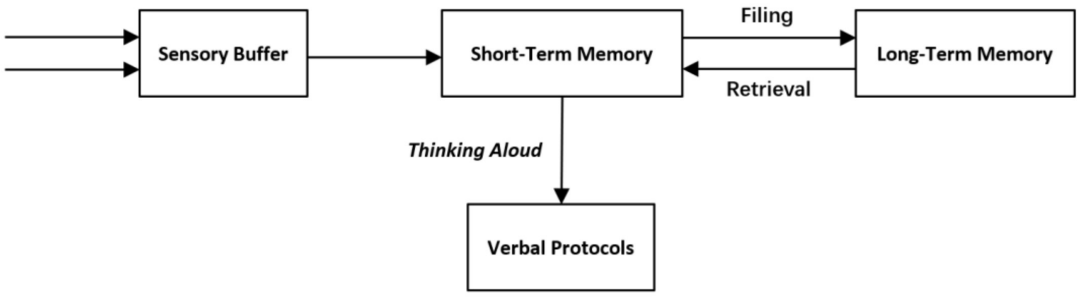


Figure 3.1: Human cognitive system and thinking aloud

When applying TAM to exploring human cognitive processes, according to Ericsson and Simon, only the information active in STM can be verbalised (see Figure 3.1), whereas the information in LTM can only be verbalised once it is retrieved to and processed in STM. Moreover, if the information in STM is already in an oral form, no additional processing time or effort is needed for the subject to verbalise the information and the verbalisation is executed almost automatically and would not interfere with the task performance. As for the information processed in STM while in a non-oral form (e.g., visual), more time and effort will be necessary for the subject to recode the information in to oral form so as to verbalise it, which slows down the task completion to a certain extent but it is believed that such recoding will not alter the structure of cognitive processes when performing the target task. Along with their arguments, Ericsson and Simon also review a great deal of TAM-based studies, showing that executing TAM does not necessarily cause influence on the amount and pattern of information processed or task outcomes.

3.1.2 Comparison between TAM and Other Approaches

Apart from TAM, several other techniques are also commonly utilised in cognitive process research. On one hand, two indirect sources of evidence are often used to interpret human cognitive process, namely the products of cognitive processes and the observations of subjects' behaviours. Research based on product analysis focuses on the

results of subjects' cognitive processes, like the hiring decisions made by interviewers in the context of employment interview, but not much information about subjects' cognitive processes could be gathered. Many other studies observe the behaviour of a subject right as it takes place either through simple observation or by using special equipment (e.g., eye-tracking, active parts of brains). Such technique is also adopted in many studies on employment interviewers' performance (Tullar, 1989; Dougherty et al., 1994; Stevens, 1998; Silvester & Anderson, 2003; Stewart et al., 2008; Madera & Hebl, 2012; Florea et al., 2019). The behaviour trace observed and recorded is called action protocols, which are analysed and interpreted to speculate on the subject's cognitive process (Moskowitz, 1986; Suen & Ary, 2014). Behavioural observation is one of the few methods that provide rich evidence about the process of a subject's task completion, whereas it does not give direct access to information about the subject's cognitive process either.

On the other hand, there are several other methods that also generate verbal reports from subjects regarding their cognitive processes while not in a concurrent manner as TAM. In addition to Introspection discussed in Section 3.1.1, typical examples also include retrospection, questioning and prompting, and dialogue observation. As for retrospection, it usually requires a subject to recall and report own thinking process after the completion of the task, which is found to be the main method utilised in IDM research to investigate interviewers' cognitive processes according to the SLR (see Section 2.2). As discussed in the literature review, drawbacks exist in retrospection to collect data about subjects' real cognitive process in terms of memory loss and memory distortion, which hamper the validity of data collected (Nisbett & Wilson, 1977; Russo et al., 1989; Harte et al., 1994; Neuman & Schwarz, 1998; Nielsen et al., 2002).

The third method is questioning and prompting, by applying which the subjects can be prompted to report or explain what they are thinking or doing at any time during task completion. The advantageous of this method is that, by deciding what to ask or prompt, data collection can be more targeted to investigate specific aspects of the subject's cognitive process. For example, scholars interested in interviewers'

ability to detect the existence of fake information in applicants' response may ask the participants whether they perceive deception or straightforwardly instruct them to indicate the information they regard as dishonest (Giordano et al., 2011; Reinhard et al., 2013; Roulin et al., 2015). However, the shortcoming is also apparent that the subject can be repeatedly interrupted when working on the task, and self-explanation is found altering (mostly in a positive way) the subject's performance because additional clues and triggering information processing are introduced into the subject's task completion (Van Someren et al., 1994; Berardi-Coletta et al., 1995; Chi et al., 1994; Renkl, 1997; Ericsson, 2006). Thus, the extent to which the process simulates the subject's real task performance (i.e., validity of data) may be questioned.

Regarding dialogue observation, it can only be applied to the situation where the task itself naturally involves dialogue (e.g., counselling), and the dialogue process will be audio- or video-taped for analysis (Rost, 1989; Riggenbach, 1991; Perry et al. 2002; Whitebread et al. 2007; Heintze et al., 2010). The advantage of this technique is that the data can be collected in real-world practice rather than laboratory settings, whereas it may not be applicable to the investigation of many tasks where the subjects do not verbalise their own thoughts. Hence, this method cannot be adopted to investigate interviewers' cognitive process regarding their decision making since interviewers are less likely to express their inner thoughts in an actual interview when communicating with the applicant.

When it comes to research applying TAM where the subject is asked to keep talking aloud own concurrent thoughts when working on a task, as discussed in Section 3.1.1, it is verified that the think aloud practice itself will not affect how the subject works on the task. Therefore, the validity of data collected through TAM should not be negatively affected by outside interruption or the subject's interpretation if carried out properly. Furthermore, TAM is particularly preferred for its three significant advantages as summarised by Payne (1994). First, the subjects taking part in the data collection procedure are required to verbally report all the thoughts without being given further instructions or probes, thus they are usually naïve about the specific interests

of the researcher and the data collected will be less affected by methodological artefacts. The second benefit of TAM lies in the time-ordered or sequential feature of the verbal protocols collected considering that "thinking is viewed as a temporal sequence of mental events (Payne, 1994)". Since the verbal protocols provided by subjects are concurrent reports of their ongoing thinking processes and reflect the exact sequence of the mental events taking place in their minds, the data allows researchers to develop and test the models of subjects' cognitive process. Thirdly, the verbal protocols collected through TAM are open-ended and believed to provide the richest and unprocessed source of data regarding human cognitive process. Researchers can focus on any specific aspects of interest when analysing the data, such as instances of certain types of thoughts or occurrence of different styles of reasoning, and the results can be computed either across individuals, tasks, or other controlled conditions. However, such characteristics also make the corresponding data analysis more time- and effort-consuming (Cotton & Gresty, 2006; Lundgrén-Laine & Salanterä, 2010). So far, no study is found applying TAM to studying interviewers' cognitive processes. Even when searching within the discipline of employment interview, only one article is identified which investigates applicants' decision making about whether or not to apply for a job by analysing the verbal protocols collected with the applicants (Barber & Roehling, 1993).

To compare TAM with these methodologies that also used to collect verbal protocols of human cognitive process, van Someren et al. (1994) propose three main factors closely related to the quality of the verbal reports, namely disturbance of process, memory errors, and interpretations. Disturbance of process refers to whether the data collection for the subject's cognitive process will interrupt the task performance of the subject. Memory errors are basically determined by the timing of data collection as either concurrent or post hoc. And interpretations usually occur when the subject is asked to interpret and even explain own cognitive processes, or when the subject is prompted with questions that does not fit the real process. The comparison among the methods discussed above regarding these three dimensions is summarised in Table 3.1. It indicates that, while the quality of verbal protocols collected through introspection,

Table 3.1: Comparing methods in cognitive process research (Van Someren et al., 1994)

Method	**Disturbance**	**Memory Errors**	**Interpretation**
Introspection	no	little	yes
Retrospection	no	yes	yes
Prompting	yes	no	little
Dialogue	*Not applicable*		
TAM	no	no	no

retrospection and prompting are negatively affected by disturbance, subjects' memory errors and interpretation to difference extents, TAM seems to be less invulnerable to these factors.

3.1.3 Requirements of TAM on the Choice of Task and Participant

In spite of the advantages of TAM, the validity of data collected through the technique can be questioned if the task and participant are not suitable to be investigated by TAM or the methodology is not implemented properly. First, whether TAM is an effective method to investigate a subject's cognitive process when performing a task is determined to a great extent by several features of the task itself, which are defined by van Someren et al. (1994) as task type, task difficulty, and task representativeness.

Task type. The requirements of TAM on the task type lie in two aspects (Ericsson & Simon, 1984; Fonteyn et al., 1993; Payne, 1994; Charters, 2003). One is that tasks involving verbal communication (e.g., traffic control, psychotherapy) are not suited for TAM in their original form, because it is impossible for participants to provide concurrent verbal report while engaging in communication at the same time. The other concern regarding task type when applying TAM for data collection is whether speed is inherent in the completion of the task. Since it takes unpredictably more time

for a subject to complete the same task when required to concurrently verbalise own thinking processes and the speaking speed also varies significantly among individuals, TAM can hardly be utilised to those tasks requiring speed of completion.

Task difficulty. In addition to task type, TAM also has particular requirements on the difficulty level of task from two perspectives (Johnson, 1992; Van Someren et al., 1994; Charters, 2003; Heerkens & Van Der Heijdenm 2005). On one hand, the task should not be too simple so that a subject can "work through" it and verbally report the cognitive process involved. Otherwise, the subject may complete the task in an automated manner (e.g., $1 + 1 =?$) without a reportable thinking process. On the other hand, the verbal reports provided by subjects can be expected to promote the understanding their cognitive processes during task completion only when they are able to complete the task to a certain degree, so the task should not be so difficult that the subjects can hardly work on it. In both situations where the task is too simple or beyond a subject's ability, it is less likely to expect the subjects to effectively provide verbal reports that reflect their cognitive processes of task completion as there is almost no such a "process".

Task representativeness. The third requirement of TAM on the feature of task suggests that the chosen task needs to be representative enough in terms of a particular type of problems investigated (Van Someren et al., 1994; Ericsson & Simon, 1998; Ericsson, 2006). This means that the task should not be unusual in any sense and a subject's cognitive process observed when working on this task should provide sufficient information about how people performing that certain kind of problems in general.

Apart from the task features discussed above, whether the data collection procedure with TAM embedded is executed properly can also affect data validity, which raises requirements on not only the subject but also the experimenter who organises and hosts the whole procedure. On the part of subjects, it is necessary to ensure the subjects involved in the data collection possess verbalisation skills that they can fluently think aloud and verbalise their cognitive processes (Fonteyn et al., 1993; Lumley, 2002; Charters, 2003; Cooke, 2010). While one may concern whether screening subjects

according to their verbalisation skills will influence the randomness of sample, it is believed that such individual property is not associated with the task investigated and the corresponding cognitive processes. Following the suggestions from Ericsson and Simon (1993) as well as most studies that investigated or adopted TAM, a warm-up session where subjects are asked to work on a mathematical problem while thinking aloud is often conducted first as a brief test of their verbalisation skills.

Meanwhile, how the procedure is carried out and how the experimenter provides instructions to ask subjects to think aloud during data collection can also influence data validity. For example, to ensure the quality of verbal report and the records, the procedure should be conducted in a one-on-one manner and in a quiet room to avoid disturbance and the experimenter is suggested to sit at the side of, rather than opposite to, the subject to minimize the possible tension caused on the subject (Van Someren et al., 1994). Before starting the data collection procedure, the experimenter should deal with the subjects' queries if there is any, particularly to make sure they understand how to think aloud. However, once the data collection procedure starts where the subject works on the task while thinking aloud, the interaction between the experimenter and the subject must be minimised to avoid influencing the subject's task performance. One exception is that, if there is a significant pause in the subject's verbal report (e.g., the subject keeps silence for more than five seconds), the experiment should remind the subject by saying "please keep speaking" (Ericsson & Simon, 1984, 1998; Gero & Tang, 2001; Cumming et al., 2002; Li & He, 2015; Han, 2017). More importantly, probing questions and requirements for explanations should be avoided since such instructions will change the subjects' task performance and cognitive process by pushing them to go beyond the information they currently focus on and report information that is unnecessary to perform the task (Ericsson & Simon, 1980; Payne, 1994; Van Someren et al., 1994).

With the clarification on the psychological base as well as the methodological issues of TAM, the method draws increasingly more attention and gain more acceptance in various disciplines, such as clinical decision making, consumer research, design, usability

Table 3.2: Application of think aloud method in different disciplines

Discipline	Literature Example
Clinical Decision Making	Fonteyn & Fisher (1995); Ericsson (2004); Funkesson et al. (2007); Lundgrén-Laine & Salanterä (2010); Forsberg et al. (2014)
Consumer Research	Dhar & Sherman (1996); Kivetz & Simonson (2000); Rayner et al. (2001); Reicks et al. (2003)
Design	Gero & Tang (2001); Jaspers et al. (2004); Perry & Krippendorff (2013); Sarkar & Chakrabarti (2013)
Usability Testing	Benbunan-Fich (2001); Roberts & Fels (2006); Cooke (2010); Olmsted-Hawala et al. (2010)
Language Learning	Kussmaul & Tirkkonen-Condit (1995); Jääskeläinen (2000); Bernardini (2001); Li (2004)
Essay Rating	Cumming et al. (2002); Lumley (2002); Li & He (2015); Han (2017)
Sports Performance	French et al. (1996); Ward et al. (2004); Nicholls & Polman (2008); Eccles & Arsal (2017)

testing, language learning, essay rating, and sports performance (see Table 3.2 for example studies).

3.2 Feasibility of Applying TAM in IDM Research

Since no empirical study is found applying TAM to employment interview scenarios, there is no such a well-developed experiment procedure that can be directly learned from when designing a TAM-based data collection procedure for IDM research. Meanwhile, recalling that TAM requires certain preconditions to be met by both the task itself (i.e., task type, task difficulty, task representativeness) and the subjects involved (i.e.,

ability to both work on the task and think aloud), it is crucial to find out whether and how the method can be effectively implemented within an interview scenario rather than rush into data collection and analysis. To check the feasibility of applying TAM in IDM research, this section first discusses the potential concerns and possibilities of introducing TAM into an interview scenario following the criteria of TAM on both the task and subjects. According to this discussion, an experiment procedure is developed and then tested through a pilot study.

3.2.1 Concerns and Possibilities of Applying TAM in IDM Research

Recalling the discussion in the previous section, TAM cannot be applied to all types of tasks to explore a subject's cognitive process and a few factors may interfere with the think aloud practice (Ericsson & Simon, 1980; Fonteyn et al., 1993; Van Someren et al., 1994; Charters, 2003; Heerkens & Van Der Heijden, 2005), including task type, task difficulty, task representativeness, and execution of TAM. Therefore, the concerns of investigating interviewers' cognitive processes with TAM are discussed first mainly according to these four criteria, of which the key points are summarised in Table 3.3 with a detailed discussion of corresponding countermeasures presented in the following context.

1) Task type

To be investigated by TAM, the task of interest is expected not to involve verbal communication or to require speed of completion. When it comes to the IDM research, the main concern is that a real interview itself usually involved verbal communication between interviewer(s) and the applicant and interviewers are less likely to verbally express their perceptions of the applicant right away during the interview process. It is therefore less likely to apply TAM to a real interview process. However, as can be noticed in the literature review, it is a common practice in most empirical studies on IDM to create a mock interview scenario where participants, rather than conducting

Table 3.3: Criteria and concerns of applying TAM to IDM research

Criteria		**Potential Concern**
Task Type	The task itself does not involve communication	• Interaction between interviewer and applicant
	The task does not require speed of completion	• Time limit of interviewer decision making
Task Difficulty	The task should be at a reasonable level of difficulty (i.e., not too easy but can still be handled)	*Participant-related:* • Interview-related knowledge and experience *Task-related:* • Job position and evaluation criteria • Length of interview • Sources of information • Informative level of applicants' responses
Task Representativeness	The task can well represent the sort of problems	• Interview questions and evaluation criteria (e.g., transferable skills)
Execution of TAM	Participants can verbalise their thinking process	• Necessity of a warm-up session

and getting involved in real interview communication themselves, play the role of interviewer and are presented with materials that simulate an interview scenario, such as premade videos, recordings, or even interview scripts in a written form. This is especially reasonable and necessary for the purpose of controlling variables. At the same time, it is also quite common in real-world practice that interviewers attending an interview without communicating with the applicant. For instance, it is possible that not all the interviewers in a panel interview communicate with the applicant, and novice interviewers may contribute to the final decision according to their observations during the interview without interacting with the applicant. Therefore, this research suggests that it is practicable to present a recorded interview scenario to participants to simulate an interview process without getting them involved in interview communication, thus allowing the implementation of TAM.

As for the speed of completion, it is hardly a requirement for an interview task either in real-world situations or in academic research, and participants are usually allowed

sufficient time to work on the task. Hence, when applying TAM to IDM research, participants acting as the interviewer should also be permitted with enough time as needed to fully report their thoughts during the process and the think aloud practice should not be interfered by any concern about time limits. Overall, the task type of employment interview can be expected to suit the requirement of TAM.

2) Task difficulty

To ensure that the verbal protocols of participants' cognitive processes can be generated when working on a task, the task should not be too simple to be completed in an automated manner or too difficult to prevent the participants from working through the task. When it comes to an interview task, the main concern is about whether participants can effectively act as an interviewer and evaluate the applicants for the target job position. Thus, it is necessary to make sure that

- the target job position is "general" enough in the sense that it is not too technical to be understood
- the materials used to simulate the interview process (e.g., job description, rating scale, applicants' responses) are informative enough without significantly causing cognitive overload on participants
- the languages used in the experiment can be effectively understood by participants

Target Job Position. Since this research aims to contribute to the general knowledge of ICP during decision making regardless of the type of job, any job position should be appropriate as long as it presents a reasonable level of task difficulty – neither a primary job position for which a job interview is even not needed, nor a technical job where employers focus more on applicants' capabilities that are evaluated through other approaches (e.g., test, work sample) instead of interview. The concern about task difficulty also requires a careful choice of interview questions and evaluation criteria, which

should be comprehensible to interviewers who are not experts in the corresponding field. Thus, similar to most mock interview experiment conducted in IDM studies (DeGroot & Motowidlo, 1999; Fox & Spector, 2000; DeGroot & Gooty, 2009), the interview scenario in this research will be created as an interview which aims to evaluate a few transferable competencies (e.g., leadership skills, organisation skills, problem-solving skills, stress management skills) of applicants for a managerial position.

Interview simulation vs. cognitive load. Due to the time restrictions as well as the limited cognitive load of participants, TAM can usually be applied to only a rather small set of tasks or problems to minimise the possibility that participants become too exhausted to complete the task or think aloud at a normal standard (Branch, 2000; Katalin, 2000; Nielsen et al.; 2002; Heerkens & Van Der Heijden 2005). Therefore, while a complete interview process is most ideal to simulate a real practice, it is necessary to consider about the trade-off between the levels at which the experiment simulates a real-world interview process and participants' cognitive load when working through the task. In particular, careful considerations are needed about what parts of an interview process should be included in the mock interview and what information should be presented to the participants.

Despite the fact that a history of over 100 years has greatly enriched the forms available of employment interview (e.g., leaderless group discussion), this research as a primary investigation into ICP will focus on the most traditional and prevalent "question & answer" (Q&A) approach, where an interviewer learns about the requirements of the target job position, conducts interview by asking an applicant interview questions and making evaluations according to the applicant's responses to the questions, and finally compares all the applicants interviewed to make a choice among them (if available). Correspondingly, experimental materials like the background information of the company, job description and evaluation criteria (e.g., rating scale) that simulate an interview task for a vacant job position should be developed.

In addition, it is essential to predetermine what information about applicants to present, based on which interviewers will choose their hiring intentions. In real-world

interviews, interviewers usually have access to two sources of information about an applicant, namely the applicant's background information (e.g., resume) and the applicant's responses to interview questions. Nevertheless, this research will only provide the latter one due to two main concerns unless such design significantly affects the interview process (according to the feedback given by each participant involved in data collection). One is about the possibility that interviewers make early decisions about an applicant according to the background information and thus "confounds the interpretation of the value of the interview (Campion et al., 1997)". The other concern is about the balance between the number of applicants to be evaluated and the amount of information provided regarding each applicant. To better simulate an interview situation, it will be appropriate to offer interviewers more than one applicant for comparison, while the amount of information provided about each applicant needs to be limited to a certain extent to minimise the possibility of interviewer cognitive overload.

Interviewers' experience level. While a subject's experience level is often used as an indicator of ability to complete a task, the standard of inclusion regarding interviewers' experience level will be relatively loose within this research due to three considerations. First, unlike some other roles or skills, it is less feasible to measure the experience level of an interviewer. Typical dimensions like time length and frequency of practice (i.e., conducting interviews) are not applicable because interviewer itself is not a job position, and for people like directors and line managers who often act as interviewers as well, interview is only a small part of their work without a fixed frequency. There is even no such a clear sign of "becoming" an experienced interviewer in real-world practice. For instance, according to the survey with a practitioner who has been working as a human resource consultant for over thirty years and frequently involves in employment interviews,

> *Normally a person who was inexperienced would merely sit in the room away from the interview group and observe interviews taking place. After they felt confident, they would then spend some time participating by asking simpler questions to begin with. Once they were more confident, knowledgeable, and*

experienced, they could play a larger part in the process and introduce their own preferred individual style into the process.

Secondly, the definition of interviewers' experience level is also absent in academic research. Many empirical studies do not distinguish the experience level of interviewers or simply recruit students as interviewer samples, suggesting that the threat to generalisability of results is minimal (Purkiss et al., 2006; DeGroot & Gooty, 2009; Florea et al., 2019). While in the studies aimed to compare the performance of experienced interviewers and that of novice ones, the boundary of the two groups is also as simple as "with experience of conducting employment interviews" (Sackett, 1982; Tsai et al., 2012; Roulin et al., 2015; Frieder et al., 2016) or "working in the personnel area" (Reinhard et al., 2013). Thirdly, as an initial investigation into ICP, the main purpose of this research is to inclusively generalise and characterise interviewers' thinking process regardless of their experience level rather than to model the cognitive progress of expert interviewers. Therefore, people who have real-world experiences as interviewers or have attended formal employment interviews (to ensure that they are aware of how the interview process works) will be considered as qualified interviewers in this research, whereas their actual task performance as well as verbalisation skills will be monitored later to decide whether the data collected should be included for analysis.

Language. Since the data collection procedure will be carried in both the UK and China, the language used throughout the whole experiment procedure is another aspect closely related to the level of task difficulty considering the diversity in interviewers' geographic origins and language skills. To avoid the potential negative impacts caused by language barrier on interviewers' task performance while controlling variables as well, the experiment will be prepared in both English and Chinese versions so that interviewers can choose the one with which they can effectively understand the information and clearly express themselves. In particular, the two versions of experimental materials are checked and approved by an expert bilingual in English and Chinese to ensure that not only the meaning but also the structure of the content are maintained the same to minimise the influence of language difference on the data collection.

3) Task representativeness

Task representativeness requires that the task chosen should not be unusual in any sense and a subject's cognitive process observed when working on this task should provide sufficient information about how people performing that certain kind of problems in general. In other words, it not only requires the mock interview developed in this research to well simulate a real interview scenario, but it also indicates that the job position and evaluation criteria selected to form the interview task should better be common and transferable. Hence, a managerial position with soft skills (e.g., leadership skills, organisation skills, problem-solving skills, stress management skills) as evaluation criteria is preferable, which is consistent with the conclusions derived from the previous discussion over task difficulty.

To enhance the level of simulation, the experimental materials like job description (with background information of the company), target competencies to evaluate (i.e., rating scale), and corresponding interview questions to ask are prepared with the assistance from professional interviewers. Specifically, a job description is created for a project manager by imitating a real one posted online. Then a pool of possible competencies and corresponding interview questions are listed and sent together with the job description to three professional interviewers for their feedback on which ones are most likely to be asked during an interview for the target job position. Following to their comments, the questions and evaluation criteria are screened and refined, which sketches the outline of the whole mock interview process. However, starting questions like "tell me about yourself", although quite prevalent in real interviews, are not included for two reasons, including (i) to protect the confidentiality of the volunteers' identity who volunteer to act as the applicants and provide responses to the interview questions, (ii) to control the overall amount of information and minimise the possibility of cognitive overload caused on interviewers.

After that, the communication between interviewer and applicants during the interview process that form the main part of the interview scenario needs to be prepared. To

achieve this, two volunteers who are preparing for interviews of similar job positions are invited to attend a mock interview. In particular, the two volunteers are interviewed separately by an interviewer with the preselected list of interview questions. The whole interview process is recorded modified, and then checked by the professional interviewers to ensure that the communication is clear, comprehensible, and provides sufficient information for interviewers to make hiring decisions for the target job position. A recording rather than video of the interview communication is used because, as a primary attempt to apply TAM to interview process, this research intends to simplify the sources of information presented to interviewers by excluding visual information like applicants' appearance, gestures, facial expression, and environmental information so that interviewers can better focus on vocal information that can be directly verbalised when executing TAM.

4) Execution of TAM

To effectively collect data of interviewers' cognitive process, it is necessary to ensure interviewers possess a sufficient level of verbalisation skills and can fluently verbalise their concurrent thoughts (Fonteyn et al., 1993; Van Someren et al., 1994; Karsenty, 2001; Charters, 2003; Cooke, 2010). To ensure the participants taking part in the experiment meet the standard, a warm-up session is provided before starting the interview task to check if a participant is capable of and comfortable with TAM. Similar to most TAM-based studies, participants are asked to work on a mathematical problem while verbally reporting their ongoing thoughts. The mathematical problem involved should have a proper level of difficulty that allowed participants to work through it to a certain degree, and participants who can correctly execute TAM during the process are regarded as ideal sample for the following mock interview task even if they cannot figure out the right answer to the mathematical problem.

5) Other concerns

Apart from the four main factors discussed above, there are a few more details of ex-

periment design that required consideration. First, since interviewers' verbal protocols during the experiment will be recorded, related information must be clearly explained in a consent form to avoid ethical problems, including what and how participants' output will be recorded, stored, and analysed and that the confidentiality of their personal information will be guaranteed.

Secondly, to ensure the quality of verbal protocols as well as the recordings, details regarding the environment where the experiment is conducted must also be paid attention to. For example, the experiment should be conducted in a quite meeting room to avoid disturbance, and it would better if the experimenter sits at the side of the participant to minimize the tension of the participant. At the same time, small and casual talks at the start of the experiment will be preferable to further ease the participant's nervousness. A possible choice can be collecting participants' background information through casual interview rather than asking them to fill up the questionnaire in written form.

Thirdly, several details as shown below in Table 3.4 need to be highlighted before each participant starts both the warm-up session and the mock interview task and the experimenter should make ensure that the participant's queries are all solved.

Fourthly, the interaction between the experiment and the participant must be minimised when the participant is working on the task. The only exception is that, if there is a significant pause in the participant's verbal protocol (e.g., the participant keeps silence for more than five seconds), the experiment should remind the participant by saying "please keep speaking" (Ericsson & Simon, 1984; Van Someren et al., 1994; Gero & Tang, 2001; Li & He, 2015; Eccles & Arsal, 2017).

Last but not the least, considering that the feedback from participants will be of great value for experiment validation and improvement, a survey concerning participants' perceptions of the whole experiment procedure is carried out following the completion of the interview task. Specifically, a list of questions presented below will be asked and the participants are instructed to indicate their perceptions with a 5-point

Table 3.4: Experimental instructions of the warm-up session and the mock interview task

Experimental Instructions
• The experiment is not aimed at judging participants' performance on the task
• Participants should be natural and honest about their reactions to the task (e.g., if they feel the task is difficult or unconformable and want to stop)
• Participants should verbalise all the thoughts that occur to them while working on the task whether the thoughts are directly related to the task or not
• If participants are reading any material, they should read it aloud
• Participants should 'act and talk as if talking to themselves'
• The experimenter will avoid interactions with the participant unless necessary (e.g., if there is a significant pause in the participant's verbalization)
• There is no time limit to complete the task and participants should feel free to take as long as needed to work on the task and think aloud during the process

Likert Scale. An exception is the question concerning impact of TAM on own task performance where a score from -5 to 5 can be given with -5 meaning strongly negative impact, 0 meaning no impact, and 5 meaning strongly positive impact.

- To what extent do you think the scenario experiment simulates a real interview?

- Do you think the length of the whole experiment is appropriate?

- Do you think the task difficulty is reasonable? Is there anything you feel hard to understand during the process (e.g., language, applicants' responses)?

- To what extent do you feel confident about your evaluations and decisions?

- Do you think TAM affects your DM? If yes, how and to what extent do you think TAM has affected your DM?

- Do you have any further advice on this experiment?

Overall, the discussion above stresses the potential concerns of applying TAM in IDM research and provides possible countermeasures as well. It justifies the feasibility of collection the data of ICP with TAM to a certain degree and lays the foundation of the next step to develop a primary design of experiment procedure. Further examination of the feasibility will be achieved later by collecting participants' feedbacks on the experiment procedure.

3.2.2 Experiment Design of Interview Task with TAM Embedded

Based on the discussion above, an experiment procedure composed of five stages is developed (see Figure 3.2), including consent form, pre-experiment survey for the participants' background information, warm-up session of TAM, mock interview task with TAM, and post-experiment survey for the participant's feedback on the experiment. In particular, the whole experiment procedure is conducted in a one-on-one and face-to-face manner in a quiet meeting room where the experimenter sat by the side of the participants (see Appendix A.).

Stage 1: Consent Form

As the first stage of the whole experiment procedure, the participant is required to read and sign a consent form to ensure the experiment-related information is understood and accepted by the participant to avoid the possibility of ethical issues. Upon the arriving of the participant, the experimenter greets the participant, briefly introduces the research, and then presents a consent form (see Appendix B.) to inform the participant about the experiment-related information. The consent form explains the content, time length, potential benefits and risks of the experiment. It also emphasises that the participant's rights and willingness are respected and that the confidentiality of the information collected is guaranteed. Queries and doubts are encouraged from the participant and solved if any, and the participant is then instructed to sign the consent form if all the information presented is accepted.

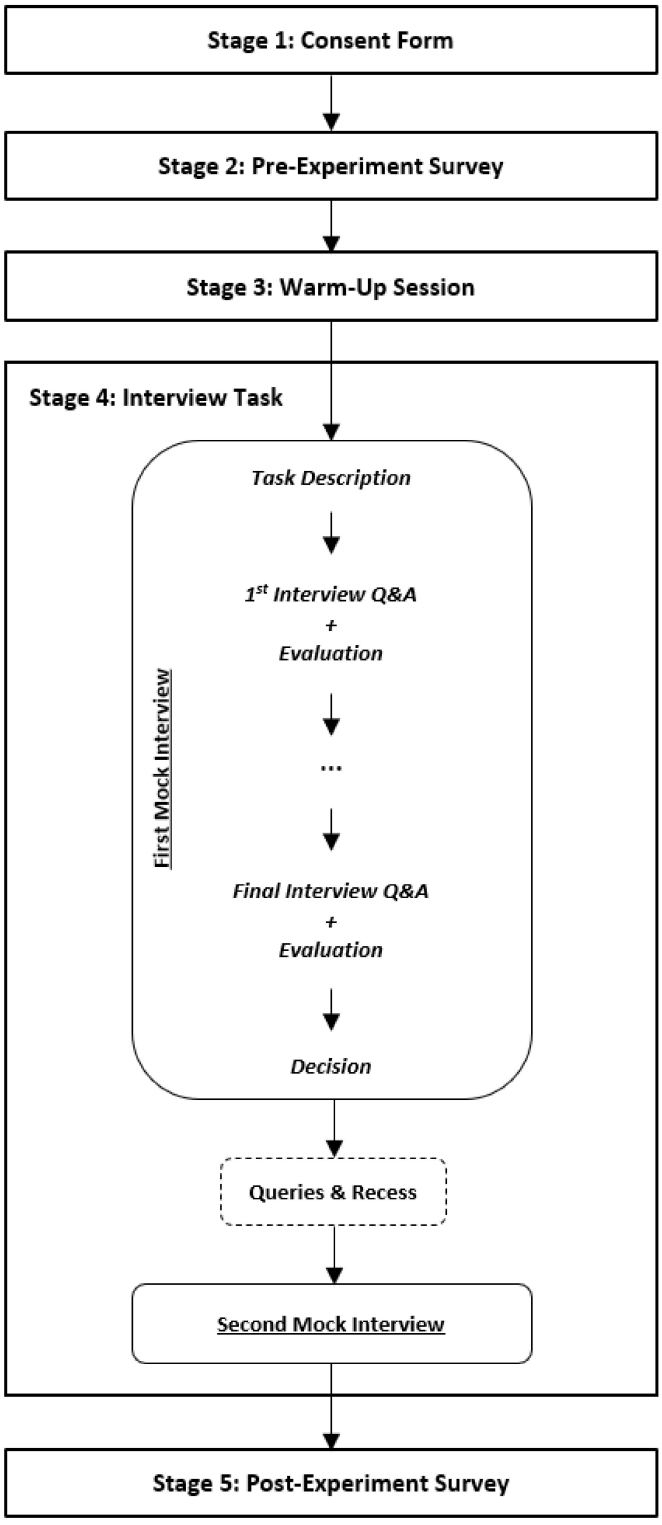


Figure 3.2: Experiment procedure

Stage 2: Pre-Experiment Survey

Following the signing of the consent form, a pre-experiment survey is carried out in order to collect the background information of the participant (see Appendix C.). The survey consists of eight questions, among which the first seven questions ask about the participant's age, gender, interview experience, and the nationalregion regarding the interview experience mentioned. The participant is asked particularly about the knowledge and experience regarding several elements of interview involved in the mock interview task, including the soft skills, types of interview questions, and types of rating scales. This is to ensure that the participant has a basic understanding of the interview task to be completed. The last question of the survey asks whether the participant knows about TAM, and a brief explanation with examples regarding the application of TAM in multiple fields are provided regardless of the participant's response. In particular, it is preferable that the survey takes a form of casual interview rather than questionnaire to ease the potential tension of the participant as previously discussed in Section 3.2.1.

Stage 3: Warm-Up Session

After finishing the pre-experiment survey while before starting the mock interview task, the participant is required to take part in a warm-up session that aims to help the participant better understand how to think aloud while completing a task. Specifically, the participant needs to work on a mathematical problem (see Appendix D.) while continuously speaking aloud own thinking process. The instructions listed in Table 3.4 are highlighted before the participant, and the experimenter then minimises the interaction with the participant throughout the warm-up session unless there is a significant pause in the participant's verbalisation. While it is possible that the participant cannot find the right answer of the mathematical problem, the suitability of the participant to the following mock interview task is judged according to the verbalisation skill. In other words, the participant is invited to participant in the mock interview task as long as the participant can effectively verbalise own thinking processes regardless of the performance on the mathematical problem. Once the completing the warm-up session, the

participant is asked by the experimenter

- whether the participant feels comfortable and natural about the process, especially concerning the think-aloud practice,
- whether the participant feels thinking aloud has affected own task performance,
- whether the participant has any doubt or concern about the experiment and,
- whether the participant feels like to take part in the mock interview task.

Stage 4: Mock Interview with TAM

The participants who are verified as possessing the ability of thinking aloud through the warm-up session and showing willingness to be involved in the mock interview task then get into the fourth stage of the experiment. In general, the participant is instructed to first goes through the task description that clearly explains the interview scenario and the details of the task, and then following the task description the participant plays the role of an interviewer who "interviews" and evaluates two applicants by listening to the recordings of interview Q&A.

The participant is first presented with a description of the interview task in written form (see Appendix E.), including background information about the company, job description, and task requirements. To help the participant get used to the interview scenario where only recordings of interview communication rather than real applicants are presented, the participant is instructed to act as one of the two interviewers who is expected to listen carefully to the conversation led by the other interviewer, take notes when necessary, and evaluate the qualification of the applicant interviewed. The participant is also informed that, while there are several applicants shortlisted for the job position, this interview task only involves two of them and the hiring intention (i.e., to accept, to accept with reservations, and to reject) chosen by the participant is not the final decision made on the applicants.

Once the participant feels clear about the mock interview task, a rating scale that contains the evaluation criteria of applicants is provided. Previous studies notice that whether detailed behaviour anchors are provided for the criteria listed in a rating scale will affect IDM (Jacobs et al., 1980; Green et al., 1993; Maurer, 2002; Klehe et al., 2008; Cambon & Steiner, 2015). To examine whether the rating scale type affects ICP, for example, by causing cognitive overload that hampers interviewers' thinking processes or providing additional information that assists their decision making, two types of rating scales are prepared in this research and randomly distributed to the participants, namely behavioural-anchored rating scale (BARS, see Appendix F.) and open grid rating scale (OGRS, see Appendix G.). The two rating scales include the same competencies and sub-competencies (i.e., attributes) to be evaluated with a 5-point scale, while BARS also presents a detailed descriptor is provided for each attribute.

After receiving all the experimental materials, the participant is allowed plenty of time to get familiar with the interview scenario and the task requirements. The participant is also encouraged to express potential concerns and uncertainties regarding the experiment procedure and materials before the mock interview starts, and the experimenter provides clear responses until the uncertainties get resolved. For example, participants may inquiry about whether they are supposed to give scores to the attributes during the interview process or after finishing all the interview questions, and the experimenter again emphasises that the participants should perform the task in the way they prefer.

Once the participant feels prepared for the scenario experiment, the mock interview process starts where the participant plays the role of an interviewer and separately evaluates two applicants according to the prepared recordings of interview communication. In the first mock interview (i.e., the first applicant interviewed), the experimenter plays the recording of first Q&A where an "interviewer" asks the first interview question and the applicant (A1) provides a response. During this procedure, the participant can either take notes or not according to own needs and preferences. As the recording of the first Q&A ends, the participant evaluates A1 according to the information received and

verbally expresses all the thoughts in minds, such as the feelings about A1, prompting questions to ask, and even the things irrelevant to the task itself. Once the participant finishes talking, the experimenter plays the recording of the next Q&A and executes the same procedure as that of the first Q&A. Such practice is repeated for each of the four interview questions, following which the participant is then required to rate A1 according to the criteria listed in the rating scale (unless the participant has already done so during the Q&A stage) while thinking aloud at the same time.

After completing the first mock interview, the experimenter checks with the participant whether there is any concern or doubt and if a break is needed before the second mock interview. When the participant feels prepared, the experimenter gives out a same set of rating scales used for A1 and carries out the interview of the second applicant (A2) in the same way as that in the first mock interview. After the second mock interview is completed, the participant is instructed to make a final decision of hiring intention while thinking aloud, then the experimenter collects all the materials used for the mock interview task.

Stage 5: Post-Experiment Survey

After finishing the whole mock interview task, a post-experiment survey is conducted for the participant's feedback on six aspects of the experiment (see Appendix H.). The main purpose of this survey is to ensure that the participants perceive both the mock interview itself and the way they work through the interview process are natural and representative of real interview scenarios, thus guaranteeing the validity of the data collected. Besides, an open question is also asked to encourage the participants to provide further suggestions on the whole experiment in case that there is any defect in the experiment design while not covered by the survey questions. Once the post-survey is completed, the participant is thanked with a small gift and advised to contact the experimenter if any concern about the experiment occurs.

To summarise, this part presents the experiment procedure developed with all the concerns and countermeasures discussed in the previous context taken into considera-

tion. The experiment procedure will then be tested through a pilot study to detect the opportunities of improvement.

3.2.3 Validating the Experiment Procedure through a Pilot Study

To test and revise the experiment procedure developed in the former step, two volunteers are invited to take part in a pilot study where they try out the whole experiment procedure and their performance as well as comments are recorded and analysed. These two volunteers have quite different backgrounds (see Table 3.5), especially regarding their experience as interviewer. Specifically, Volunteer 1 has worked as recruitment consultant for over ten years with sufficient experience of conducing employment interviews (i.e., experienced interviewer, EI), whereas Volunteer 2, has attended a few formal interviews within the year and is familiar with interview process (i.e., novice interviewer, NI). With such distinction in the volunteers' background, the pilot study is expected to examine the feasibility of the experiment procedure among participants with different levels of experience as employment interviewer.

The pilot study is carried out strictly following the experiment designed (see Figure 3.2) and the two volunteers work through the whole experiment quite smoothly. The

Table 3.5: Background information of the two volunteers in the pilot study

Background Information	**Volunteer 1**	**Volunteer 2**
Age	42	28
Gender	Female	Male
National Region (Interview experience)	UK	China
Experience Level (As interviewer)	Experienced	Novice

performance observed from and the comments produced by the two volunteers at each stage of the experiment are summarised in Table 3.6, which not only prove the feasibility of the experiment but also provide valuable information for the revision of the experiment procedure.

Table 3.6: Performance and comments of the volunteers throughout the pilot study

Experiment Stage	**Performance and Comments**	
	Volunteer 1 (EI)	*Volunteer 2 (NI)*
Pre-Experiment Survey		
Knowledge of interview	• Know well about interview-related elements and terms	• Has not heard about the term but understand the content
Knowledge of TAM	• Have not heard about the term, but can understand the method when the experimenter presents examples for explanation	
Warm-Up Session		
Task performance	• Starts with a random guess and then tests and narrows it down until getting the right answer	• Adopts multivariate linear equations and successfully works out the right answer with a few steps of calculations
Think-aloud practice	• Provide sufficient verbal data during the process and suggest that they clearly understand how to TA and feel confident to carry it out in the mock interview	
Mock Interview Task		
Task performance	• Feel easy to understand the task requirements and have no significant doubt	
	• Gives a score to each attribute and competency listed in the rating scale after the whole Q&A stage	• Rates the attributes throughout the Q&A stage, occasionally makes modifications to previously given scores, and finally evaluates the competencies after finishing all the interview questions
Think-aloud practice	• Continuously provide verbal reports as required and no significant pause occurs throughout the whole process	

(continued table)

Experiment Stage	Performance and Comments	
	Volunteer 1 (EI)	*Volunteer 2 (NI)*
Post-Experiment Survey		
Simulation level	• The experiment well simulates a real interview	
		• *"Usually, a CV of the candidate was provided and the candidate could be observed during interview, but these elements did not influence the interview process"*
Overall difficulty	• **Get "a little bit exhausted when it comes to the second half" due to the overall length of the experiment (i.e., about 2 hours)**	
		• **the BARS contains too much information**
Confidence in own decisions	• Confident about own evaluations and decisions	
		• **Expects for more information by asking follow-up questions**
Perceived impact of TAM	• No impact on the decisions made *"I feel it only differs from a real one in the sense that we (interviewers) do not discuss with each other during an interview. However, we do share our thoughts after that. So, in real practice, I would definitely have done something similar."*	• Slightly positive influence on the decisions made *"It encouraged me to go through each of the criteria and really think about it rather than just checking things off on the list. It did influence my process a little bit, talking out aloud as I went along, but in a positive way, because it's more thorough."*
Further advice	• **To shorten the length of experiment by cutting down the number of Q&A**	
	• **To provide an additional piece of paper for note taking**	

In general, the observations and comments from the two volunteers in the pilot study indicate that the overall experiment design is feasible in multiple aspects. First, not only EI but also NI without working experience as interviewer can understand the requirements of the mock interview and successfully work through the task while

providing rich verbal reports of own thinking process. This indicates a proper level of task difficulty (i.e., task difficulty) as well as a sufficient simulation of real-world interview (i.e., task representativeness) achieved by the experiment, which are again confirmed by the two volunteers' feedback provided in the post-experiment survey. Secondly, while the two volunteers have not heard about TAM ahead of the experiment, they can not only understand how the technique works but also carry it out in the warm-up session and the mock interview task. Particularly, both of them continuously provide verbal reports as required without no significant pause throughout the whole process (i.e., execution of TAM) and perceive little or only a slightly positive impact of TAM on their task performance, suggesting that TAM can be effectively embedded in interview scenario and carried out by interviewers (i.e., task type).

While the pilot experiment verifies the feasibility and validity of applying TAM to investigating ICP to a great extent, it also sheds a light on the opportunities to revise the experiment procedure from two perspectives. The first and the most crucial modification that needs to be made, according to the volunteers' performance and feedbacks, is to cut down the overall length of the mock interview task. Otherwise, participants may get exhausted and their task performance can be hampered by cognitive overload. Two significant criteria are applied here that (i) the applicants' responses should be informative enough that enable interviewers to make and justify their judgements and decisions, and (ii) the whole interview process should still be natural and smooth. Thus, the overall interview length is shortened by picking up the four Q&A that directly evaluate the four most essential and relevant soft skills (leadership skills, organisation skills, problem-solving skills, stress management skills) to the job position as suggested by the three professional interviewers, while the Q&A that are either less informative or less related to the target soft skills are excluded. The second revision, detailed but still essential, is to offer participants sufficient papers to take notes during mock interview as suggested by EI.

As for the comment made by NI that the BARS contains too much information, however, no corresponding adjustment is made to the experiment, because BARS has

been widely applied in real-world practice and the rating scale utilised in this research is verified by the professional interviewers as well as EI. NI also suggests an expectation for more information by asking follow-up questions to strengthen the confidence in own decisions. No change is made particularly regarding this comment, because it is noticed that NI does verbally report the prompting questions he would like to ask during the mock interview. Additionally, NI also explains that he is expressing an expectation for the applicants' further responses rather than suggesting a modification in the experiment procedure when queried by the experimenter.

Overall, the experiment procedure developed is tested through a pilot study where two volunteers with disparate backgrounds take part in, especially in terms of their previous experience as employment interviewer. The volunteers' performance is observed and their comments on the experiment procedure are collected, of which a detailed report as well as discussion are presented. The results verify the overall feasibility of applying TAM to IDM research while also indicating the chances to refine the experiment procedure.

3.2.4 Participants' Feedback on the Experiment

After being revised (i.e., screening the Q&A presented to shorten the overall length of interview task and offering participants sufficient papers to take notes), the experiment which still complies with the procedure shown in Figure 3.2 is applied for data collection, and the feedback on the experiment is collected from the 29 participants that complete the mock interview task (see Section 3.3 for detailed information of the participants). As a further verification of the feasibility of applying TAM to an interview scenario to investigate ICP, the outcomes of the post-experiment survey are summarized in Table 3.7.

Specifically, the averaged simulation level as rated by the participants is 4.86 ($SD = 0.44$), indicating an agreement that the experiment well simulates a real interview scenario. The scores given by all the participants to both the length of experiment

and task difficulty are 5.00, suggesting a proper level regarding these two dimensions. The participants also indicate a high level of confidence (i.e., 4.34) in their evaluations and decisions, where two thirds of participants ($N = 19$) choose a score of 4 rather than 5 for various reasons, such as an expectation for more information, a lack of own experience as interviewer, or simply a preference not to give the highest score. As for the impact caused by TAM, most participants ($N = 20$) perceive little influence caused by TAM on their decision making, while the other nine suggest that TAM results in a slightly positive impact on their performance by helping them "think properly" and "adding quality to the decision made". Nevertheless, three participants express their self-preference regarding either the types of interview questions asked or the rating scale utilised. Specifically, one participant prefers to ask more open-ended questions and the other two prefer not to use BARS. Considering the prevalence and the proved reliability and validity of both the questions and BARS, this research suggests that the data validity is not affected and the protocols collected can be investigated to study ICP regarding their decision-making process.

To summarise, the results of the post-experiment survey indicate that the participants perceive the experiment procedure as (i) effectively simulating a real interview, (ii) having a proper overall length where the completeness is achieved without exhaust-

Table 3.7: Descriptive statistics of the participants' feedback on experiment

Dimension	***M***	***SD***	***Max***	***Min***
Simulation Level (1-5: highly simulated)	4.86	0.44	5	3
Length of Experiment (1-5: appropriate)	5.00	0.00	5	5
Task Difficulty (1-5: appropriate)	5.00	0.00	5	5
Confidence in Own Decisions (1-5: confident)	4.34	0.55	5	3
Perceived Impact of TAM (-5-0-5: strongly positive impact)	1.03	1.59	0	4

ing the participants, (iii) involving an interview task at a proper difficulty level which can be effectively completed by the participants and produce rich information of their cognitive processes, and (iv) the execution of TAM does not impose significant influence on the participants' performance on the interview task.

3.3 Data Collection

The experiment procedure developed and validated in Section 3.2 is applied for data collection, of which the details are reported in this section, including the information of the participants, experiment settings and procedure as well as the data collected.

3.3.1 Participants

As the purpose of this research is to inclusively examine the possibility of ICP regarding their decision making rather than to focus on a certain group of interviewers, participants with diversified backgrounds are involved in the experiment. Specifically, the data collection is carried out in both the United Kingdom and China while basically in two major cities (i.e., Glasgow, UK and Sichuan Province, China) because the experiment must be conducted in a face-to-face manner so that the experimenter can monitor the procedure and record the verbal protocols to guarantee the quality of the data collected. Invitations to the experiment are widely sent out through multiple channels, such as LinkedIn, organisations, universities, emails, and networks.

To enhance the randomness of sampling, all the respondents are welcomed to take part in the experiment as long as they meet the two indispensable criteria, namely interview experience (i.e., whether they possess the ability to complete the interview task) and verbalisation skills (i.e., whether they can effectively think aloud when working on a task). The former is ensured by confirming with the respondents that they either have previous experience as employment interviewers (i.e., experienced interviewer) or have at least attended formal interviews of a similar kind and are aware of how an

Table 3.8: Backgrounds of the participants involved in data collection

Background Info.		**No. of Participants (%)**
Age	18 – 30	17 (58.62%)
	31 – 40	3 (10.34%)
	41 – 50	4 (13.79%)
	Over 50	5 (17.24%)
Gender	Male	6 (20.69%)
	Female	23 (79.31%)
National Region (Interview experience)	UK	14 (48.28%)
	CN	15 (51.72%)
Experience Level (As interviewer)	Experienced	11 (37.93%)
	Novice	18 (62.07%)

Note. The total number of participants is 29.

interview process works (i.e., novice interviewer). Respondents with the interview experience needed are invited to the experiment where they are required to work through a warm-up session to test if they are able to continuously think aloud when completing a task. After careful examinations, a total of 29 respondents are proved possessing both interview experience and verbalisation skills and are thus invited to take part in data collection as participants. The demographic information of these 29 participants is shown in Table 3.8.

3.3.2 Experiment Settings and Procedure

The face-to-face experiment is conducted in a quiet room with each participant (see Appendix A.). Facilitates and the environment are examined in advance to minimise

disturbance (e.g., noise) and discomfort (e.g., chair, temperature). The experiment follows the procedure shown in Figure 3.2 and lasts between 60 and 90 minutes, depending on the quantity of thoughts expressed by each participant when thinking aloud.

After greeting the participant, the experimenter sits next to the participant and starts a short casual talk with a brief introduction of the research to ease the potential nervousness of the participant. The participants are made aware of the neutrality of the research that their thinking processes are conceptually studied without any judgement or evaluation. A consent form (see Appendix B.) is read and signed before they proceed with the experiment procedure.

If the participant agrees to participate in the experiment after checking the consent form, a pre-experiment survey (see Appendix C.) is then conducted to gather the information of the participant's background as shown in Table 3.8, following which a brief explanation of TAM is provided. After that, a warm-up session is scheduled that asks the participant to work on a mathematical problem (see Appendix D.) while verbally expressing their concurrent thoughts. At the beginning of this practise session, several experimental instructions (see Table 3.4) are emphasised and explained by the experimenter to help the participants better understand the requirements of TAM.

Once the verbalisation skills of the participants are ensured, the experimenter introduces the mock interview task where the participants act as the interviewer and evaluate two applicants by listening to the audiotaped interview communication prepared in advance (see Section 3.2.2 for details). At the start of the mock interview task, the participants are presented with task materials and allowed plenty of time to get familiar with both the interview scenario and the task requirements. During the interview process, the participants verbally express their current thoughts after listening to the recording of each Q&A, whether they are comments on the applicant, following-up questions to ask, or something irrelevant to the interview task itself. When finishing the four Q&A of an applicant, the participants give comprehensive evaluation to the applicant and choose their hiring intention both on the rating scale and through

verbal reports. Besides, the participants are checked with their feelings and potential queries and asked whether a short break is needed before proceeding to the second mock interview. When the participants are working on the mock interview task, the experimenter intends to minimise the interaction with the participants unless they stop thinking aloud, which does not happen for all the 29 participants. This indicates that the possibility of interference and bias caused by the experimenter on the participants' task performance is limited to the minimum.

As soon as the participants finish the two mock interviews, a post-experiment survey is executed to collect in-depth feedback on the experiment (see Appendix H.), after which the participants are thanked with a small gifted and encouraged to contact the experimenter if they have any concern about the experiment.

3.3.3 Data Collected

Two groups of data are collected from the experiment for different aims of data analysis. One is the audiotaped think aloud protocols produced by the participants during the mock interview task, which is processed and analysed to explore the elements and characteristics of ICP. The other group of data refers to the information of the four contextual factors collected to examine their potential associations with ICP, namely interviewers' gender, experience level, national region and type of rating scale used in the interview task. Specifically, it has been found in previous research that interviewers' gender (Parsons & Liden, 1984; Raza & Carpenter, 1987; Chapman & Rowe, 2001), experience level (Hess, 2013; Roulin et al., 2015; Frieder et al., 2016) and type of rating scale used during interview (Fay & Latham, 1982; Maurer, 2002; Klehe et al., 2008; Melchers et al., 2011) have an impact on IDM. As for the national region of the interviewers' experience, while having not yet been investigated straightforwardly in the existent interview studies, the factor is introduced to the data analysis considering that differences in culture (Sanchez-Burks et al., 2006; Manroop et al., 2013) and employment laws (Williamson et al., 1997; Morgeson et al., 2008) may affect IDM.

3.4 Data Analysis

The data analysis in this research refers to an investigation into the recorded verbal protocols which are produced by the participants as the interviewer (interviewers hereafter) during the mock interview task. The analysis consists of two main stages that are carried out sequentially, including protocol processing and protocol analysis. During protocol processing, the recorded verbal protocols collected are processed into an analysable style and the elements occurring in ICP are identified and defined. Later through protocol analysis, the coded protocols are further analysed to explore the characteristics of ICP and how interviewers' judging behaviour can be monitored at a cognitive level. Finally, the associations between the four contextual factors and ICP in terms of the ICP characteristics and the measurements of judging behaviour defined in the protocol analysis.

3.4.1 Protocol Processing

As the preparation stage of data analysis, the recorded verbal protocols are first processed into an analysable style (i.e., segmented and coded protocols). A standard procedure of verbal protocol processing (see Figure 3.3) is commonly adopted in TAM-based studies (Van Someren et al., 1994; Davison et al., 1997; Charters, 2003; Heerkens & Van Der Heijden, 2005; Ericsson, 2006; Eccles & Arsal, 2017), which is also consulted as the guideline in this research. As can be seen in Figure 3.3, a psychological model of subjects' task performance is often brought into the procedure to help develop a coding scheme that predicts subjects' cognitive processes, while such a model of ICP is absent in existent studies on IDM. Therefore, the procedure is partly adjusted in the sense that a group of randomly selected verbal protocols (i.e., sample protocols) and the knowledge of interview process (i.e., task information) are utilised as a substitute of "psychological model".

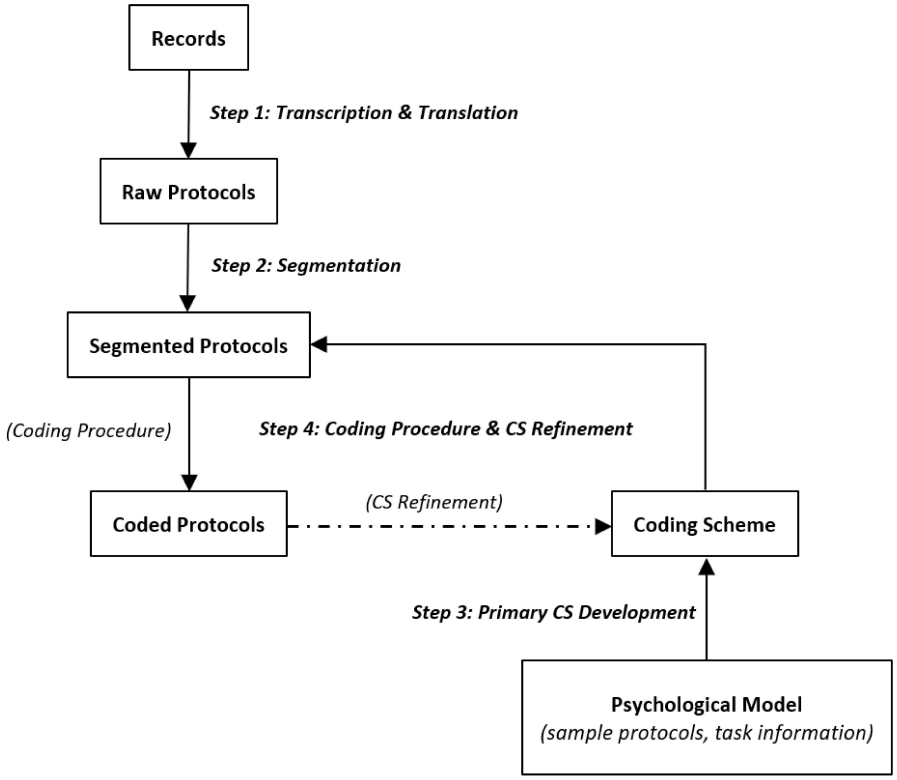


Figure 3.3: Procedure of protocol processing

Step 1: Transcription and Translation

Similar to all the other research applying TAM, the recorded verbal protocols are first transcribed into written form (i.e., transcription). The experimenter completes all the transcription work for the protocols collected from the 29 interviewers, and each transcript is verified and revised by the corresponding interviewer who has produced that particular verbal report to assure that there is no misunderstanding (see Appendix I. for a sample part of raw protocol). An additional work of translation is carried out within this step as well where the protocols collected from Chinese interviewers are translated into English by the experimenter and then proofread by the bilingual volunteer who has previously assisted in the translation of the experimental materials to ensure that not only the meaning but also the structure of the content are maintained to minimise the impact of language difference on data analysis (see Appendix J. for

a sample part of translated protocol). Secondly, the content of the protocols is divided into segments (i.e., segmentation) according to the criteria established, like the pauses and the linguistic structure identified in the protocols and shifts in interviewers' attention to different information.

The think-aloud protocols collected turn out to be extensive with a total count of 111,741 words of all protocols. Specifically, length of each protocol (i.e., the verbal protocol produced by an interviewer during one single mock interview) ranges from 553 to 4,920 words ($M = 1927.57, SD = 1075.13$), and the total length of the two protocols produced by the same interviewer ranges from 1,515 to 8,780 words ($M = 3853.14, SD = 2076.31$). Such richness can be expected to enhance the contribution of the data collected and the data analysis in later stage to the knowledge of ICP regarding decision making.

Step 2: Segmentation

In the second step of protocol processing, the content of each protocol is segmented into separate and comparable units "to facilitates identification and further analysis of the parts of the text and of the protocols" (Joseph & Patel, 1990, p.33), and the segments are then labelled with continuous numbers for the convenience of discussion. The segmentation procedure is usually conducted following certain criteria, of which the details vary with the tasks investigated and different purposes of research (Ericsson & Simon, 1984; Fonteyn et al., 1993; Cumming et al., 2002; Nielsen et al., 2002; Li & He, 2015).

When developing the segmentation criteria in this research, three terms are defined first, namely information type, operator, and cognitive action, which not only serve as the indicators of segmentation but also form the foundation of the following protocol analysis. To be specific, decision-making is in essence an information-processing procedure (see Chapter 2) and interviewers as the decision maker of employment interview gather and deal with information in certain manners and steps until final decisions are made. Each of these steps consists of two major elements, including the information

processed (i.e., either received from the environment or retrieved from LTM) and the operator adopted by the interviewer to process a piece of information (Fonteyn et al., 1993). When scanning the verbal protocols, it can be noticed that the information processed by the interviewers covers a wide range of types, such as task-related information (e.g., task requirements, job description, interview questions, rating criteria), information generated from applicants' responses (e.g., applicants' work performance, features of responses), personal sources of information (e.g., personal experience, knowledge, beliefs), and the information produced by the interviewers themselves (e.g., comments, scores, prompting questions, inferences). Hence, the information involved in ICP is identified and categorised into different information types. Particularly, the information is distinguished into general types without specifying the exact content of the information processed by the interviewers when discussing the elements of ICP, because this research aims to inclusively explore ICP during decision making regardless of the specific contexts of an interview scenario (e.g., industry, company, job position, target competence, evaluation criteria).

All the types of information are processed by the interviewers through various operators, like reading, interpreting, comparing and prompting. However, it is noticed that the knowledge of ICP provided by the operators themselves is quite limited. Meanwhile, it is noticed that the same operator can be executed to different types of information, while the same type of information can be processed by different operators as well. For instance, an interviewer may "evaluate" (i.e., the same operator) either an applicant's "work performance" or "the features of a response" (i.e., different types of information), and an interviewer may either "revise", "describe" or "explain" (i.e., different operators) own previous rating scores (i.e., the same type of information). To enhance the clarity of discussion when investigating and describing the elements of ICP, the concept of cognitive action is introduced to specify different combinations of operator and information type recognised in ICP (i.e., "evaluate an applicant's work performance" and "evaluate an applicant's response" are two different cognitive actions, "give a score" and "revise a score" are also two different cognitive actions).

Based on the concept of cognitive action, the first criterion of segmentation in this research is developed that

> *Criterion (i):* A segmentation will be marked if there is a clear alter in the interviewer's cognitive action.

Taking part of the verbal protocols provided by Interviewer 7 as an example (see Table 3.9), four segmentations are marked according to the alteration in the interviewer's cognitive actions. Specifically, except that the segmentation marked between Segment 52 and Segment 53 is due to the interviewer's focus on different piece of information, the other three segmentations are derived from the alterations in both information types and the operators adopted (i.e., Segment 51 - make a suggestion; Segment 52 - represent the applicant's response; Segment 53 - represent the applicant's response; Segment 54 - characterise the applicant's work performance; Segment 55 - make an inference).

Nevertheless, it is possible in some cases that interviewers implement a single operator with a great bunch of information involved, such as recalling own experience as shown below (see Table 3.10) or reading a large part of experimental materials (not found in the protocols). While such content may consist of more than one sentence and

Table 3.9: Sample protocol of segmentation (1)

No.	Segment
51	He could have given a little bit more about his particular role in that case,
52	because what he said was that he had an overview, he got the big picture, he understood what the issues were.
53	He also identified that the original plan could not be achieved within the time scale and the modifications would have to be made.
54	And his role was pretty much to go in particular with that of the team leader,
55	so the communication between him and the leader was good.

Table 3.10: Sample protocol of segmentation (2)

No.	Segment
73	It also reminds me of my first job. I was responsible for organising an opening ceremony that would go live. Everything was fine in the rehearsal, but you know, unexpected situations always happen. And I was really nervous at that time and could hardly control my temper.
74	So I think this candidate, according to his story, I don't think he has performed like a manager or a leader, I didn't quite get that.

plenty of information, it is suggested that such a bunch of information as presented in Segment 73 should be treated as a whole and the segmentation should be marked both before and after it rather than anywhere else within this part of the content. Hence, no segmentation is marked within Segment 73, which involves the same operator (i.e., associating) without an alteration in information type (i.e., own experience). In contrast, a segmentation is marked between Segment 73 and 74 where the cognitive action of the interviewer turns from 'associating with own experience' to 'making an evaluation'. A corresponding criterion is proposed that

> *Criterion (ii):* A segmentation will be marked both before and after the interviewer reads a bunch of information from the experimental materials, represents a bunch of information from the applicant' response, or narrates an excerpt of personal experience.

Besides, considering the possibility that an interviewer keeps silence for 5 seconds or more (although not found in this research) and is reminded by the experimenter to keep talking, a criterion commonly applied in TAM-based studies is that

> *Criterion (iii):* A segmentation will be marked if there is a pause of 5 seconds or more.

Last but not the least, there are also some sentences that are either modal particles (e.g., "Umm. . . ") or incomplete, which are usually not involved in data analysis (Joseph

& Patel, 1990; Fonteyn et al., 1993; Van Someren et al., 1994) unless a particular interest is in these observations. Thus, the corresponding criterion of segmentation is as follow

> *Criterion (iv):* No segmentation will be marked for modal particles and incomplete sentences.

It is found that most content of the protocols can be segmented according to these criteria, whereas some parts of the protocols contain more than one cognitive action and cannot be easily segmented into separate units. For example, while two cognitive actions can be identified in Segment 30 as shown in Table 3.11, including 'comparing with another applicant' and 'identifying a characteristic of the applicant's response', it is not appropriate to split the content into two segments because, once segmented, no cognitive action can be identified within the content "both of them" alone. However, since a count of coded cognitive actions will be generated though the coding procedure in the following steps and such count rather than the number of segments will be used for data analysis, it is suggested in this research that no significant impact will be caused by this small part of "unsegmented" content on the data analysis.

Overall, the segmentation procedure results in a total number of 5,814 segments, ranging from 60 to 381 segments produced by an interviewer throughout the two mock interviews, or from 25 to 210 segments produced by an interviewer within one single mock interview process.

Table 3.11: Sample protocol of segmentation (3)

No.	Segment
30	when we're talking about distressing, both of them are very much talking about their personal life.

Step 3: Development of a Primary Coding Scheme

When exploring a subject's cognitive process when working on a task, the most primary and indispensable step is to identify the elements of interest that occur in the subject's cognitive process (e.g., different categories of information type and operators). A common practice to achieve this objective is to develop a coding scheme that presents the elements identified, according to which the segmented protocols can be coded for a more in-depth investigation into the subject's cognitive process (Harte et al., 1994; Chi, 1997; Cumming et al., 2002; Nielsen et al., 2002; Eccles and Arsal, 2017). In the standard procedure of protocol analysis, a coding scheme is usually developed according to the psychological model of subjects' performance on the task investigated, which is less feasible in this research since a psychological model of ICP regarding decision making is lacking. Hence, this part of the procedure is revised in the sense that a group of randomly selected protocols (i.e., sample protocols) representing ICP and the knowledge of interview process (i.e., task information) are set as the starting point of coding scheme development.

On the basis of the information provided by both the sample protocols and task information, predictions of the elements occurring in ICP are made and listed in a primary coding scheme (see Table 3.12) in the form of cognitive actions (i.e., operator + information type). An explanatory example, if identified in the sample protocols, is also provided for each of the cognitive action. As for the cognitive actions anticipated from the general knowledge of employment interview while not recognised in the sample protocols, a label of "anticipated" is marked and these cognitive actions remain to be checked when the rest protocols are brought into analysis. Correspondingly, this primary coding scheme will be revised as well in the following stage.

Step 4: Coding Scheme Refinement and Coding Procedure

The fourth step aims to complete two major tasks, including refining the coding scheme and coding all the segments in the protocols according to the coding scheme, from which the coded protocols generated can be analysed to study ICP. Since the

Table 3.12: Primary coding scheme

No.	Cognitive Action	Example
1	Read or represent experimental materials (i.e., job description, interview questions, rating scale)	*So, "Leadership and conflict management, treats everyone with dignity, respect, and fairness"*
2	Represent the content of the answer	*... he talked about his own scheduling methods and that if people are looking for some flexibility and coordinate.*
3	Identify a particular characteristic of the response	*That wasn't really about handling stress* (relevance) *He used the word "we" a lot...*(language)
4	Make an inference (NOT judgment) as a supplement to the answer (e.g., situation, action, result)	*It suggested that he'd been making the decision on whether or not people got the opportunity to provide the feedback.*
5	Utter specific follow-up questions	*A question I would ask would be how do you deal with priorities on a daily basis, what do you actually do.*
6	Articulate a (positive/negative) evaluation or rating score	*So, the stressful example one has been quite weak (response)* *And I feel like he has been maybe a little bit arrogant and close-minded to other people's opinions. (applicant)*
7	Make or revise a hiring decision	*I would like to accept him.*
8	Match the question or (part of) the answer with the dimensions/attributes in the rating scale	*I'll take that to be work plan and implementation, coordination.*
9	Associate with own knowledge/experience/ preference/beliefs, etc.	*As an HR you would want someone in a project management role to be able to manage the individuals and the team...*
10	Compare or associate with the applicant's previous response(s)	*Yeah, I think this answer was less successful than the first one.*
11	Compare or associate with previous applicant(s)	(anticipated)
12	Summarise or recall previous judgements	(anticipated)
13	Reflect on own personal cognitive status/habits/ preferences, etc.	(anticipated)
14	Point out an own knowledge gap	*I don't know much about what he called "syndicated loans".*

coded protocols form the conceptual and elementary base of protocol analysis, it is vital to ensure that the coding scheme developed meets four criteria of quality, namely completeness, grain size, unambiguousness, and context independence (Fonteyn et al., 1993; Van Someren et al., 1994; Charters, 2003; Heerkens & Van Der Heijden, 2005). Specifically, completeness requires the coding scheme to comprehensively cover the target items occurring in the subjects' cognitive process (i.e., cognitive action in this research). Grain size means the size of codes defined must correspond to that of segments in protocols so that one segment can be labelled with single code. As for unambiguousness, it stresses that the codes must be defined clear enough to be understand and used by outsiders, while the context independence indicates that each code defined should be able to be recognised in the protocols without the context in which it appears.

In order to reach the four standards of the coding scheme as well as to guarantee the objectivity of the coding procedure, two PhD researchers without particular knowledge of and interests in this research are invited into the work of coding scheme refinement and coding procedure as co-coders. The basic logic of coding protocols and refining the coding scheme is that different coders should separately code the same pool of protocols, solve the disagreements, revise the coding scheme, apply the revised coding scheme to another pool of protocols and repeat this procedure until an acceptable level of inter-coder agreement is achieved. The inter-coder agreement is defined as the proportion of segments labelled by different coders with consistent codes, of which a satisfying level is set as 90.0% in most cases (Harte et al., 1994; Chi, 1997; Li & He, 2015; Eccles & Arsal, 2017).

Following this guideline, the protocol coding and coding scheme refinement is carried out according to the procedure shown in Figure 3.4. Besides, a coding form is designed as well (see Appendix K.) to facilitate the coding process, which consists of the segmented protocols, the serial number of each segment to help promptly locate the concerned segment during coding and analysing procedure, and space to mark the code assigned to each segment. In particular, the coders are required to distinguish whether each segment of the interviewer's cognitive process is focusing on the applicant's work

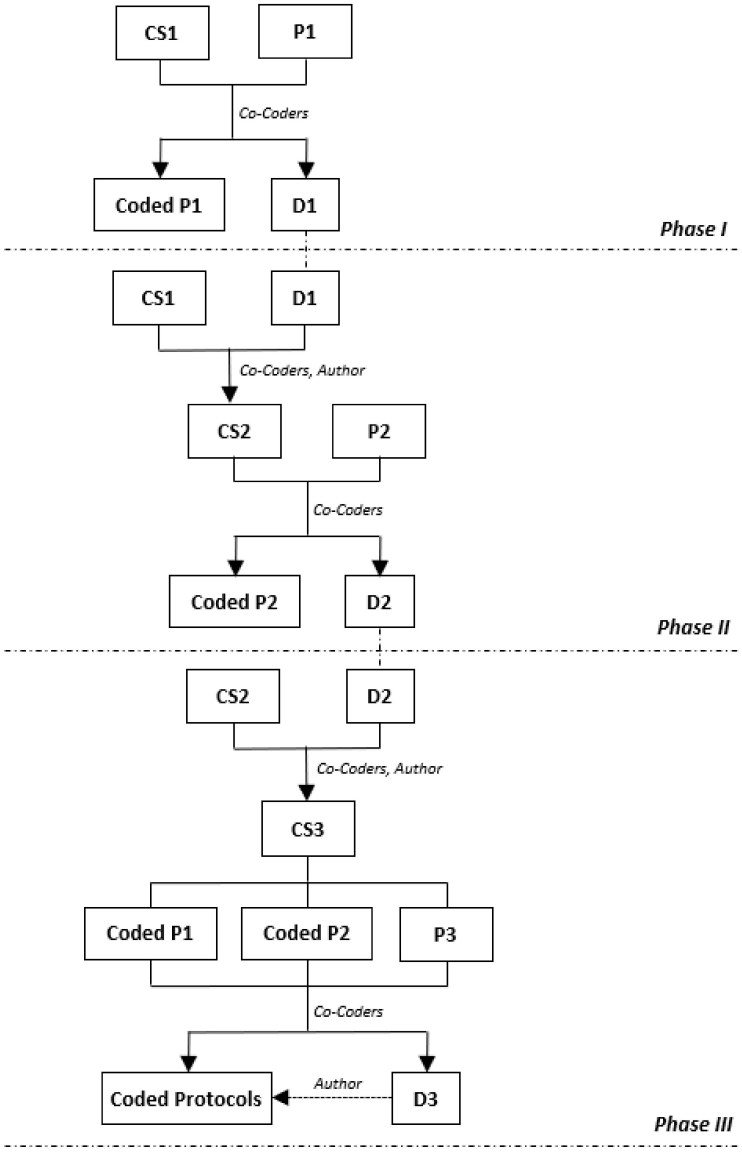


Figure 3.4: Coding procedure and coding scheme refinement

performance (i.e., "performance"), interview responses (i.e., "response"), or neither of them (i.e., "other", like experimental materials or environmental circumstances), which are defined as interviewers' different evidential focus in this research.

1) Phase I

In Phase I, the two co-coders (i.e., A & B) are first instructed about how to code the protocols with a coding scheme. Particularly, to ensure the context independence of the coding scheme, only minimal information about the research purpose is provided to the co-coders. For the same consideration, the protocols are divided into small episodes that contain no more than five segments and then mixed to minimize the information the co-coders can perceived about the context (Van Someren et al., 1994). The same group of protocol episodes (P1) are distributed to the two co-coders to code independently with the primary coding scheme (i.e., CS1) developed.

The inter-coder agreement level achieved in Phase I is 78.1%. According to the results in the coded protocols (i.e., Coded P1), all the segments are assigned with at least one codes, which means the codes defined in CS1 sufficiently cover all the cognitive actions occur in the current group of protocol episodes, thus indicating the completeness of CS1 to a certain degree. However, ambiguousness of CS1 is noticed that leads to the disagreements between the two co-coders (i.e., D1) where they label some of the segments with different codes. This indicates that, after solving the disagreements, corresponding adjustments need to be made to clarify and distinguish the definitions of these codes.

2) Phase II

In Phase II, the experimenter works together with the two co-coders to solve D1, from which it is found that the discrepancies in their coding are mainly caused by the overlap among several codes' definitions and the vagueness in the definitions of several other codes. Thus, CS1 is revised according to the discussion (i.e., CS2), which is then applied to another randomly selected group of protocol episodes (i.e., P2). A higher level of inter-coder agreement level (i.e., 88.3%) is reached with CS2 compared with that in Phase I, whereas the disagreements (i.e., D2) existing in the coded P2 suggest that the coding scheme needs further revision.

3) Phase III

When discussing the disagreements occur in Phase II, it is noticed that D2 are basically derived from the occurrence of several cognitive actions that are not observed in P1 and are thus not included in CS2. Therefore, as a start of Phase III, refinements are made to CS2 by including the newly identified cognitive actions, which leads to CS3. Since the application of CS2 has already resulted in a relatively high level of inter-coder agreement (i.e., 88.3%), it can be expected that CS3 will achieve a satisfying result by precisely covering most possibilities of the cognitive actions occur in ICP. Thus, the two co-coders are instructed to code all the rest protocol episodes (i.e., P3) with CS3, from which an inter-coder agreement level of 91.3% is reached. At the same time, considering that the coding results based on CS3 meets the target level of inter-coder agreement, CS3 is regarded as the finalised coding scheme (see Table 4.1 in for details) and applied to P1 and P2 to check and revise the codes assigned.

Comparing the two versions of coding scheme, four major changes are made. First, with more protocols investigated beyond the sample protocols selected when developing the primary coding scheme, the categories of cognitive actions identified and listed are enriched. For example, it is noticed that the interviewers may suggest their follow-up actions (e.g., "Now, I think at this stage, it's probably a good time to compare him to the previous candidate"), which are not found in the sample protocols. At the same time, the definitions of several codes are modified not only by using more precise language but also through the integration of similar cognitive actions. For instance, the interviewers are found occasionally suggesting improvements in an applicant's work performance or interview responses (e.g., "You need to maybe relax a little bit and let other people take responsibility") rather than directly making a positive or negative evaluation, which are thus integrated into the same code of "articulate a (positive/negative) evaluation or rating score" instead of defining them as different codes to avoid redundancy and overlap.

Secondly, while the ultimate goal of all the cognitive actions performed by an inter-

viewer is to make a final decision, it can be perceived that these cognitive actions serve different immediate purposes, which is defined as information-processing strategies in this research. Following the work of Cumming et al. (2002) that investigates essay rating through TAM, the cognitive actions are further classified into different groups of information-processing strategies to investigate the structure of ICP at a higher level. Thirdly, several types of comments recognised in the interviewers' protocols are categorised as "irrelevant comments", including the comments that are incomplete expressions, vague in meaning regarding the decision-making task, or concerning about the experiment design which are less likely to happen in a real interview scenario. While these irrelevant comments are not the main focus of this research, they are detected and listed in the coding scheme with explanatory examples to inclusively present the possibilities of ICP. Besides, sufficient explanatory examples are provided for each of the cognitive actions, especially those anticipated from the knowledge of employment interview while not recognised in the sample protocols. To assist the coders to better match the segments with the codes, more than one example is presented for some of the codes to include the different possibilities identified. Taking "identify a particular characteristic of the response" as an instance, examples where the interviewers comment on various features of applicants' responses (e.g., relevance, comprehensiveness, detailed level, tone, and wording) are listed and labelled separately.

Prevalence and Sequence of the Elements Occurring in ICP

On the basis of the coded protocols, the prevalence of the elements occurring in ICP is investigated first. Since it is noticed that each interviewer focuses on all the three types of information (i.e., performance-related information, response-related information, and other information) during both interviews, only the occurrence of each cognitive action and information-processing strategy is checked to examine how commonly they are utilised among the interviewers (i.e., prevalence of a cognitive action/information-processing strategy). In other words, analysis is carried out to investigate whether these cognitive actions and information-processing strategies are widely adopted by most of the interviewers or only by few of them. Therefore, a

corresponding measurement is developed as follow:

$$P(X) = \frac{No.\ of\ Interviewers\ Adopting\ X}{Total\ No.\ of\ interviewers} \tag{3.1}$$

where X refers to a particular category of cognitive action or information-processing strategy. The value of prevalence ranges from 0 to 1, where a higher value implies that the cognitive action or information-processing strategy is executed by more interviewers (i.e., higher prevalence) and 1 means that all the interviewers perform the cognitive action or information-processing strategy.

To obtain the data of "the number of interviewers adopting X", the occurrence of the cognitive actions in each of the verbal protocols is checked first. If a category of cognitive action occurs for at least once in a protocol, that particular cognitive action is marked as 1 for that protocol, otherwise 0. The results are then integrated to examine if an interviewer adopts a particular category of cognitive action or not. Specifically, if a category of cognitive action occurs in at least one out of the two protocols produced by an interviewer, that cognitive action is marked as 1 for the corresponding interviewer, otherwise 0. Afterwards, the total number of interviewers marked as 1 for a cognitive action (i.e., the number of interviewers utilising the cognitive action) is counted to represent the number of interviewers adopting that particular category of cognitive action. The same procedure is repeated for each category of information-processing strategy as well.

Additionally, while the cognitive actions and information strategies are detected and summarised, no sequence is presumed in which they may occur throughout the interviewers' thinking processes during decision making. Hence, the coded protocols produced by different interviewers are examined and compared to check if particular sequence of the cognitive actions and information-processing strategies can be recognised, according to which a detailed procedural model of ICP during decision making can be developed.

Overall, two significant outputs are generated within the stage of protocol processing. One is the finalised coding scheme that summarises the elements occurring in ICP. In particular, the prevalence and sequence of the cognitive actions and information-processing strategies among the interviewers are investigated as well. The other output of protocol processing is the coded protocols acquired, which are brought into the next stage of verbal protocol analysis.

3.4.2 Protocol Analysis

A multi-dimensional analysis is conducted into the coded protocols obtained through protocol processing, from which further findings of ICP during decision making is generated from three perspectives, including characterising ICP based on the elements identified (RQ2a-d), monitoring interviewers' judging behaviour at a cognitive level (RQ3a-d), and examining the associations between ICP and the four contextual factors of interest (RQ4).

3.4.2.1 Characterising ICP based on the Elements Identified

Identification of ICP Characteristics

Several features of the verbal protocols can be easily noticed during protocol processing, indicating the possible dimensions that can be defined to characterise ICP. First, the most notable difference existing among different protocols is the overall protocol length, which to a certain extent reflects how cognitively active an interviewer is when completing the interview task (i.e., cognitive activeness).

Secondly, the ICP characteristics identified regarding the categories of cognitive actions and information-processing strategies adopted by the interviewers are explored. While cognitive actions are inclusively recognized and defined from the 58 protocols produced by 29 interviewers, it is noticed during protocol processing that individual interviewers do not necessarily adopt all kinds of cognitive actions. Such observation

leads to queries about the categories of cognitive actions and information-processing strategies performed by interviewers from three perspectives, including the diversity of cognitive actions occurring in individual ICP (i.e., cognitive diversity) and the difference in the frequency of each cognitive action (and information-processing strategy) performed by the interviewers (i.e., frequency of cognitive action/information-processing strategy).

In addition to the characteristics concerning the cognitive actions and information-processing strategies adopted by interviewers, the extent to which they focus on different kinds of clues during decision making is another significant characteristic worth investigation. Fundamentally, information itself plays a fundamental role in decision making (Keller & Staelin, 1987) it has also long been a core research aim of many studies to explore what information provided by applicants are taken into consideration by interviewers when making hiring decisions (Middendorf & Macan, 2002; Van Dam, 2003; Nikolaou, 2011). Nevertheless, few studies mention interviewers' interest in the features of applicants' responses, which also draw attention from interviewers during decision making according to the observations in protocol processing. In particular, it is found that most coded actions can be classified as either focusing on performance-related information (i.e., information related to applicants' work performance as stated in their responses) or response-related information (e.g., the features of applicants' responses as perceived by interviewers, like comprehensiveness, relevance, tone), while the rest without specific focus on either of these two categories are labelled as Other (e.g., read or retell experimental materials). Hence, a concept of evidential focus is defined to characterise interviewers' focus on different types of clues.

Measurements of the ICP Characteristics

In order to describe and compare the characteristics of individual ICP in a more concrete manner, quantitative measurements are developed for the four characteristics identified. The measurement of each characteristic is presented in Table 3.13 together with the definition of the characteristic, followed by a detailed explanation of formula construction.

Table 3.13: Characteristics of interviewers' cognitive processes

Characteristic	Definition	Measurement
Cognitive Activeness	The activeness level of an interviewer's cognition	$CA = Modified\ No.\ of\ Coded\ Cognitive\ Actions$
Cognitive Diversity	The diversity level of cognitive actions adopted by an interviewer	$CD = \frac{No.\ of\ Cognitive\ Action\ Categories\ Adopted}{Total\ No.\ of\ Cognitive\ Action\ Categories}$
Frequency of Cognitive Action/IPS	The proportion of ICP adopting a particular cognitive action or IPS (i.e., X)	$F(X) = \frac{MN(X)}{CA}$
Evidential Focus	The proportion of ICP focusing on performance-/response-related information (i.e., Y)	$EF(Y) = \frac{CA(Y)}{CA}$

1) Cognitive activeness

As for cognitive activeness, two possible indicators closely related to the cognitively active level of an interviewer are considered first, namely the word count of a protocol and the number of segments of a protocol. However, these two indicators are waived for different reasons. On one hand, it is not appropriate to measure and compare interviewers' cognitive activeness with the word counts of protocols because the figure can be greatly influenced by an individual's verbosity (Van Someren et al., 1994; Chi, 1997). For example, some interviewers may be less fluent in their expressions and frequently repeat some words, or they may verbalise own thoughts with words like "umm" "you know" and "I would like to say", whereas others express themselves more briefly. Such difference can be caused by various factors, like the differences in the interviewers' expression habits or perception of tension, which affect the word counts of protocols but may not necessarily indicate different levels of the interviewers' cognitive activeness during decision making.

On the other hand, the number of segments of a protocol is considered because it significantly mitigates the impact caused by individual verbosity. However, it is not a

satisfying choice either considering that some interviewers prefer to use longer sentences that are indivisible with more than one cognitive action recognised, whereas others break their thoughts down into short sentences. In this case, while the interviewers preferring longer sentences express own thoughts with fewer segments, it can hardly be claimed that these interviewers have a lower cognitive activeness than the latter group.

Alternatively, the measurement of interviewers' cognitive activeness can be derived from the number of coded actions, which is closely correlated with the overall length of a protocol without being affected by differences in individual verbosity or preference for complex sentences. However, it is noticed that some interviewers do not always verbally report their rating actions while others do, thus a modification is made to the protocols by patching them with separate segments stating "a score of [specific number] was given here" with a code of J1 marked (see Appendix L. for an example). The number of coded actions generated after modification, i.e., modified number of coded cognitive actions of the protocol, is used to measure the cognitive activeness of an interviewer:

$$CA = Modified\ No.\ of\ Coded\ Cognitive\ Actions \tag{3.2}$$

2) Cognitive diversity

Since the concept of cognitive diversity aims to measure the diversified level of the cognitive actions adopted by an interviewer, the number of cognitive action categories occurring in a protocol is counted first and then divided by the total number of categories to calculate the proportion of the total cognitive action categories adopted by an interviewer:

$$CD = \frac{No.\ of\ Cognitive\ Action\ Categories\ Adopted}{Total\ No.\ of\ Cognitive\ Action\ Categories} \tag{3.3}$$

The value of the indicator ranges from 0 to 1, where a higher value means that the

interviewer performs more categories of cognitive actions during decision making and 1 implies that all types of cognitive actions defined are adopted by the interviewer.

3) Frequency of cognitive action and information-processing strategy

The measurement of cognitive diversity concentrates on the occurrence of different cognitive actions, i.e., whether a cognitive action is adopted, while it is not clear at what level of frequency each cognitive action is used. Therefore, a measurement is developed to examine the weights of different cognitive actions adopted by interviewers. To be specific, the modified number of coded cognitive actions is counted for each category first, i.e., CA(X) and then divided by the overall cognitive activeness. Percentages rather than absolute numbers of frequency are used in order to mitigate the influence caused by individual differences in verbosity (Van Someren et al., 1994):

$$F(X) = \frac{CA(X)}{CA} \tag{3.4}$$

where X refers to a particular category of cognitive action or information-processing strategy. Similarly, the proportion of ICP implementing each kind of information-processing strategies is calculated as well to investigate ICP constitution at a higher level, which can be generated either by summing up the percentages of each cognitive action under that strategy or by dividing the modified number of coded actions regarding an information-processing strategy by the overall cognitive activeness. The value of the indicator ranges from 0 to 1, where a higher value implies that the corresponding category of cognitive action or information-processing strategy is executed more frequently by the interviewer, whereas 0 means that the cognitive action or information-processing strategy is not utilised by the interviewer throughout the process.

4) Evidential focus

The evidential focus of an interviewer is measured by the proportions of an inter-

viewer's ICP processing different types of clues (i.e., performance-related information, response-related information, other information), based on which they make judgements or propose following-up actions on an applicant. To calculate the proportion of ICP focusing on a certain type of clues, the modified number of coded actions classified to that clue type is counted and then divided by the overall cognitive activeness. Percentages rather than exact numbers of coded actions are used to avoid the impact caused by differences in the interviewers' cognitive activeness on the measurement and comparison of interviewers' evidential focus.

$$EF(Y) = \frac{MN(Y)}{CA} \tag{3.5}$$

where Y refers to a particular type of clues (i.e., performance-related information, response-related information, other information). The value of the characteristic ranges from 0.00% to 100.00%, and a higher value means that a larger proportion of an interviewer's attention is paid to the type of clues examined and 100.00% implies that the interviewer focuses solely on that type of information throughout the process.

Statistical Analysis of ICP Characteristics

Descriptive analysis is conducted first for each of the four ICP characteristics (i.e., cognitive activeness, cognitive diversity, frequency of information-processing strategy, and evidential focus) as separately observed in the two interviews. Additionally, the observations of each ICP characteristic performed by the same interviewers in the two different interviews are called paired samples. Therefore, in order to investigate whether the interviewers' performance in terms of the ICP characteristics significantly differs in the two interviews, paired-samples t-test or Wilcoxon signed-rank test is conducted on the paired samples according to the normality of dependent variables. Specifically, if a dependent variable is normally distributed under both conditions, paired-samples t-test is adopted to compare the means of the dependent variable in the groups, otherwise the medians of the dependent variable are compared through Wilcoxon signed-rank test.

If no statistically significant difference is detected in the corresponding characteristic between the two interviews, the averaged performance of each interviewer in these two interviews is calculated and a descriptive analysis is conducted on the results.

3.4.2.2 Monitoring IDM at a Cognitive Level

As indicated by the review findings concerning the evidence sources adopted by the process research on IDM, the conclusions of most studies (80.73%) are based on the products produced by interviewers (e.g., scores, decisions) rather than a more direct investigation into how interviewers reach their final decisions. Particularly, since the chief task of interviewers as the decision maker is to make judgements about applicants' qualification based on the information received, a corresponding interest in academic research has arisen to improve interviewers' decisions by structuring the criteria they followed (Campion et al., 1997; Siddique, 2004; Huffcutt, 2011) and the type of evidence they base own decisions on (Taylor & Small, 2002; Day & Carroll, 2003; Klehe & Latham, 2006; Culbertson et al., 2017).

Nevertheless, the studies investigating the impact of structured interview components on IDM rely on the comparison of interviewers' decisions generated in different interview settings through controlled trials, while little is known about their actual performance during interview process. Meanwhile, regardless of the abundant studies on IDM patterns and bias (Parsons & Liden, 1984; Fox & Spector, 2000; Barrick et al., 2012), the majority of the findings are supported by decision analysis as well with little direct evidence derived from the observations of IDM process. Hence, the third step of data analysis aims to promote the monitoring of IDM through the investigation into ICP from two perspectives, namely measuring interviewers' judging behaviour at a cognitive level and detecting typical decision patterns and bias in IDM from ICP.

Cognitive Measurements of Interviewers' Judging Behaviour

1) Constructing cognitive measurements of interviewers' judging behaviour

When developing cognitive measurements of interviewers' judging behaviour, the particular focus is in three aspects, namely to what extent interviewers evaluate applicants according to the given criteria (i.e., criteria-compliance level), to what extent interviewers intend to justify own judgements with evidence (i.e., evidence-based level), and, as a particular requirement of past behaviour questions, to what extent interviewers base own judgements on behavioural clues (i.e., behavioural-based level). The measurement of each dimension of interviewers' judging behaviour is shown in Table 3.14 with the corresponding definition, followed by a detailed explanation.

Criteria-compliance level (CCL). To reduce personal bias and enhance interviewers' decision quality, interviewers are often expected to follow the same given criteria when evaluating applicant (Reilly et al., 2006; Klehe et al., 2008; Melchers et al., 2011; Cambon & Steiner, 2015), while it is not clear whether interviewers are consistent in the way they utilise the given criteria. According to the verbal reports produced by the interviewers, it is noticed that some interviewers may not strictly follow all the dimensions and attributes listed in the rating scale. Thus, a dimension called criteria-

Table 3.14: Dimensions of interviewers' judging behaviour

Dimension	Definition	Measurement
Criteria-Compliance Level	The extent to which an interviewer follows the structured criteria	$CCL = \frac{No.\ of\ Criteria\ Evaluated}{No.\ of\ Criteria}$
Evidence-Based Level	The extent to which an interviewer intends to justify own judgements	$EBL = \frac{No.\ of\ Judgements\ Justified}{No.\ of\ Criteria\ Evaluated}$
Behavioural-Based Level	The extent to which an interviewer makes judgements based on applicants' behavioural clues	$BBIU = \frac{No.\ of\ Judgements\ Based\ on\ Behavioural\ Clues}{No.\ of\ Criteria}$

Note. No. of Criteria is 23 in this research.

compliance level (CCL) is defined in this research, which indicates the extent to which an interviewer follows the given criteria when evaluating applicants. Specifically, CCL is measured by the proportion of the given criteria evaluated by an interviewer:

$$CCL = \frac{No.\ of\ Criteria\ Evaluated}{No.\ of\ Criteria} \tag{3.6}$$

The value of CCL ranges from 0 to 1, where a higher value means the interviewer takes more given criteria into consideration when evaluation an applicant and 1 means that the interviewer follows all the given criteria.

Evidence-based level (EBL). In addition, while interviewers are expected to derive their judgments and decisions from target information that are closely related to the requirements of the job position to avoid unfairness and risk of lawsuits (Campion et al., 1997; Hackett et al., 2004; Madera & Hebl, 2012), it is found that they may not always provide supporting evidence to justify own judgements. Hence, the second dimension called evidence-based level (EBL) is developed to describe the extent to which an interviewer intends to justify own judgements with evidence. A special situation is recognised where interviewers make judgements based on their perceptions that they cannot identify sufficient evidence from the applicant's response. In this case, the interviewers are regarded as "intending to justify own judgments" and the corresponding judgments are marked as "justified". The formula constructed to measure EBL is as shown below:

$$EBL = \frac{No.\ of\ Judgements\ Justified}{No.\ of\ Criteria\ Evaluated} \tag{3.7}$$

The value of the dimension ranges from 0 to 1, where a higher value indicates that more judgments made by an interviewer on the given criteria are supported by specific evidence and 1 means that the interviewer provides evidence for all the judgments made on the given criteria.

Behavioural-based level (BBL). A number of studies suggest that past behaviour question, which is adopted in the experiment within this research, is based on the premise that the best predictor of future behaviour is past behaviour and can improve the quality of interviewers' decisions (Moscoso & Salgado, 2001; Harel et al., 2003; Bangerter et al., 2014). Despite the possible advantages of this question type, it is not clear whether interviewers follow the corresponding requirements and evaluate applicants based on the behavioural clues extracted from their past performance. Therefore, the third dimension of interviewers' judging behaviour concerns the extent to which an interviewer makes judgements based on applicants' behavioural clues, which is defined as behavioural-based level (BBL) and measured by the formula below:

$$BBL = \frac{No.\ of\ Judgements\ Based\ on\ Behavioural\ Clues}{No.\ of\ Criteria} \tag{3.8}$$

The value of BBL ranges from 0 to 1, where a higher value implies that an interviewer makes judgments about an applicant based on behavioural clues for a larger proportion of the given criteria and 1 indicates that the interviewer evaluates an applicant with all the given criteria and provides justifications for each of the criteria according to behavioural clues.

2) Detecting target judging behaviour in the protocols

To obtain the corresponding data of the indicators within these formulas, the protocols are again scrutinised by the two co-coders separately to identify, code and count interviewers' target behaviours following the logic shown in Figure 3.5. For each of the protocols, the co-coders first check through the protocol and mark the criteria used by the interviewer to evaluate the applicant. Particularly, it is noticed that the interviewers may sometimes refer to their own criteria (e.g., personal preferences or experiences), which are idiosyncratic in nature and unable to validate. Thus, the discussion on the interviewers' judging behaviour is constrained to the criteria listed in the rating scale (i.e., four dimensions and nineteen attributes, see Appendix F. and G.).

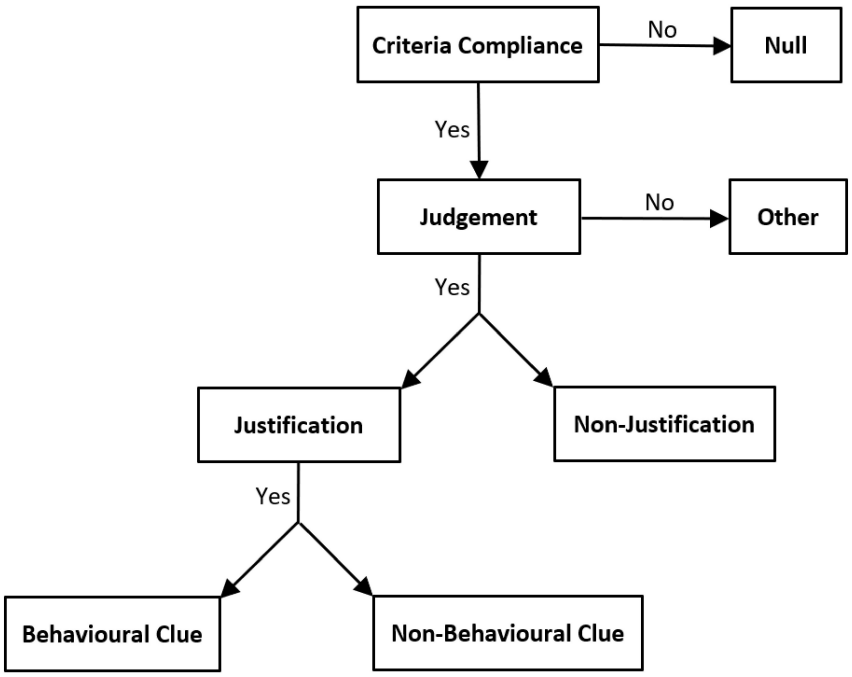


Figure 3.5: Logic of coding for interviewers' judging behaviour

Secondly, the co-coders screen out the criteria that are only mentioned without a judgment made by the interviewer and then count the total number of criteria evaluated. Thirdly, by checking whether the interviewer provides or at least intends to provide supporting evidence to justify own judgements, each of the evaluated criteria is marked as either with or without justification and the number of judgements justified is counted. At the final step, the nature of evidence used by the interviewers to support own judgements on each criteria evaluated is checked and categorised as either based on behavioural clues or non-behavioural clues, and the number of judgments based on behavioural clues is counted. The two co-coders complete coding independently and achieve an inter-coder agreement level of 94.2%, indicating that the outcomes are appropriate for further analysis after solving the discrepancies.

3) Statistical analysis of interviewers' judging behaviour

Similar to the statistical analysis of ICP characteristics, the descriptive statistics of

the three dimensions of interviewers' judging behaviour (i.e., criteria-compliance level, evidence-based level, and behavioural-based level) are reported first, followed by a significance test of difference between the interviewers' judging behaviour performed in the two different interviews (i.e., paired samples). Either paired-samples t-test or Wilcoxon signed-rank test is adopted depending on the results of normality test of each dependent variable. If no statistically significant difference is identified in a dimension between the two interviews, a descriptive analysis is conducted on the averaged performance of each interviewer on that particular dimension.

Cognitive Evidence of Interviewers' Decision Patterns and Bias

In addition to the cognitive measurements developed for IDM, the protocols are also searched for direct evidence from the protocols to verify the existence of interviewers' typical decision-making patterns and bias (e.g., cue weighting, halo effect, contrast effect, similar-to-me effect). However, it is foreseeable that a few decision patterns and bias, although heatedly discussed in interview literature, are in nature out of discussion in this research due to the experiment design. For example, interviewers' judging behaviour related to factors like applicants' demographics, appearance and non-verbal cues that are not revealed in the experiment is less likely to be recognised in ICP. Similarly, interviewers' questioning style, like question types and sequence, is also limited since the interview communication is determined in advance. After screening out the aspects that are less likely to be explored within the protocols collected, evidence is sought for within the verbal protocols to check whether four types of decision patterns (i.e., stereotypes of ideal applicants, anchoring-and-adjusting, deception detection, and cue weighting) and three types of decision bias (i.e., similar-to-me effect, halo effect, and contrast effect) exist according to their definitions (see Table 2.3). The results are reported in the form of quotes from the interviewers' verbal protocols.

3.4.2.3 Examining the Associations between ICP and Contextual Factors

Considering the wide interest of employment interview research in the impact of contextual factors on interviewers' performance during interview process, the potential associations between ICP during decision making and several contextual factors are examined.

Dependent Variables

The dependent variables include the four ICP characteristics (i.e., cognitive activeness, cognitive diversity, frequency of information-processing strategy, and evidential focus) and the three measurements of interviewers' judging behaviour (i.e., criteria-compliance level, evidence-based level, and behavioural-based level) defined in this research. Recalling the significance test of difference for both ICP characteristics and interviewers' judging behaviour in the two interviews, the averaged results of interviewers' performance throughout the two interviews are applied to the association analysis as the data of the dependent variables if no significant difference is recognised, otherwise the data of the interviewers' performance in the first interview will be utilised.

Independent Variables

Four contextual factors are taken into investigation as independent variables, of which three factors are chosen because existent studies have approved their impact on interviewers' performance and decision quality, including interviewers' gender (Parsons & Liden, 1984; Raza & Carpenter, 1987; Chapman & Rowe, 2001), experience level as interviewer (Hess, 2013; Roulin et al., 2015; Frieder et al., 2016), and type of rating scale (Fay & Latham, 1982; Maurer, 2002; Klehe et al., 2008; Melchers et al., 2011). The national region of previous interview experience as the fourth factor, although having not been introduced into employment interview research yet, is also taken into the analysis considering that differences in regional culture (Sanchez-Burks et al., 2006; Manroop et al., 2013) and employment laws (Williamson et al., 1997; Morgeson et al., 2008) may affect IDM.

Among the four independent variables of interest, three factors are binary variables in nature, including interviewers' gender (male vs. female), national region of interview experience (UK vs. CN), and rating scale type used in the interview task (BARS vs. OGRS). As for the experience of working as an interviewer, although quantified information is collected (i.e., the number of years working as an interviewer, the frequency of conducting employment interviews) from the interviewers, the results vary greatly among individuals from zero to over 30 years or interviewing hundreds of applicants per year. Considering the small sample size in this research as well as the fact that a clear definition of "experienced interviewer" is absent, experience cannot be analysed as a continuous variable or ordinal categorical variable, thus it is transformed into a binary variable by labelling interviewers with at least one year of working experience as an interviewer with "experienced", otherwise "novice".

Statistical Analysis

To examine if each of the four contextual factors leads to statistically significant difference in the ICP characteristics and the three dimensions of interviewers' judging behaviour, significance test of difference between independent samples is conducted on the observations of the 29 interviewers. According to the normality of each dependent variable in the two groups of a contextual factors, either independent-samples t-test or Mann-Whitney U test is carried out to investigate if the dependent variable observed in the paired groups of a binary variable are significantly different with each other. Specifically, independent-samples t-test is adopted to compare the means of a dependent variable if the two groups of samples are normally distributed, otherwise the medians of a dependent variable observed in the two groups are compared through Mann-Whitney U test.

3.5 Justification of Research Trustworthiness

This research is carefully designed and conducted in terms of its choice of methodology, development of experiment procedure, data collection and data analysis to ensure that the widely accepted standards (i.e., credibility, transferability, dependability and confirmability) in pursuit of the trustworthiness of qualitative research are met (Guba, 1981; Shenton, 2004; Williams & Morrow, 2009; Connelly, 2016).

1) Choice of methodology and development of experiment procedure

The methodology utilised in this research aiming to exploring interviewers' cognitive processes during decision making is called think aloud method, which is regarded as the most direct method applied in the investigation into human cognitive process and is well established on solid psychological base. It is commonly applied across various disciplines while has not yet been used to study IDM. Therefore, the mechanism of TAM is explained first, followed by a comparison between TAM and other research methods of human cognition which demonstrates the superiority of TAM (see Section 3.1). Specifically, the data of human cognitive process collected through TAM is less vulnerable to external disturbance as well as the subject's memory errors and self-interpretation, all of which distort the data provided by the subject from the real performance on the task investigated. Correspondingly, the honesty of informants and the quality of data collected should be ensured by the nature of TAM. On one hand, the concurrent reports represent interviewers' real-time thinking processes and the possibility of memory distortion is restrained to the least. On the other hand, the interaction between interviewers and the experimenter is minimised during the mock interview task to avoid the potential impact caused by the experimenter on interviewers' task performance.

Apart from the advantages of TAM, the potential concerns of applying the method in employment interview scenario is discussed in detail based on the common require-

ments of TAM on both the nature of the task investigated and the essential quality of participants (see Section 3.2). Based on the discussion, the experiment procedure of an interview task with TAM embedded is carefully designed, tested and revised. Specifically, the experimenter learns from massive literature about how to create a mock interview scenario and how to effectively implant TAM to the task investigated. The stimulus materials are developed and approved with the assistance of expert interviewers. The overall experiment procedure is tested through a pilot study, which confirms the feasibility and validity of the experiment to collect data regarding interviewers' cognitive processes during decision making. In addition, feedback on the experiment is collected from the interviewers involved in data collection as well to re-confirm the validity of the experiment design for data collection.

2) Data collection

In addition to the development of the experiment, the trustworthiness of data collection (see Section 3.3) is monitored carefully as well. First, the interviewers are selected through random sampling in the sense that the invitation is sent through multiple channels and all the respondents are included as long as they are capable of completing the interview task while thinking aloud. Second, rather than to generate common rules among large samples, TAM-based studies aim at an in-depth exploration of individual cognitive process with a small sample size (Kuipers & Kassirer, 1984; Fonteyn et al., 1993) and, as put by Lewis (1982), "the thinking-aloud approach can be used with a very small number of participants and still give valuable results ... data from even one person may give useful feedback to the designer". The number of interviewers involved in this research is 29 including 11 interviewers possessing rich working experience as interviewer, which should be sufficient to provide data for the research questions about IDM. In addition, except for the national region of the interviewers' previous interview experience, which is controlled by collecting data in both UK and China, no restriction is set on any other background feature of interviewers and a broaden variety of interviewers in terms of industries, types of organisations and job positions take part in the

data collection. This is believed to ensure the richness of information as well as to help achieve triangulation via a wide range of data sources.

Meanwhile, since interviewers with different regional backgrounds (i.e., UK or China) take part in the research, the experiment materials in both English and Chinese are available so that interviewers can choose the language version that they can effectively understand and express themselves. Particularly, the experiment materials in the two languages are checked and approved by an expert bilingual in English and Chinese to ensure that not only the meaning but also the structure of the content are maintained the same to minimise the influence of language difference on the data collection.

3) Data analysis

The data analysis of this research basically follows the standard procedure of protocol analysis (see Section 3.4), through which the research trustworthiness is strengthened in several aspects. First, when transcribing the recorded protocol into written form, the accuracy of transcription is confirmed with the corresponding interviewer who produces the protocol. In addition, the part of the protocols collected in Chinese are translated first by the experimenter and then proofread by the bilingual expert who has previously assisted in the translation of the experimental materials to ensure the structure and meaning of the content are not altered. When it came to the coding procedure and refinement of the coding scheme, two volunteer co-coders who are PhD students from different departments without particular interest in or knowledge of this research are involved to minimise the possibility of subjective interpretation. The common threshold of inter-coder agreement is 90.0%, while an inter-coder agreement of 91.3% is achieved in this research and all the disagreements are solved.

In addition, the coded protocols are analysed with the guidance from supervisor. Particularly, we arrange regular meetings to discuss over the possible interpretations of the data to avoid subjectivity. Besides, a detailed ongoing journal is developed as a self-reflective record of how the whole research process evolves, which is regularly

reviewed both to monitor the potential risk of subjectivity and to seek for the possibility of improvement. Recalling Section 3.2 as an example where the development of experiment procedure is reported and explained step by step, such practice can ensure the transparency of the research and enable the whole procedure to be transferred or repeated in a different context. At the same time, the research is shared and discussed in several international conferences with worldwide scholars who have provided valuable feedback facilitating the refinement of the research.

3.6 Summary

This chapter demonstrates the overall design of the present research, where a psychological method called TAM is adopted and the verbal protocols are processed and analysed to seek answers to the research questions raised.

Noticing that TAM has not yet been applied to employment interview research and no well-developed paradigm with TAM embedded in interview scenario can be referred to, the choice of method is justified first by starting with the history, mechanism and advantages of TAM. Afterwards, the feasibility of applying TAM to exploring ICP is examined. On the basis of a detailed discussion over the potential concerns and corresponding countermeasures, an experiment procedure is developed where interviewers are required to execute think aloud practice during decision making. Then the experiment design is tested and revised through a pilot study, which basically confirms the feasibility of the experiment.

The experiment established is then carried out for data collection. Twenty-nine interviewers with diversified backgrounds take part in the data collection acting as interviewers and provide sufficient verbal data that represents their thinking processes when working on the interview task. Their feedback on the experiment further supports the reliability and validity of the research design and the data collected.

The verbal protocols obtained as the data of ICP are processed following a stan-

dard procedure and analysed to seek for answers to the research questions. The coding scheme developed during protocol processing presents the elements identified in ICP (RQ1), and the coded protocols generated through protocol processing is analysed to define ICP characteristics (RQ2), construct dimensions of interviewers' judging behaviour (RQ3a-c), and detect cognitive evidence of typical decision patterns and bias in IDM (RQ3d). In addition, the associations between the four contextual factors and ICP are examined though significance test of difference (RQ4). The trustworthiness of the research is justified with details of the research design as well.

Chapter 4

Research Findings

In this chapter, the research findings are reported in accordance with the research questions proposed. Specifically, the findings regarding the elements (RQ1a&b) and characteristics of ICP (RQ2a-d) are presented in **Section 4.1** and that of interviewers' judging behaviour (RQ3a-d) are shown in **Section 4.2**. Then in **Section 4.3**, the results of association test between the contextual factors and ICP (RQ4) are discussed. Finally in **Section 4.4**, a summary of the present chapter is provided.

4.1 Identifying the Elements and Characteristics of ICP

4.1.1 Types, Prevalence and Sequence of ICP Elements

1) Types of the elements in ICP

The major elements of ICP during decision making are identified and defined at three different levels during protocol processing, including elementary elements (i.e., information type, operator), combined elements (i.e., cognitive actions in the form of "information type + operator"), and information-processing strategies. At the same time, several types of irrelevant comments are also summarised to inclusively conclude

the possibilities of ICP. A coding scheme is developed to present the corresponding results (see Table 4.1).

Table 4.1: The finalised coding scheme

Code	**Example**
Interpreting Strategy: Interviewers interpret and extract information from experimental materials or applicants' responses	
I1: Read or represent experimental materials/settings (i.e., job description, interview questions, rating scale)	*Okay, so the question was how he handled multiple tasks or projects when the time was tight.*
I2: Represent the content of applicants' response; Or present an existing fact happened during the interview	*... he talked about his own scheduling methods and that if people are looking for some flexibility and coordinate.* *He had to be prompted and challenged for a specific example again.*
I3: Identify a particular characteristic of the response	*However, it sounded more like his answer was based on what he will do in the future.* (behavioural-based level) *That wasn't really about handling stress.* (relevance) *...but he didn't specifically talk about a time where there was a specific time pressure that he had to achieve.* (comprehensiveness) *...he was really specific when he talked about his experience.* (detailed level) *He used the word "we" a lot...*(wording) *He definitely sounds deflated thinking about the case.* (tone) *...because to use in an interview, it didn't have a successful outcome.* (result)
I4: Make an inference (NOT evaluation) as a supplement to the response (e.g. situation, action, result) Or define (NOT evaluate) the behaviour of the applicant	I just feel that they were maybe not so capable that he might need to train his team a bit more. (inf.-situation) It suggested that he'd been making the decision on whether or not people got the opportunity to provide the feedback. (inf.-action) … he is very quick to go over people's heads to managers to force other people to do work may then cause tensions between himself and other employees. (inf.-result) I've written down that he invests in his team. (define)
I5: Utter follow-up questions; Or express an intention to do so	*So what I would have liked him to say is, for example, I was handling a project and this is what happened and this was complicated, etc.*

(continued table)

Code	Sample
Judging Strategy: Interviewers make judgements or justify the judgements they have made	
J1: Articulate a (positive/negative) evaluation or rating score; Or make a suggestion for improvement	*So the stressful example one has been quite weak* (evaluation – answer) *And I feel like he has been maybe a little bit arrogant and close-minded to other people's opinions.* (evaluation – applicant) *I think I would give him a 3, but he does need improvement.* (rating)
J2: Make or revise a hiring decision; Or express an intention to do so	*I'm not sure that I would want to employ this person.*
J3: Comment on the informative level of applicant's responses regarding a particular criterion	*There wasn't anything about improvement of his stress management. (Um, so I just give that a 3.)*
J4: Associate the evaluation of one criterion to that of another (especially due to a lack of direct evidence)	*(He is open to different ideas,) but that does not necessarily mean he is good at handling conflicts or promoting teamwork.* *(I think he can't be bad at handling conflicts,) because he has openly and patiently considered about others' opinions.*
J5: Associate with own knowledge/ experience/preference/beliefs as criteria, etc.	*As an HR you would want someone in a project management role to be able to manage the individuals and the team…*
J6: Compare or associate with the applicant's responses to other questions	*Yeah, I think this answer was less successful than the first one.*
J7: Compare or associate with previous applicants	*What the first candidate does for me, which the second didn't, is that the first demonstrated valuing his workers.*
J8: Give a further explanation or example for own judgements	*(He does seem to be able to find solutions.) I mean, I didn't think any of them were highly creative, but they were decent.*
Self-Monitoring Strategy: Interviewers reflect on own decision-making process	
S1: Reflect on own personal cognitive status/habits/preferences when performing the task, etc.	*To be honest, I'm also thinking about his early examples as well…* *Maybe I'm just a little bit biased based on some of his other responses…*
S2: Reflect on own previous actions/comments/ratings	*I did mention earlier on that I'd be interested in hearing example from applicant one.*
S3: Suggest follow-up actions (except for asking promoting questions)	*…and I'm just going to have that as a benchmark.* *Now, I think at this stage, it's probably a good time to compare him to the previous candidate.*
S4: Point out an own knowledge gap or doubt; Or confirm the information available	*I don't know if there was sort of an evaluation tool.*

(continued table)

Code	Sample
Irrelevant Comments	
Incomplete expressions	*I'm not ... I mean...*
Vague statements for decision making	*So that's interesting.* *He mentioned that he does do housework, so maybe he can come over and clean my house when he is stressed.*
Comments on the experiment design	*...there's quite a distinction between the various levels between 1, 3, or 5 in the attributes.*

Information type. According to the nature of TAM, the verbal protocols produced by the interviewers reflect how they mentally process information within STM to make final decisions. It is found that the interviewers pay attention to four major types of information when working on the interview task. The first information type is closely related to the design of interview task, of which examples include task requirements, job description, interview questions, and rating criteria. The second type of information processed by the interviewers is derived from applicants' responses, which can be further classified as either an applicant's work performance stated in a response (e.g., "*... he went on to talk about not wanting to ask colleagues for help or advice, because it might be, um, an imposition on their work schedule, or it may have an impact on their performance appraisal*") or the features of responses as perceived by the interviewer (e.g., "*... he was really specific when he talked about his experience*"). Thirdly, the interviewers also process the information obtained before the interview, such as personal experience, knowledge, and beliefs (e.g., "*... as a project manager, you absolutely have to give people the flexibility and the opportunity to work independently*", "*... interviewees should be prepared with examples that showcase what they've done when going for an interview*"). In addition, some information produced by the interviewers earlier during decision making may be further processed. For example, interviewers may take the inferences and evaluations made about the applicant earlier in the interview into consideration when making their final decisions (e.g., "*I did mention earlier on that I'd be interested in hearing example from applicant one*").

These four types of information processed in STM basically comes from two information sources, including external environment and the interviewers' long-term memory. On one hand, the interviewers receive rich information from the external environment throughout the interview process, particularly the task requirements (e.g., job description, evaluation criteria) and the applicant's responses to the interview questions. While applicants' non-verbal cues like behaviours, gestures and facial expressions are often found affecting IDM (DeGroot & Gooty, 2009; Tsai et al., 2012; Nguyen & Gatica-Perez, 2015), they are out of investigation in this research due to the design of data collection. Notably, not all the information received from external environment through the interviewers' sensory receptors (e.g., eyes, ears) are processed in STMs. Such situation may form in two major mays as a result of the interviewers' limited cognitive load, including passive information loss (Van Merrienboer & Sweller, 2005; Sepp et al., 2019) and proactive information filtering (Savolainen, 2007; Saxena & Lamest, 2018; Jones & Kelly, 2018), for example, according to the perceived relevance of the information to the decision-making task (Bawden & Robinson, 2020). On the other hand, a variety of information processed in the interviewers' STM is retrieved from their LTM, such as the information about own beliefs, preferences and past experience, the information received and memorised earlier in the current interview (e.g., previous responses provided by the applicant, previous evaluation and scores given to the applicant), as well as the information related to the previous applicants.

Operator. The other group of elementary elements of ICP is the operators adopted by the interviewers to process each piece of information, of which typical examples include reading, representing, recalling, interpreting, evaluating, comparing and prompting. Nevertheless, the knowledge of ICP generated from each operator alone is quite limited, especially when considering the fact that the same operator can be executed to different information and the same information can be processed by different operators as well. Therefore, instead of taking a further look into the operators themselves, the different combinations of information type and operator are investigated to explore ICP.

Cognitive action. A combination of the information processed and the operator used to process the information is defined as a cognitive action. As can be seen in the coding scheme, seventeen cognitive actions are recognised in ICP that cover all the possibilities of the interviewers' mental activities during decision making. For example, the interviewers may deal with task-related information through "reading or representing experimental materials or settings" (I1), extract information from the applicant's response by "representing the content of the applicant's response" (I2) or "identifying a particular characteristic of a response" (I3), evaluate the applicant by "articulating a (positive/negative) evaluation or rating score, or make a suggestion for improvement" (J1), and finally "make or revise a hiring decision" (J1). At the same time, the interviewers may also occasionally monitor own decision-making process, for example, by "reflecting on own previous actions, comments or ratings" (S1) or "pointing out an own knowledge gap or doubt" (S4).

Information-processing strategy. The cognitive actions identified are further classified into different categories according to their immediate purpose from a perspective of information processing, which is defined as information-processing strategy in this research. In particular, the interviewers adopt the cognitive actions following three different information-processing strategies, namely interpreting strategy, judging strategy, and self-monitoring strategy. Interpreting strategy covers the five cognitive actions (I1-I5) identified where the interviewers interpret and extract information from experimental materials or the applicant's responses. Judging strategy is adopted by the interviewers when making judgements or justifying own judgements, of which eight cognitive actions (J1-J8) are found. As for self-monitoring strategy with four corresponding cognitive actions (S1-S4) recognised, the concentration of ICP is on the monitoring of own decision making and the interviewers describe, explain, or reflect on own cognitive processes.

Irrelevant comments. Apart from the elements reported above, a special category called "irrelevant comments" is included in the coding scheme as well to present the interviewers' comments that are incomplete expressions, vague in meaning or rele-

vance to the decision-making task, or less likely to happen in a real interview situation. For example, the interviewers are found occasionally producing incomplete expressions (e.g., "*I'm not ... I mean...*") or vague comments on the applicant (e.g., "*He mentioned that he does do housework, so maybe he can come over and clean my house when he is stressed*"), of which the relevance to the decision-making task can hardly be recognised. At the same time, the interviewers may also comment on the experiment design (e.g., experiment procedure, experimental materials), which is less likely to happen in a real interview situation and is thus regarded as less relevant to the interview task itself.

2) Prevalence and sequence of the ICP elements

The prevalence of a cognitive action or information-processing strategy is measured by the percentage of the interviewers adopting each cognitive action and information-processing strategy, of which the results are presented in Table 4.2 together with the prevalence rank of each cognitive action (i.e., 1 is assigned to the most prevalent cognitive action). As can be interpret from the outcomes, all the 29 interviewers adopt the three categories of information-processing strategy ($P = 1.00, n = 29$), except that self-monitoring strategy ($P = 0.97, n = 28$) is not recognised in the cognitive process of one interviewer (i.e., Interviewer 15). As for the prevalence of each cognitive action, I2, I3, I4, J1, J2 and J5 are the most prevalent cognitive actions and are executed by all the 29 interviewers ($P = 1.00, n = 29$), followed by I1 ($P = 0.93, n = 27$), J8 ($P = 0.90, n = 26$), I5 ($P = 0.86, n = 25$), S1 ($P = 0.83, n = 24$), J3 ($P = 0.79, n = 23$) and S4 ($P = 0.79, n = 23$), which are also widely used by most interviewers. In contrast, cognitive actions like J6 ($P = 0.41, n = 12$), S2 ($P = 0.38, n = 11$) and J4 ($P = 0.24, n = 7$) are less prevalent among the interviewer.

The sequence of the cognitive actions and information-processing strategies in the verbal protocols are examined and compared to figure out if a universal model can be constructed to illustrate the process of IDM at a cognitive level. In general, it is found that the interviewers integrate cognitive actions into complex and interactive

Table 4.2: Prevalence of cognitive actions and information-processing strategies

Cognitive Action/Strategy	**Prevalence (*n*)**	**Rank**
Interpreting	***1.00 (29)***	
I1	0.93 (27)	7
I2	1.00 (29)	1
I3	1.00 (29)	1
I4	1.00 (29)	1
I5	0.86 (25)	9
Judging	***1.00 (29)***	
J1	1.00 (29)	1
J2	1.00 (29)	1
J3	0.79 (23)	11
J4	0.24 (7)	17
J5	1.00 (29)	1
J6	0.41 (12)	15
J7	0.59 (17)	13
J8	0.90 (26)	8
Self-Monitoring	***0.97 (28)***	
S1	0.83 (24)	10
S2	0.38 (11)	16
S3	0.55 (16)	14
S4	0.79 (23)	11

combinations which are applied to various pieces of information received until a final decision is made. It is also noticed that cognitive actions adopted by the interviewers when commenting on a piece of information (or several pieces of information where they are associated or compared with each other) do not occur in an isolated manner but are logically related to each other. This is especially apparent where conjunctions like "because" "so" "but" "however" and "also" are used in the expressions. However, no fixed sequence can be identified for either the occurrence of cognitive actions or the information-processing strategies. This is not surprising considering that the interviewers are less likely to be highly structured when expressing themselves, especially when they are instructed to "speak out whatever comes to their minds" so that the verbal protocols can reflect their real thinking processes when working on the interview task. At the same time, the analysis of cognitive diversity shows that the interviewers differ in not only the number but also the type of cognitive actions executed. Consequently, it seems less feasible to develop a procedural model with the cognitive actions or information-processing strategies consolidated in a detailed and fixed manner to elaborate ICP during decision making.

4.1.2 Characteristics of ICP

This part reports the findings of the four ICP characteristics defined from two perspectives, including the descriptive statistics in each of the two interviews (see Table 4.3) and the comparison between each pair of characteristics observed in the two interviews based on the significance test of difference (see Table 4.4).

1) Cognitive activeness

According to the results of descriptive analysis, the cognitive activeness of the interviewers during the first interview ($M = 107, SD = 50$) is close to that in the second interview ($M = 102, SD = 57$). A paired-samples t-test is conducted on the 29 interviewers to determine if a difference exists in the mean cognitive activeness

Table 4.3: Descriptive statistics of ICP characteristics in the two interviews

	Interview	*M*	*SD*	*Min*	*Max*
Cognitive Activeness	1	107	50	29	210
	2	102	57	25	216
Cognitive Diversity	1	0.66	0.13	0.29	0.94
	2	0.67	0.16	0.29	0.88
Frequency of IPS					
Interpreting	1	46.90%	9.71%	10.34%	60.48%
	2	46.84%	9.37%	29.03%	63.49%
Judging	1	46.81%	10.90%	34.44%	89.66%
	2	47.84%	9.72%	34.92%	70.97%
Self-Monitoring	1	4.78%	3.52%	0.00%	14.46%
	2	4.02%	3.26%	0.00%	10.53%
Irrelevant	1	1.51%	2.10%	0.00%	10.00%
	2	1.30%	1.86%	0.00%	7.24%
Evidential Focus					
Performance	1	68.63%	11.36%	37.78%	93.10%
	2	71.95%	11.84%	42.55%	91.89%
Response	1	19.70%	12.50%	1.72%	50.00%
	2	18.09%	12.60%	1.45%	55.32%
Other	1	11.67%	7.42%	0.00%	26.92%
	2	9.97%	6.90%	0.00%	23.19%

Note. $N = 29$

between the two interviews, while no statistically significant difference is recognised $[t(28) = 0.669, p = .509]$. Hence, descriptive analysis is carried out on the interviewers' averaged cognitive activeness of the two interviews. The shows that the mean cognitive activeness of the interviewers is 108 ($SD = 50$), and the cognitive activeness of individual interviewers varies greatly in the sense that the lowest cognitive activeness is 30 while the highest reaches 206.

Table 4.4: Comparing ICP characteristics in the two interviews

	Test	***p***	***M***	***SD***	***Min***	***Max***
Cognitive Activeness	PST	.509	108	50	30	206
Cognitive Diversity	WSR	.529	0.66	0.12	0.29	0.88
Frequency of IPS						
Interpreting	WSR	.871	46.87%	8.25%	19.69%	59.45%
Judging	WSR	.538	47.32%	8.98%	36.16%	80.31%
Self-Monitoring	WSR	.412	4.40%	2.64%	0.00%	10.04%
Irrelevant	WSR	.528	1.41%	1.69%	0.00%	5.79%
Evidential Focus						
Performance	PST	.081	70.29%	10.50%	50.48%	91.71%
Response	WSR	.214	18.89%	11.54%	2.39%	43.84%
Other	PST	.064	10.82%	6.76%	0.00%	25.06%

Note. 1. *PST* refers to paired samples t-test, and *WSR* refers to Wilcoxon signed-rank test; 2. $N = 29$.

2) Cognitive diversity

The descriptive statistics of the interviewers' cognitive diversity in the two interviews are quite similar. Specifically, the mean cognitive diversity is 0.66 in the first interview ($SD = 0.13$) and 0.67 in the second interview ($SD = 0.16$). Besides, the maximums of cognitive diversity in both interviews are less than 1.00, indicating that no interviewer adopts all the seventeen categories of cognitive actions in a single interview. The results of a Wilcoxon signed-rank test imply that no statistically significant difference exists in the medians of the interviewers' cognitive diversity between the two interviews ($Mdn_1 = 0.65, Mdn_2 = 0.71, Z = -0.629, p = .529$). Overall, the mean of the interviewers' averaged cognitive diversity in the two interviews is 0.66 ($SD = 0.12$), while difference can be perceived among individual interviewers ($Min = 0.29, Max = 0.88$).

However, it should be noted that the discussion over cognitive diversity concentrates on the number of cognitive actions category adopted by the interviewers, regardless of what categories of cognitive actions are adopted. In other words, different categories of cognitive actions may be executed by interviewers with the same level of cognitive diversity. For example, while the cognitive diversity of both Interviewer 5 and Interviewer 7 is 0.82 in the second interview, their adoption of four types of cognitive actions differs. Specifically, J2 and J4 are performed by Interviewer 8 but not by Interviewer 5, whereas J7 and S2 are identified in the cognitive process of Interviewer 5 but not Interviewer 8.

3) Frequency of cognitive actions and information-processing strategy

The frequency of each cognitive action and strategy is calculated separately for each interviewer during the two interviews. However, noticing that each category of cognitive action accounts for a relatively small proportion of ICP and the counts of many cognitive actions are zero, investigation is conducted only on the frequencies of the three information-processing strategies and irrelevant comments rather than the cognitive actions. The descriptive statistics are reported in Table 4.3, from which it can be seen that mean frequency of each category observed in the two interviews is close to each other, which has been confirmed by Wilcoxon signed-rank tests as well that no significant difference exists in the median frequency of interpreting strategy ($Z = -0.162, p = .871$), judging strategy ($Z = -0.616, p = .538$), self-monitoring strategy ($Z = -0.820, p = .412$) or irrelevant comments ($Z = -0.631, p = .528$) in the two interviews. Therefore, the averaged performance of the interviewers in these two interviews regarding the frequency of each information-processing strategy and irrelevant comments is figured out. According to the descriptive statistics listed in Table 4.4, the mean frequency of interpreting strategy ($M = 46.87\%, SD = 8.25\%$) is similar to that of judging strategy ($M = 47.32\%, SD = 8.98\%$), whereas the mean frequency of both self-monitoring strategy ($M = 4.40\%, SD = 2.64\%$) and irrelevant comments ($M = 1.41\%, SD = 1.69\%$) are much lower.

4) Evidential focus

Descriptive analysis is conducted separately on the three types of information that the interviewers focus on during decision making (i.e., performance-related information, response-related information, and other information), of which the corresponding results are presented in Table 4.3. In general, the interviewers pay much more attention to performance-related information in both interviews ($M_1 = 68.63\%, SD_1 = 11.36\%; M_2 = 71.95\%, SD_2 = 11.84\%$) than to response-related ($M_1 = 19.70\%, SD_1 = 12.50\%; M_2 = 18.09\%, SD_2 = 12.60\%$) or other information ($M_1 = 11.67\%, SD_1 = 7.42\%; M_2 = 9.97\%, SD_2 = 6.90\%$). To find out if the interviewers' evidential focus in the two interviews differs significantly, significant test of difference is conducted on the paired samples for each type of information. The comparison of both the interviewers' focus on performance-related $[t(28) = -1.812, p = .081]$ and other information $[t(28) = 1.927, p = .064]$ is on the basis of paired-sample t tests, which shows that the means of the paired samples are not significantly different from each other. As for the interviewers' focus on response-related information, a Wilcoxon signed-rank test is carried out to compare the medians of the paired samples, while no statistically significant difference can be recognised either ($Z = -1.243, p = .214$). Consequently, the averaged levels of the interviewers' focus on these three types of information within the two interviews are calculated and summarised in Table 4.4. Overall, the majority of the interviewers' attention is paid to performance-related information ($M = 70.29\%, SD = 10.50\%$), whereas they focus much less on response-related information ($M = 18.89\%, SD = 11.54\%$) and other information ($M = 10.82\%, SD = 6.76\%$).

4.2 Monitoring Interviewers' Judging Behaviour at a Cognitive Level

In addition to exploring the black box of ICP by demonstrating the elements occurring, the characteristics perceived as well as how IDM unfolds at a cognitive level,

ICP is also investigated to promote the monitoring of interviewers' judging behaviour. In particular, three dimensions (i.e., criteria-compliance level, evidence-based level, and behavioural-based level) are constructed to measure interviewers' corresponding judging behaviour of interest and the verbal protocols are scrutinised to seek for direct evidence of typical decision patterns, of which the findings are presented in this section.

4.2.1 Cognitive Measurements of Interviewers' Judging Behaviour

The findings of the investigation into the three dimensions of interviewers' judging behaviour are reported in this part. For each of the dimensions, the descriptive statistics of the observations in each of the two interviews are presented first (see Table 4.5), and then the results of significance test of difference in these dimensions between the paired samples regarding the two interviews are discussed (see Table 4.6).

1) Criteria-compliance level

As indicated by the descriptive statistics, the overall performance of the interviewers' criteria-compliance level during the two interviews turns out to be quite close to each other with similar means and standard deviations ($M_1 = 0.72, SD_1 = 0.35; M_2 =$

Table 4.5: Descriptive statistics of interviewers' judging behaviour in the two interviews

	Interview	***M***	***SD***	***Min***	***Max***
Criteria-compliance level	1	0.72	0.35	0.17	1.00
	2	0.70	0.37	0.17	1.00
Evidence-based level	1	0.62	0.31	0.00	1.00
	2	0.56	0.36	0.00	1.00
Behavioural-based level	1	0.51	0.29	0.00	1.00
	2	0.44	0.33	0.00	1.00

Note. $N = 29$

Table 4.6: Comparing interviewers' judging behaviour in the two interviews

	Test	p	Mean	SD	Min	Max
Criteria-compliance level	WSR	.260	0.71	0.35	0.17	1.00
Evidence-based level	WSR	.389	0.59	0.30	0.00	0.96
Behavioural-based level	WSR	.174	0.48	0.29	0.00	0.90

Note. 1. *PST* refers to paired samples t-test, and *WSR* refers to Wilcoxon signed-rank test; 2. $N = 29$.

$0.70, SD_2 = 0.37$) as well as the same extremes ($Min = 0.17, Max = 1.00$). To find out whether a statistically significant differences exists, a Wilcoxon signed-rank test is conducted on the paired samples and the results show that no significant difference ($Z = -1.127, p = .260$) is recognised when comparing the medians of criteria-compliance level in the two interviews ($Mdn_1 = 0.96, Mdn_2 = 0.96$). The descriptive statistics of the interviewers' averaged criteria-compliance level between the two interviews are thus worked out (see Table 4.6) that the criteria-compliance level varies from 0.17 (i.e., the interviewers make judgments about an applicant on 4 out of 23 given criteria) to 1.00 (i.e., the interviewers make judgments about an applicant on all the 23 given criteria) among different individuals, of which the mean is 0.71 ($SD = 0.35$).

2) Evidence-based level

According to the descriptive statistics, the evidence-based level of the interviewers in the first interview ($M = 0.62, SD = 0.31$) is slightly higher than that in the second interview ($M = 0.56, SD = 0.36$). The results of a Wilcoxon signed-rank test are shown in Table 4.6, implying that the difference between the medians of evidence-based level in the two interviews ($Mdn_1 = 0.75, Mdn_2 = 0.67$) is not statistically significant ($Z = -0.861, p = .389$). Based on the averaged evidence-based level of each interviewer in the two interviews, the mean of the dimension is 0.59 ($SD = 0.30$) for the 29 interviewers, while great difference can be perceived among individual interviewers concerning the value of the extremums ($Min = 0.00, Max = 0.96$).

3) Behavioural-based level

The descriptive statistics of behavioural-based level observed in the two interviews shows that the interviewers make more judgments based on behavioural clues when evaluating the first applicant ($M = 0.51, SD = 0.29$) than the second one ($M = 0.44, SD = 0.33$). A Wilcoxon signed-rank test is executed on the 29 interviewers to determine if a statistically significant difference exists in the median behavioural-based level between the two interviews, while no significant difference is recognised ($Z = -1.538, p = .174$). Hence, the averaged behavioural-based level of each interviewer between the two interviews is calculated, based on which a descriptive analysis is conducted on the overall performance of the 29 interviewers. As can be seen in Table 4.6, the mean behavioural-based level of the interviewers is 0.48 ($SD = 0.29$), which means that the interviewers make judgments about an applicant based on behavioural clues for nearly 11 out of 23 given criteria. It can also be interpreted that, while some interviewers follow up to 21 given criteria out of 23 and provide behavioural evidence to justify own judgments ($Max = 0.90$), others may scarcely comment on the given criteria with behavioural evidence ($Min = 0.00$).

4.2.2 Cognitive Evidence of Interviewers' Decision Patterns and Bias

Cognitive evidence is found in the interviewers' cognitive processes during decision making for all the judging behaviour investigated, including four types of decision patterns (i.e., stereotypes of ideal applicants, anchoring-and-adjusting, deception detection, and cue weighting) and three types of decision bias (i.e., similar-to-me effect, halo effect, and contrast effect). The results are reported mainly in a form of direct quotes from the interviewers' verbal protocols of own cognitive processes.

1) Stereotypes of ideal applicants

In spite of the job description and given criteria that interviewers are supposed

to follow when evaluating an applicant, interviewers may have their own beliefs of what an ideal applicant should be like as well (i.e., stereotypes of ideal applicants, see for example, Anderson, 1992; Weichselbaumer, 2004; Rice & Barth, 2016, 2017), which can vary a lot among individual interviewers. Previous studies on interviewers' stereotypes of ideal applicants are mainly based on interviewers' self-reports, while scholars like Schmitt (1976) suggest that "it would be more profitable to learn what these interviewers actually weighted in reaching their decision" rather than what they say they will weight.

When sorting through the protocols gathered from the experiment, it can be clearly seen that many interviewers rely at least part of their evaluation and decision making on own understanding and criteria regarding what kind of applicants they are (not) expecting for the target job position. For example, Interviewer 1 comments on how managers should balance between "being open to teammates' opinions" and "ensuring the completion of work" by saying that

> *It's important to recognize where people's ideas are useful, but it's also important to understand how to get people to do what is required of them ... sometimes it can mean that you've got to be really direct, very honest with teammates and let them know where the position is and they have got to work on what's been agreed rather than what they see themselves.*

And the interviewer does not appreciate the applicant's seeking for help from managers when teammates seem less cooperative. In contrast, the interviewer believes that

> *Um, if you're trying to build consensus, you're trying to get people on side, then talking to people first and getting their agreements to things is important rather than going to a manager and having people feel that their involvement is imposed on them rather than coming from a willingness to want to able to be involved.*

Interviewer 8 states a similar point as well and says that

> *So, I would possibly reject him at this point, um, primarily based on the point he made about going to his manager. Um, as a project manager, I would want someone to be able to take that responsibility and make decisions and delegate and be confident in themselves to do that. And that really raised a red flag with me in terms of this particular role in that particular client.*

The differences among individual beliefs regarding ideal applicants turn out to be even more apparent when interviewers express completely opposite opinions to the same piece of information extracted from an applicant's response. For example, when the second applicant mentions own preference for exercise as a relief of stress, some interviewers agree with the idea:

> *Interviewer 3: With regards to how he deals with stress, exercise is a really important and basic way of dealing with stress in the stress response cycle, and it's really good and healthy that he's noticed that.*

> *Interviewer 8: He (i.e., the applicant) mentioned physical exercise, and that can actually be a good strategy in the workplace, because going for a walk, getting out of the situation, giving yourself time to calm down can actually be quite a good strategy.*

In contrary, other interviewers may be less satisfied about the response:

> *Interviewer 5: . . . having exercises is not the answer I was looking for.*

> *Interviewer 18: It seems like this applicant prefers to divert his attention away from what has caused him pressure rather than to think about why he feels stressful or how to deal with it.*

2) Anchoring-and-adjusting

It is found in some studies that, rather than give a score after collecting all the information about an applicant, interviewers may start off with an initial judgement and adjust own perceptions of the applicant based on this starting point as they receive more related information (Madera & Hebl, 2012; Derous et al., 2016; Buijsrogge et al., 2021). Such proposition is tested mainly by comparing an interviewer's ratings given to the applicant at different stages of interview, whereas in this research exploring the cognitive process of interviewers during decision making, their rating styles can be directly observed to examine if an anchoring-and-adjusting approach is adopted.

Cognitive evidence is recognised supporting the existence of this decision-making pattern. For example, Interviewer 21 intends to rate an applicant on all the traits listed in the rating scale following the applicant's response to the first interview question. Then, after listening to the response to each of the subsequent interview questions, the interviewer checks the scores given according to the new information received and make revisions if necessary:

> *. . . He did not consider about the potential risk, and he did not think about a Plan B to ensure the completion of the task. So, I'm afraid I might have given the score a little bit early, and I would like to lower it down.*

3) Detection of deception

As stated in the literature review, applicants are found using fake information (i.e., deceptive impression management) to convince interviewers that they have the desirable traits as required by the job position. Therefore, another aspect of interviewers' judging behaviour drawing wide attention from scholars is their ability of detecting misinformation from applicants' responses (DePaulo & Pfeifer, 1986; Mann et al., 2004; Van Iddekinge et al., 2005; Reinhard et al., 2013; Roulin et al., 2015). Most of the studies investigating whether and how interviewers can detect applicants' deception collect

related data through questioning and prompting where interviewers are required to work on mock interviews and asked to point out where they think an applicant provides misinformation or whether they believe the applicant tells the truth in general. Recalling the potential drawbacks of the technique (see Section 3.1), this research investigates the interviewers' detection of applicant deception through the verbal protocols that represent the interviewers' ongoing thinking processes during interview. Thus, the process avoids warning the interviewers about the possibility of applicants' faking, of which the influence on interviewers' deceptive detection is not clear.

According to the content of the verbal protocols, it can be observed that some interviewers show an intention to question the reliability of the applicants' responses, especially when they perceive a lack of information they are expecting for. For example, Interviewer 13 proposes a need of asking prompting questions almost after each of the responses given by the first applicant:

> *I feel like he has portrayed himself as someone who gets a lot of work, but I don't know his role enough. He has mentioned that he would arrange reviews and make sure there were no gaps, so I want to ask something more like how do you interact with your team to ensure projects are on track in terms of like team meetings, one to one project tracking. Yes, I feel that more probing is needed on how he manages the team on a daily or weekly basis so I could understand the support he's providing.*

Similarly, Interviewer 8 doubts that part of the responses given by the second applicant is something cited from a book:

> *Ok, I think, with that one, he obviously had done some background reading for handling multiple tasks or projects... Should be. It was almost like he was reciting something from a book or a paper that he'd read, rather than talking very specifically about something that actually happened to him ... So, it's very much about 'this is what I should do' 'this is what I would do'*

> *'this is what good project managers should do in this situation'. It wasn't a very personal answer and I would have liked something a lot more personal. So, I would prompt and challenge a lot more for a specific example where he put what he said in theory into practice.*

4) Cue weighting

It is verified in many studies that interviewers may not give the same weight to different cues (Mayfield et al., 1980; Dougherty et al., 1986; Graves & Karren, 1992; Van Dam, 2003; DeGroot & Gooty, 2009). Perhaps the most straightforward evidence is the criteria-compliance level defined and investigated in this research which shows that 17 out of the 29 interviewers do not give a rate to all the traits listed in the rating scale. Such results imply that interviewers are likely to make hiring decisions based on only part of the given criteria. Furthermore, even for the traits considered and rated by the interviewers when evaluating an applicant, the interviewers may not necessarily assign the same weight to them. For instance, Interviewer 15 suggests that the given criteria are not equally significant when evaluating applicants for the target position by saying that

> *For me, I'd say planning and organisation skills and problem-solving skills are more essential for the current job position. So... Um... I would choose "accept with reservations".*

5) Similar-to-me effect

Plenty of studies prove the existence of similar-to-me effect in IDM that interviewers tend to rate applicants with similar biographical backgrounds, attitudes or personalities to themselves more favourably (Baskett, 1973; Frank & Hackman, 1975; Anderson, 1992; Sears & Rowe, 2003). While biographical information and personalities are not collected investigated in this research and the interviewers' final decisions

are not brought into analysis, evidence confirming the similar-to-me effect in IDM can be found at a cognitive level where interviewers recognise and highlight a similarity between themselves and an applicant in terms of particular experience or attitude. For instance, when learning about the case of the stressful working situation described by the first applicant, Interviewer 24 expresses an empathy for the applicant and recalled a similar experience of own:

> *It also reminds me of my first job. I was responsible for organising an opening ceremony that would go live. Everything was fine in the rehearsal, but you know, unexpected situations always happen. And I was really nervous at that time and could hardly control my temper.*

Further supporting examples can be the previously discussed comments where interviewers share the same preferences or opinions with an applicant upon the idea of stress management through exercising.

6) Halo effect

A positive trait or characteristic of an applicant can influence interviewers' judgements on other unrelated factors of the applicant in a favourable way (Dougherty et al., 1986; Fat, 2000; Levine & Feldman, 2002; Sears & Rowe, 2003), which is defined as halo effect. In an early study on halo effect in employment interview, Crissy and Regan (1951) suggested two ways that could be used to test the existence of halo in interviewers' DM. One was correlation analysis, and the other one in a more straightforward way referred to an analysis of the evidence cited by interviewers to support their judgments. Most studies on the halo effect in IDM adopt the former one (Parsons & Liden, 1984; Zysberg & Nevo, 2004), while this research proves the existence of halo effect by examining the clues cited by the interviewers. For example, while Interviewer 21 claims that the second applicant does not provide sufficient information for certain criteria, the interviewer would like to give a modest score to these traits due to a generally positive feeling about the applicant:

So, planning and organisation, I don't think he has talked much about that. But I would probably give him a 3, because he has given quite good answers to these questions, especially the two projects he has mentioned. Yes, maybe I'll do the same to other items as well. ... Productivity improvement, I didn't quite get that, but I will give a score of two.

7) Contrast effect

While interviewers are expected to interview and evaluate each applicant individually and separately, results of numerous empirical studies confirms that their ratings are usually affected by the preceding applicant's quality, which is called contrast effect (Wexley et al., 1972; Landy & Bates, 1973; Kopelman, 1975; Schuh, 1978; Anderson, 1992; Fat, 2000). Such impact can be especially strong if the performance of the previous applicant is extremely poor or good. Most studies of contrast effect in IDM them are based on controlled trials where the quality of the previous applicant is manifested and interviewers' ratings given to the current applicant are analysed and compared. In contrary, this research seeks for evidence of contrast effect that is directly extracted from ICP. For example, Interviewer 10 keeps comparing the two applicants throughout the interview process:

Um, comparing these two, you know, the first applicant gave an example where something didn't go to plan, but he was actually able to rescue the situation. Whereas in this case, the project sort of um, failed and that was the end of it. Nothing could be done about it. So that's a shame.

or

To me, it sounds like the first applicant is a lot more of a people person ... whereas the second applicant seems to be a lot more driven by the outcome ... Um, he doesn't strike me as a person who likes to be disagreed with. And

he hasn't spoken at all about, um, I mean beyond, you know, making some sort of matrix, he hasn't spoken an awful lot about how he handles the people that work for him. So, I think that's the key difference between these two, one is a people person, one seems to be more of a business person. So, this second applicant, I think, would be very good if he's working in such a way that he can just do what he needs to do, but I'm not sure that's the skill set I would want for someone who's in charge of others.

4.3 Examining the Associations between ICP and Contextual Factors

In this section, the findings regarding the possible associations between ICP (i.e., four ICP characteristics and three dimensions of interviewers' judging behaviour) and the four contextual factors are summarised, namely interviewers' gender (gender), national region of interview experience (national region), experience level as interviewer (experience), and rating scale type.

4.3.1 Impact of Contextual Factors on ICP Characteristics

Cognitive activeness. To determine if a statistically significant difference exist in the interviewers' cognitive activeness between the two groups of each contextual factor, an independent-samples t-test is conducted for gender, national region, and experience, while the impact of rating scale type on interviewers' cognitive activeness is examined by a Mann-Whitney U Test. According to the results in Table 4.7, a statistically significant difference is detected only in mean cognitive activeness of the interviewers due to the different national regions of interview experience $[t(27) = 2.276, p = .031]$. Particularly, the cognitive activeness of UK interviewers ($M = 125, SD = 45$) is higher than CN interviewers ($M = 86, SD = 49$). No statistically significant difference is recognised in the means of cognitive activeness between male and female interviewers

Table 4.7: Comparing interviewers' cognitive activeness in different contextual groups

Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
Gender	M	6	103	46	IST	-	.923
	F	23	105	52			
Region	UK	14	125	45	IST	-	**.031***
	CN	15	86	49			
Experience	E	11	112	44	IST	-	.531
	N	18	100	55			
Rating Scale Type	BA	15	104	50	MWU	80	.813
	OG	14	106	52		103	

Note. IST refers to independent samples t-test, and *MWU* refers to Mann-Whitney U Test.

$[t(27) = -0.098, p = .923]$ or between experienced and novice interviewers $[t(27) = 0.635, p = .531]$, and the median cognitive activeness of the interviewers using the two different types of rating scale is not significantly different either ($U = 99.5, Z = -0.24, p = .813$).

Cognitive diversity. An independent-samples t-test is carried out for each contextual factor to check whether it leads to statistically significant difference in mean cognitive diversity of the interviewers. As can be seen in Table 4.8, no statistically significant difference is found in mean cognitive diversity between the two groups of gender $[t(27) = 0.311, p = .758]$, national region $[t(27) = 1.613, p = .118]$, experience $[t(27) = 0.658, p = .516]$, or rating scale type $[t(20.009) = 0.276, p = .785]$.

Frequency of information-processing strategy. The association is examined separately between the contextual factors and the frequency of each information-processing strategy as well as the irrelevant comments, of which the outcomes are shown in Table 4.9. As for interpreting strategy, its associations with interviewers' gender and experience are investigated through Wilcoxon signed-rank tests, while an independent-

Table 4.8: Comparing interviewers' cognitive diversity in different contextual groups

Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
Gender	M	6	0.68	0.09	IST	-	.758
	F	23	0.66	0.13			
Region	UK	14	0.70	0.08	IST	-	.118
	CN	15	0.63	0.15			
Experience	E	11	0.68	0.09	IST	-	.516
	N	18	0.65	0.14			
Rating Scale Type	BA	15	0.67	0.09	IST	-	.785
	OG	14	0.66	0.16			

Note. IST refers to independent samples t-test.

samples t-test is adopted for the other two contextual factors. The results $[t(27) = 3.062, p = .005]$ imply that the interviewers possessing interview experience in UK ($M = 51.13\%, SD = 4.59\%$) execute interpreting strategy at a significantly higher frequency than those with interview experience in CN ($M = 42.90\%, SD = 9.02\%$). In contrast, gender ($U = 42, Z = -1.454, p = .158$), experience ($U = 66, Z = -1.483, p = .146$) and rating scale type $[t(27) = 0.396, p = .695]$ seem to cause little difference in the frequency of interpreting strategy. At the same time, the frequency of judging strategy is also found affected by the interviewers' national regions of interview experience according to the results of a Mann-Whitney U test ($U = 37, Z = -2.968, p = .002$) but in an opposite trend. Specifically, UK interviewers use judging strategy less ($Mdn = 42.70\%$) than CN interviewers ($Mdn = 52.40\%$). Besides, an independent-samples test is performed on the frequency of self-monitoring strategy and a Mann-Whitney U test is conducted on the frequency of irrelevant comments, whereas no statistically significant association is recognised between the four contextual factors and either of these two dependent variables.

Evidential focus. Independent-samples t-tests are utilised to find out if the four contextual factors result in the means of interviewers' focus on the three different

Table 4.9: Comparing frequency of information-processing strategy in different contextual groups

	Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
Int.	***Gender***	M	6	43.89%	4.52%	MWU	42.18%	.158
		F	23	47.65%	8.88%		49.96%	
	Region	UK	14	51.13%	4.59%	IST	-	**.005***
		CN	15	42.90%	9.02%			
	Experience	E	11	49.40%	7.39%	MWU	50.41%	.146
		N	18	45.33%	8.56%		47.12%	
	Rating Scale Type	BA	15	47.47%	6.55%	IST	-	.695
		OG	14	46.23%	9.97%			
Jud.	***Gender***	M	6	48.02%	4.21%	MWU	47.57%	.414
		F	23	47.14%	9.92%		45.39%	
	Region	UK	14	42.40%	4.49%	MWU	42.70%	**.002***
		CN	15	51.91%	9.79%		52.40%	
	Experience	E	11	44.72%	8.44%	MWU	41.83%	.134
		N	18	48.91%	9.16%		46.93%	
	Rating Scale Type	BA	15	46.58%	6.36%	MWU	45.39%	.949
		OG	14	48.11%	11.35%		46.25%	
SM	***Gender***	M	6	5.59%	3.55%	IST	-	.222
		F	23	4.09%	2.36%			
	Region	UK	14	4.85%	2.49%	IST	-	.388
		CN	15	3.98%	2.80%			
	Experience	E	11	4.36%	2.39%	IST	-	.951
		N	18	4.42%	2.86%			
	Rating Scale Type	BA	15	4.61%	2.23%	IST	-	.663
		OG	14	4.17%	3.10%			

types of information (i.e., performance-related information, response-related information, other information) during decision making, except that the association between

(continued table)

	Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
Irr.	***Gender***	M	6	2.50%	1.89%	MWU	2.34%	.071
		F	23	1.12%	1.56%		0.30%	
	Region	UK	14	1.62%	2.08%	MWU	0.83%	.847
		CN	15	1.21%	1.28%		0.89%	
	Experience	E	11	1.52%	2.22%	MWU	0.30%	.877
		N	18	1.34%	1.35%		1.12%	
	Rating Scale Type	BA	15	1.34%	1.72%	MWU	0.78%	.847
		OG	14	1.48%	1.73%		1.01%	

Note. IST refers to independent samples t-test, and *MWU* refers to Mann-Whitney U Test.

interviewers' national region and focus on response-related information whose medians are compared through a Mann-Whitney U test (see Table 4.10). According to the test results, interviewers' focus on performance-related information is associated with both interviewers' national region $[t(27) = -2.798, p = .009]$ and experience $[t(27) = -2.739, p = .011]$. On one hand, the interviewers with previous interview experience in CN focus significantly more on the information about an applicant's work performance ($M = 75.01\%, SD = 8.51\%$) than the UK interviewers (M = 65.22%, SD = 10.30%). On the other hand, the novice interviewers ($M = 74.05\%, SD = 8.65\%$) are found paying more attention to performance-related information than the experienced interviewers ($M = 64.13\%, SD = 10.69\%$). Meanwhile, statistically significant association is recognised between the interviewers' focus on response-related information and the types of rating scale they use during interview process $[t(27) = 2.071, p = .048]$. To be specific, interviewers pay significantly more attention to response-related information when evaluating an applicant with BARS ($M = 22.95\%, SD = 12.05\%$) than with OGRS ($M = 14.55\%, SD = 9.55\%$). No statistically significant association is found between any other paired variables.

Table 4.10: Comparing interviewers' evidential focus in different contextual groups

	Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
Per.	***Gender***	M	6	67.09%	11.51%	IST	-	.413
		F	23	11.51%	10.33%			
	Region	UK	14	65.22%	10.30%	IST	-	**.009***
		CN	15	75.01%	8.51%			
	Experience	E	11	64.13%	10.69%	IST	-	**.011***
		N	18	74.05%	8.65%			
	Rating Scale Type	BA	15	68.25%	11.17%	IST	-	.286
		OG	14	72.47%	9.64%			
Res.	***Gender***	M	6	21.96%	14.38%	IST	-	.475
		F	23	18.10%	10.92%			
	Region	UK	14	22.06%	12.63%	MWU	20.89%	.186
		CN	15	15.94%	9.94%		15.29%	
	Experience	E	11	22.59%	13.82%	IST	-	.228
		N	18	16.64%	9.62%			
	Rating Scale Type	BA	15	22.95%	12.05%	IST	-	**.048***
		OG	14	14.55%	9.55%			
Oth.	***Gender***	M	6	10.95%	5.68%	IST	-	.960
		F	23	10.79%	7.13%			
	Region	UK	14	12.72%	5.67%	IST	-	.146
		CN	15	9.04%	7.38%			
	Experience	E	11	13.28%	6.69%	IST	-	.128
		N	18	9.32%	6.53%			
	Rating Scale Type	BA	15	8.80%	4.15%	IST	-	.109
		OG	14	12.98%	8.38%			

Note. IST refers to independent samples t-test, and *MWU* refers to Mann-Whitney U Test.

4.3.2 Impact of Contextual Factors on Interviewers' Judging Behaviour

The test results of whether the four contextual factors lead to statistically significant difference in the three dimensions of interviewers' judging behaviour are presented in Table 4.11.

Criteria-based level. Mann-Whitney U tests are utilised to compare medians criteria-based level between the two groups of gender ($U = 57, Z = -0.671, p = .546$), national region ($U = 73.5, Z = -1.429, p = .172$), experience ($U = 68, Z = -1.448, p = .173$) or rating scale type ($U = 77.5, Z = -1.247, p = .234$), whereas no statistically significant difference is found.

Evidence-based level. Mann-Whitney U tests are also conducted on the interviewers' evidence-based level except for rating scale type, which is examined by an independent-samples t-test. The results imply that no significant impact is caused by gender ($U = 64, Z = -0.269, p = .813$), national region ($U = 61, Z = -1.922, p = .057$), experience ($U = 89, Z = -0.450, p = .674$) or rating scale type [$t(27) = 1.970, p = .059$] on the dimension.

Behavioural-based level. For behavioural-based level, the corresponding means in the two groups of gender [$t(27) = -0.030, p = .976$], experience [$t(27) = -0.827, p = .416$] and rating scale type [$t(27) = 2.352, p = .026$] are compared based on independent-samples t-tests, while a Mann-Whitney U test is adopted to examine whether the median behavioural-based level significantly differs between the two national regions ($U = 73, Z = -1.397, p = .172$). Statistically significant association is confirmed only between behavioural-based level of the judgments made by the interviewers and the corresponding types of rating scale they use during interview. It can be interpreted from the results that the interviewers using BARS turn out to provide behavioural clues for own judgments on significantly more given criteria ($M = 0.59, SD = 0.25$) than those adopting OGRS ($M = 0.36, SD = 0.28$).

Table 4.11: Comparing interviewers' judging behaviour in different contextual groups

	Factor	Group	*N*	*M*	*SD*	Test	*Mdn*	*p*
CCL	***Gender***	M	6	0.79	0.32	MWU	0.97	.546
		F	23	0.68	0.36		0.93	
	Region	UK	14	0.76	0.33	MWU	1.00	.172
		CN	15	0.66	0.38		0.93	
	Experience	E	11	0.82	0.27	MWU	1.00	.173
		N	18	0.63	0.39		0.88	
	Rating Scale Type	BA	15	0.62	0.35	MWU	0.63	.234
		OG	14	0.80	0.34		1.00	
EBL	***Gender***	M	6	0.57	0.30	MWU	0.55	.813
		F	23	0.60	0.31		0.73	
	Region	UK	14	0.70	0.23	MWU	0.79	.057
		CN	15	0.48	0.32		0.48	
	Experience	E	11	0.58	0.25	MWU	0.61	.674
		N	18	0.59	0.33		0.67	
	Rating Scale Type	BA	15	0.69	0.25	IST	-	.059
		OG	14	0.48	0.32			
BBL	***Gender***	M	6	0.47	0.28	IST	-	.976
		F	23	0.48	0.29			
	Region	UK	14	0.57	0.27	MWU	0.68	.172
		CN	15	0.39	0.29		0.37	
	Experience	E	11	0.42	0.25	IST	-	.416
		N	18	0.51	0.31			
	Rating Scale Type	BA	15	0.59	0.25	IST	-	**.026***
		OG	14	0.36	0.28			

Note. IST refers to independent samples t-test, and *MWU* refers to Mann-Whitney U Test.

4.4 Summary

This chapter reports the key findings in accordance with the research questions proposed, including the elements occurring in ICP, the descriptive statistics of the four ICP characteristics identified and the three dimensions of judging behaviour, the cognitive evidence detected for four decision patterns and three decision bias, and the associations recognised between the four contextual factors and ICP.

1) Elements occurring in ICP

As presented in the coding scheme (see Table 4.1) developed, 17 categories of cognitive actions occurring in ICP during decision making are identified, which can be further categorised into three information-processing strategies, namely interpreting strategy, judging strategy, and self-monitoring strategy. The prevalence of these cognitive actions among the interviewers varies greatly in the sense that, while several cognitive actions are found in the cognitive processes of all the 29 interviewers, some other cognitive actions are adopted by much fewer interviewers. The three information-processing strategies are recognised in the cognitive processes of almost every interviewer except that one has not executed self-monitoring strategy. As for the information that the interviewers pay attention to during decision making, four types of information are found in ICP, including information related to the design of interview task, information derived from applicants' responses, personal sources of information obtained before the interview, and information produced by the interviewers earlier during decision making. Besides, a few types of irrelevant comments made by the interviewers during decision making are also found. These elements occur in ICP in a complex and interactive combinations, of which no fixed or regular sequence is recognised.

2) ICP characteristics

The investigation into the four ICP characteristics shows that the performance of

interviewers varies among individuals while no significant difference is found between their performances within the two different interviews. In general, the averaged cognitive activeness of the interviewers is 108 ($SD = 50$) as measured by the modified number of coded cognitive actions. The mean cognitive diversity of the interviewers is 0.66 ($SD = 0.12$), which means that about 11 out of 17 categories of the cognitive actions are adopted by the interviewers during an interview. As for the frequency of information-processing strategies, the interviewers execute interpreting strategy in around 46.87% of ICP ($SD = 8.25\%$), make or justify judgements in around 47.32% of ICP ($SD = 8.98\%$), monitor own decision-making performance in about 4.40% of ICP (SD = 2.64%), whereas the rest 1.41% of ICP ($SD = 1.69\%$) produces irrelevant comments. Besides, the interviewers concentrate much more on the information about applicants' work performance ($M = 70.29\%, SD = 10.50\%$) compared to the information related to the features of applicants' responses ($M = 18.89\%, SD = 11.54\%$) and other information ($M = 10.82\%, SD = 6.76\%$).

3) Interviewers' judging behaviour

Similar findings are also acquired regarding the three dimensions of interviewers' judging behaviour that the interviewers' performance is relatively stable during different interviews whereas discrepancies can be perceived among individuals. Overall, the criteria-compliance level is 0.71 ($SD = 0.35$), which means that the interviewers follow 71.0% of the given criteria during decision making. An averaged level of evidence-based level among the interviewers is 0.59 ($SD = 0.30$), implying that around 59.0% of the judgements that the interviewers make on the given criteria are justified with specific evidence. As for the behavioural-based level which measures the extent to which the interviewers base own judgements on behavioural clues as required by past behaviour questions, it is found that only nearly a half of the given criteria ($M = 0.48, SD = 0.29$) are evaluated by the interviewers according behavioural clues. Besides, cognitive evidence is recognised in ICP verifying the existence of all the four decision patterns and three decisions bias.

4) Associations between IDM and the contextual factors at a cognitive level

When examining the associations between the four contextual factors and IDM at a cognitive level, the impact on either the ICP characteristics or the dimensions of interviewers' judging behaviour is identified for interviewers' national region of interview experience, experience level and type of rating scale used during interview. To be specific, the interviewers with previous interview experience in UK turn out to be more cognitively active, adopting interpreting strategy more frequently and pay relatively less attention to performance-related information during decision making when compared to CN interviewers. The influences of rating scale type on ICP are twofold in the sense that the interviewers using BARS with informative behavioural-anchored descriptors pay a relatively greater attention to performance-related information and evaluate more given criteria based on behavioural clues than those using OGRS during interview. When comparing the performance of experienced and novice interviewers, only the interviewers' focus on performance-related information is found that a significantly greater proportion of novice interviewers' cognitive processes concentrates on performance-related information, whereas the focus of experienced interviewers' cognitive processes on either performance-related or response-related information are more balanced. No statistically significant difference caused by the interviewers' gender in the cognitive variables is found.

Chapter 5

Discussion and Conclusions

While decision making is in essence a human cognitive process where a series of information-processing activities take place, the existent studies on IDM pay little attention to interviewers' cognitive processes during decision making, especially based on the data that can effectively represent interviewers' real thinking processes. Such a gap may conceal the chance to improve interview settings and the outcomes of IDM, thus this research aims to explore interviewers' cognitive process when making hiring decisions by seeking for the answers to the four sets of research questions listed in the introduction chapter. A scenario experiment with TAM embedded in a mock interview task is developed and applied to data collection, where 29 interviewers with diversified background are involved and produce data of ICP in the form of verbal protocols. A multi-stage processing and analysing procedure is conducted on the protocols and the findings generated provide the answers to all the research questions raised. As the ending chapter of the dissertation, this part of context consists of three sections, including **Section 5.1** discussing the key findings generated, **Section 5.2** highlighting both the contributions to academic knowledge and practical implications of this research, and **Section 5.3** discussing the potential limitations and opportunities of future research.

5.1 Discussion of the Key Findings

The findings obtained in this research provide the answers to all the research questions brought up, for which a discussion is presented below.

5.1.1 Elements of ICP

Resembling many other TAM-based studies on human cognitive process, the primary interest in ICP lies in its constitution, such as the information processed by the interviewers and how the information is processed (Eveland Jr & Dunwoody, 2000; Cumming et al., 2002; Reicks et al., 2013; Han, 2017). Four types of information (RQ1a) are found in ICP, including that related to the design of interview task (e.g., task requirements, job description, interview questions, and rating criteria), derived from applicants' responses (e.g., applicants' work performance, features of responses), personal sources of information obtained before the interview (e.g., personal experience, knowledge, beliefs), and produced by the interviewers earlier during decision making (e.g., comments, scores, prompting questions, inferences). Nevertheless, further investigation is not conducted into the specific information processed by the interviewers which may vary greatly with various factors (e.g., industry, company, job position, target competence, evaluation criteria) while this research intends to explore ICP during decision making in general regardless of the specific contexts.

As for RQ1b, the coding scheme (see Table 4.1) developed during protocol processing summarises 17 categories of cognitive actions identified in ICP to describe how the interviewers process the information in their cognitive processes. Moreover, in order to investigate the structure of ICP at a higher level, each cognitive action is further categorised into different information-processing strategies according to its immediate purpose. Following the work of Cumming et al. (2002) that investigates essay rating through TAM, three information-processing strategies are identified covering all the 17 cognitive actions, namely interpreting strategy (i.e., interpret and extract information

from experimental materials or the applicant's responses), judging strategy (i.e., make or justify a judgment on the applicant), and self-monitoring strategy (i.e., reflect on own decision-making process). Besides, a few types of irrelevant comments made by the interviewers during decision making are also found.

The prevalence and sequence of both the cognitive actions and information-processing strategies are examined, where a difference in ICP among individuals is recognised that not all these cognitive actions are adopted by each interviewer and the proportion of interviewers executing each cognitive action varies from 24.1% to all the twenty-nine interviewers. While at a strategic level, the prevalence of each information-processing strategy is similar among the interviewers in the sense that almost all the interviewers perform the three categories of strategies. It is also found that these elements occur in ICP in a complex and interactive combinations, of which no fixed or regular sequence is recognised. The phenomenon may partially result from the lack of structure in the interviewers' decision making especially when the task itself is less structured (Campion et al., 1997). Another possible cause of the intertwined elements of ICP observed is individual expression styles which are less likely to be highly regular and fixed (Van Someren et al., 1994).

5.1.2 Characteristics of ICP

In addition to the prevalence among different interviewers, four indicators are constructed to describe the characteristics of ICP at an individual level, namely cognitive activeness (RQ2a), cognitive diversity (RQ2b), frequency of information-processing strategy (RQ2c) and evidential focus (RQ2d).

The decision time of interviewers is often examined in IDM research to imply how quickly interviewers reach their final decisions (Tullar et al., 1979; Judge et al., 2000; Huffcutt, 2010; Frieder et al., 2016) and it is also a common indicator of cognitive effort a subject makes when completing a task (Cooper-Martin, 1994). Such a variable, however, may be affected by irrelevant factors (e.g., mental distractions) especially

when the data collection process is not strictly monitored and controlled. Hence, a characteristic of ICP called cognitive activeness is proposed to measure how cognitively active an interviewer is during decision making, which to certain extent implies the cognitive efforts an interviewer made to complete the task (Bettman & Park, 1980; Cooper-Martin, 1994; Cumming et al., 2002; Li & He, 2015). The result shows that the cognitive activeness varies among individual interviewers, thus it can be interpreted that the cognitive efforts spent by the interviewers on the same interview task differ a lot.

While the prevalence of cognitive action examines whether the cognitive action is adopted by all the interviewers, cognitive diversity investigates whether an interviewer executed more or fewer types of cognitive actions. The outcomes indicate that, while some interviewers' cognitive processes are more diversified with a wider range of cognitive actions executed to facilitate their decision making, other interviewers may be more simplex in information processing.

Following existent studies on cognitive processes (e.g., Kivetz & Simonson, 2000; Cumming et al., 2002; Jaspers et al., 2004; Barkaoui, 2011), the other two ICP characteristic are defined to explore the constitution of ICP by figuring out how frequently an interviewer adopts each of the three strategies (i.e., frequency of information-processing strategy) or focusing on different types of clues (i.e., evidential focus). It is found that, on average, the interviewers balance their cognitive processes that aims to interpret and extract information (46.87%) with that concentrate on making or justify judgements (47.32%) while also occasionally monitoring own decision-making performance (4.40%). A small proportion of ICP (1.41%) where the interviewers produce irrelevant comments is also recognised. As for evidential focus, the interviewers pay the majority attention of own (70.29%) to the information about applicants' work performance while the features of applicants' responses (18.89%) and other information (10.82%) are considered much less during decision making.

In general, the results suggest that discrepancies exist among individual interviewers within the same interview, which may to a certain extent cause differences in final

decisions. However, the interviewers' performance regarding each of the four ICP characteristics turns out to be similar between the two interviews, suggesting a relatively stable cognitive performance when interviewers work on different decision-making task.

5.1.3 Interviewers' Judging Behaviour at a Cognitive Level

Vast numbers of studies have been conducted to figure out how IDM can be improved, and structured interview seems to be the most impactful one widely utilised in real-world practice (Campion et al., 1997; Chapman & Zweig, 2005; Huffcutt et al., 2013; Kausel et al., 2016). Nevertheless, it is noticed that the measurement of IDM performance (McDaniel et al., 1994; Conway et al., 1995; Moscoso, 2000; Woods et al., 2020) and the identification of their decision patterns and bias (Graves & Powell, 1995; Hebl & Kleck, 2002; Florea et al., 2019; Buijsrogge et al., 2021) are constrained to the decision outcomes with little attention paid to the process through which interviewers reach their decisions. To promote a more in-depth monitoring of IDM, this research constructs three dimensions to measure whether interviewers make judgements in the style as expected and searches for direct evidence from ICP that verifies the existence of various heatedly discussed decision patterns and bias in IDM.

Since interviewers are expected to structure their decision making following the given criteria (Campion et al., 1997, Levashina et al., 2014), the first dimension of interviewers' judging behaviour called criteria-compliance level (RQ3a) is developed to measure the extent to which the interviewers evaluate an applicant with the criteria provided. It is found that, while some interviewers take all the given criteria into consideration when making decisions, others may base the assessment of an applicant almost completely on the standards or preferences of own, which may lead to bias and unfairness in the recruitment. On average, the interviewers are found following most of the given criteria.

Interviewers are also expected to rely their judgements on specific clues rather than intuitions (Highhouse, 2008), whereas it is repeatedly confirmed that their decisions

are vulnerable to multiple bias, like similar-to-me effect (Anderson & Shackleton, 1990; Anderson, 1992; Sears & Rowe, 2003), halo effect (Parsons & Liden, 1984; Zysberg & Nevo, 2004; Thomas & Reimann, 2022) and contrast effect (Hakel et al., 1970; Schuh, 1978; Fat, 2000). Thus, ICP is examined to find out whether interviewers justify own judgements with specific evidence, i.e., evidence-based level (RQ3b), and the results show that only a slightly more than half of the interviewers' judgments made on the given criteria are supported by certain clues. Moreover, some interviewers justify nearly all of the judgments, but it is also possible that interviewers evaluate an applicant completely based on general perceptions and intuitions.

The third aspect of interest concerning interviewer's judging behaviour is that, while past behaviour questions are widely adopted because applicants' past behaviour is believed to be the best predictor of their future performance (Barrick et al., 2000; Huffcutt et al., 2001; Moscoso & Salgado, 2001; Bangerter et al., 2014), interviewers may not always make judgments according to behavioural clues. To quantitatively measure the corresponding performance of an interview, a dimension called behavioural-based level (RQ3c) is constructed and it is found that interviewers evaluated around half of the given criteria based on behavioural clues identified in applicants' responses. Individual performance differs greatly in the sense that the proportion of given criteria judged with behavioural clues ranges from zero to a hundred percent.

As the most primary responsibility of interviewers to judge an applicant's qualification for the target job, interviewers' judging behaviour turns out to vary greatly among the individuals in terms of the three dimensions constructed. Such discrepancies in judging behaviour indicate that not all the interviewers structure own decision-making process as expected, which may impede interrater reliability of decision outcomes (Conway et al., 1995; Campion et al., 1997; Chen et al., 2010; Huffcutt et al., 2013) and even lead to unfairness in recruitment (Bragger et al., 2002; McCarthy et al., 2010; Levashina et al., 2014). Meanwhile, the comparison between the two interviews shows that an interviewer's judging behaviour within different interviews are similar, implying the possibility that the characteristics of ICP are inherent in interviewers' cognitive style

when working on interview tasks (at least of the current kind adopted in the research) rather than being determined by the different applicants interviewed.

Besides, while a broad range of decision patterns and bias are heatedly discussed in IDM research, the conclusions of most corresponding studies rely significantly on comparing the decisions made by interviewers in controlled trials while direct evidence is almost absent. When investigating ICP as presented in interviewers' verbal protocols, cognitive evidence is searched for four decision patterns (i.e., stereotypes of ideal applicants, anchoring-and-adjusting, detection of deception in applicants' responses, cue weighting) and three decisions bias (i.e., similar-to-me effect, halo effect, contrast effect) from interviewers' thinking processes during interview (RQ3d). The findings support the existent studies by verifying the existence of these decision patterns and bias with direct evidence quoted from ICP.

5.1.4 Impact of Contextual Factors on ICP during Decision Making

As suggested by massive empirical studies on IDM, the decisions made by interviewers can be affected by various contextual factors, among which interviewers' gender (Parsons & Liden, 1984; Raza & Carpenter, 1987; Chapman & Rowe, 2001), experience level as interviewer (Hess, 2013; Roulin et al., 2015; Frieder et al., 2016), and type of rating scale used during interview process (Fay & Latham, 1982; Maurer, 2002; Klehe et al., 2008; Melchers et al., 2011) are examined for their associations with ICP (RQ4). Considering the possible impact of regional culture (Sanchez-Burks et al., 2006; Manroop et al., 2013) and employment laws (Williamson et al., 1997; Morgeson et al., 2008) on IDM, another factor called national region of interview experience is also brought into analysis.

The impact on ICP during decision making is recognised for all the contextual factors except for the gender of interviewer. As for the national region of interview experience, it has been hardly taken into the investigation of IDM and, as shown in the literature review, many empirical studies do not even specify the source of data regard-

ing the geographic origins where the studies take place or the background information of the interviewers involved in data collection. Notably, the statistical results generated in this research indicate the potential impact caused by interviewers' national region of interview experience on their decision-making performance at a cognitive level. First, it is found that the interviewers with previous interview experience in UK are more cognitively active, spending relatively more efforts on collecting information through interpreting strategy more frequently during decision making when compared to CN interviewers. Such difference may be derived from the difference in the legislation in the two countries that protects applicants' equal employment opportunity. Specifically, while the development of anti-discrimination laws in China is still at an embryonic stages (Ayalew, 2020) and even Chinese people' capacity to perceive discrimination is limited (Lu, 2014), a series of related laws and regulations are well established in UK, such as the Sex Discrimination Act 1975 (SDA), the Race Relations Act 1976 (RRA), the Disability Discrimination Act 1995 (DDA), the Employment Equality (Religion or Belief) Regulations 2003 and the Employment Equality (Gender) Regulations 2006. These norms and acts requires recruitment practitioners like interviewers to pay great attention not only to outcome fairness but also to procedural justice (Kossek & Pichler, 2006; Schleicher et al., 2006; Alonso et al., 2017), thus encouraging them to make judgements based on related evidence and keep a record of the procedure (e.g., note-taking) to defend own decisions.

The influences of rating scale type on ICP are twofold. On one hand, interviewers using a behaviourally-anchored rating scale with behaviourally-anchored descriptors presented for each criterion are found paying more attention to the features of applicants' responses than those using an open grid rating scale. This may because the detailed descriptors serve as reminders of the target information to seek for when evaluate a particular trait of an applicant (Schwab et al, 1975; Tziner et al., 2000; Debnath et al., 2015), thus it encourages interviewers to query or comment more on the response features, such as comprehensiveness and relevance. On the other hand, interviewers using BARS also present a higher behavioural-based level and evaluate more given criteria based on behavioural clues of applicants. Such results confirm the proposed

advantage of BARS that, by specifying the target behaviours of ideal applicants, the behaviourally-anchored descriptors presented in BARS make it easier for interviewers to focus more on applicants' concrete behavioural clues and match them with evaluation criteria (Jacobs, et al., 1980; Schmitt et al., 1991; Debnath et al., 2015).

When comparing the decision-making performance of experienced and novice interviewers at a cognitive level, difference is only recognised in their concentration on performance-related information in the sense that the majority of novice interviewers' attention is paid to performance-related information while experienced interviewers keep an on the quality of applicants' responses. This difference observed might be a result from the less knowledge and well-developed standards established by novice interviewers about the features of ideal responses to interview questions, which restrains the chance for them to comment on the response quality.

5.2 Contributions and Implications

The findings generated in this research can both contribute to the academic knowledge of IDM and be used to facilitate the real-world practice from several perspectives. The contributions to academic knowledge lie in four aspects. First, this research becomes the first to provide an overall prospect of interviewers' thinking processes during decision making, which is absent in previous research on IDM. In particular, the information types and cognitive actions involved in ICP are unveiled, which can be further investigated to serve various interests in IDM, for example, to model the decision-making path of expert interviewers.

Secondly, the characteristics of ICP from several angles are identified and quantified, which not only enrich the knowledge of IDM process but can also be introduced as variables into future research on IDM either as indicators of IDM performance (e.g., cognitive effort) or as intermediates to examine how various factors affect interviewers' decision outcomes by altering the characteristics of ICP.

Thirdly, responding to the situation that interviewers are expected to follow certain rules when making decisions (e.g., comply with given criteria, base own judgements on evidence, evaluate applicants according to behavioural clues), dimensions are constructed to measure interviewers' judging behaviour at a cognitive level, which break the limits of investigations and evaluations of IDM on the properties of decision outcomes.

Fourthly, the associations are examined between four contextual factors and ICP, which not only demonstrates how the knowledge of ICP generated can facilitate the monitoring of IDM at a cognitive level, but also provides hints about how factors like national regions of interview experience, experience level and rating scale type may lead to different hiring decisions by affecting ICP. Moreover, this is the first research that propose to investigate the national regions of interview experience on IDM with statistically significant impact recognised on several aspects of interviewers' ICP during decision making (e.g., cognitive activeness, frequency of information-processing strategy, evidential focus), which points out a valuable direction to exploit in IDM research.

Additionally, different from previous studies claiming the existence of typical decision patterns and bias in IDM based on the analysis of decision outcomes, this research supports the findings by providing direct evidence extracted from interviewers' thinking processes during decision making. Such observations can not only be used as auxiliary evidence in the identification of these decision patterns and bias, but the observations themselves can be further analysed to explore and evaluate IDM performance, for instance, how frequently these bias occur in IDM and how it is influenced by different factors (e.g., interview settings).

Last but not the least, as a pioneering work that introduces TAM to the study on IDM, this research not only confirms the feasibility of applying TAM to investigate ICP, but it also demonstrates a detailed procedure of experiment development and validation that can be transferred to many other interview contexts (e.g., different industries, job positions, interview types) to serve specific academic interests in IDM.

On the part of practical implications, the output of this research can be expected to assist the promotion of IDM in real-world practice in two ways. On one hand, the knowledge regarding cognitive process of decision making can assist the development of various decision-based systems (Wang & Ruhe, 2007), like cognitive informatics, software agent systems, expert systems, and decision support systems, and so is the knowledge of ICP. In particular, there is a growing interest in applying artificial intelligence (AI) to recruitment process since it speeds up the hiring process at lower costs while enhancing the efficiency and fairness at the same time (Sylva & Mol, 2009; Upadhyay & Khandelwal, 2018; Johnson et al., 2020). However, while AI is believed to provide cognitive insights that facilitate decision making (Duan et al., 2019; Edwards et al., 2000), the exploration of AI-based tools assisting IDM process is quite limited and the focus of existent research on the application of AI into interview procedure lies in three main aspects, including decision aids in identifying and screening applicants at the pre-interview stage (Nikolaou, 2014; Garg et al., 2018; Cowgill, 2018), chatbots that substitute for human interviewers to engage with and collect information from applicants (Xiao et al., 2020; Han et al., 2021; Majumder & Mondal, 2021), and analysing tools to investigate the features of applicants' interview performance, like facial expression, voice tone and language proficiency (Naim et al., 2016; Kharkovyna, 2018). In contrast, it is not until the recent years that the possibility of embedding AI into IDM process starts to be uncovered (see for example, Yakkundi et al., 2019; Siswanto et al., 2022), which may at least partially due to the deficiency of knowledge regarding how interviewers' decisions are progressively formed. Hence, the investigation into ICP from which the patterns of IDM process can be learned and modelled can facilitate the development of AI techniques that facilitate IDM.

On the other hand, in spite of suggested superiority of particular interview settings (e.g., consistent evaluation criteria, past behaviour questions, behavioural-anchored rating scale) concerning the enhancement of decision quality, it is unclear whether interviewers carry out the interview process as expected due to a lack of information of their performance during decision making (Latham & Saari, 1984; Di Milia & Gorodecki, 1997). The rich findings of ICP in this research, such as interviewers' eviden-

tial focus, judging behaviour and associations between the contextual factors and IDM at a cognitive level, can be utilised to monitor interviewers' decision-making process so that corresponding opportunities of improvement can be recognised. In addition, the observations of IDM process can also provide guidance for both the improvement of interview settings and the design of training courses for interviewers.

5.3 Limitations and Future Research

Apart from the rich findings and contributions, two limitations need to be noticed in this research. One is the small sample size due to the nature of TAM. Because of the particular focus on individual performance and the difficulty and intensity of protocol analysis, TAM-based studies aiming to qualitatively explore a subjects' cognitive process usually involve a sample size. While the number of interviewers involved in this research (i.e., 29) should be sufficient to seek for conceptual knowledge of ICP, it may constrain the generalisation of the statistical findings concerning the associations between the contextual factors and ICP. The other potential limitation lies in the fixed interview scenario adopted in the experiment and the single source of data collected and investigated. In particular, a single source of data (i.e., verbal protocols) of ICP is considered, which is collected from a fixed interview condition in terms of the job description, questions and responses, as well as the evaluation criteria provided. Hence, it is desirable if the findings can be tested in different contexts or double checked through other methods (e.g., survey, behavioural observations).

Future research is expected from various perspectives. First, similar empirical studies can be conducted in different interview scenarios and with a larger sample size involved so that both the conceptual and statistical findings in this research can be tested. Secondly, while the impact of only four contextual factors on ICP are investigated in this research, a broad range of additional factors examined in IDM research can be brought into analysis. Some possible examples include interviewers' personality (Lazar et al., 2004; De Kock, 2015), Cognitive complexity (Schneier, 1977), cognitive

style (Hunt et al., 1989), and initial impressions of an applicant (Macan & Dipboye, 1990; Swider et al., 2016; Florea et al., 2019) as well as applicants' appearance (Cable & Judge, 1997; Tsai et al., 2012; Johnson et al., 2014), accents (Purkiss et al., 2006; Deprez-Sims & Morris, 2013) and impression management tactics (Chen et al., 2010; Roulin et al., 2015; Kleinmann et al., 2016). Thirdly, research on various topics of IDM can be accelerated based on the knowledge of ICP, for example, to what extent the ICP during decision making can be structured by increasing the structure of interview settings and what specific clues is considered by interviewers when assessing applicants for a particular job position. Besides, with the cognitive processes of expert interviewers uncovered especially when interviewing applicants for certain job positions, their decision paths can be modelled to promote the development of AI-based decision aids to assist with (at least part of) IDM for the target job position.

References

Adkins, C. L., Russell, C. J., & Werbel, J. D. (1994). Judgments of fit in the selection process: The role of work value congruence. *Personnel Psychology*, *47*(3), 605–623.

Agerström, J., & Rooth, D.-O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*, *96*(4), 790.

Alliger, G. M., Feinzig, S. L., & Janak, E. A. (1993). Fuzzy sets and personnel selection: Discussion and an application. *Journal of Occupational and Organizational Psychology*, *66*(2), 163–169.

Alonso, P., Moscoso, S., & Salgado, J. F. (2017). Structured behavioral interview as a legal guarantee for ensuring equal employment opportunities for women: A meta-analysis. *The European journal of psychology applied to legal context*, *9*(1), 15–23.

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. *English linguistics research*, *3*(1), 39–45.

Amaral, A. A., Powell, D. M., & Ho, J. L. (2019). Why does impression management positively influence interview ratings? the mediating role of competence and warmth. *International Journal of Selection and Assessment*, *27*(4), 315–327.

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

Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International journal of selection and assessment*, *18*(3), 291–304.

Anderson, N., & Shackleton, V. (1990). Decision making in the graduate selection interview: A field study. *Journal of Occupational Psychology*, *63*(1), 63–76.

Anderson, N. R. (1992). Eight decades of employment interview research: A retrospective meta-review and prospective commentary. *The European Work and Organizational Psychologist*, *2*(1), 1–32.

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

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research 1. *Personnel psychology*, *35*(2), 281–322.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.

Ayalew, D. T. (2020). Re-visiting the chinese 'land-mark cases' of employment discrimination in light of theories of employment discrimination and american adjudication practice. *Available at SSRN 3718881*.

Azuma, R., Daily, M., & Furmanski, C. (2006). A review of time critical decision making models and human cognitive processes. In *2006 ieee aerospace conference* (pp. 9–pp).

Bangerter, A., Corvalan, P., & Cavin, C. (2014). Storytelling in the selection interview? how applicants respond to past behavior questions. *Journal of Business and Psychology*, *29*, 593–604.

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. does the verbalisation method affect learning? *Metacognition and Learning*, *3*, 39–58.

Barber, A. E., & Roehling, M. V. (1993). Job postings and the decision to interview: A verbal protocol analysis. *Journal of applied psychology*, *78*(5), 845.

Barkaoui, K. (2010). Variability in esl essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, *7*(1), 54–74.

Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, *28*(1), 51–75.

Baron, R. A. (1983). " sweet smell of success"? the impact of pleasant artificial scents on evaluations of job applicants. *Journal of Applied Psychology*, *68*(4), 709.

Baron, R. A. (1986). Self-presentation in job interviews: When there can be "too much of a good thing" 1. *Journal of Applied Social Psychology*, *16*(1), 16–28.

Baron, R. A. (1987). Interviewer's moods and reactions to job applicants: The influence of affective states on applied social judgments 1. *Journal of applied social psychology*, *17*(10), 911–926.

Barr, S. H., & Hitt, M. A. (1986). A comparison of selection decision models in manager versus student samples. *Personnel Psychology*, *39*(3), 599–617.

Barrick, M. R., Dustin, S. L., Giluk, T. L., Stewart, G. L., Shaffer, J. A., & Swider, B. W. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology*, *85*(2), 330–352.

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, *53*(4), 925–951.

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, *94*(6), 1394.

Basch, J. M., Melchers, K. G., Kurz, A., Krieger, M., & Miller, L. (2021). It takes more than a good camera: which factors contribute to differences between face-to-face interviews and videoconference interviews regarding performance ratings and interviewee perceptions? *Journal of Business and Psychology*, *36*, 921–940.

Baskett, G. D. (1973). Interview decisions as determined by competency and attitude similarity. *Journal of Applied Psychology*, *57*(3), 343.

Bawden, D., & Robinson, L. (2020). Information overload: An overview.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & management*, *39*(2), 151–163.

Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 205.

Bernardini, S. (2001). Think-aloud protocols in translation research: Achievements, limits, future prospects. *Target. International Journal of Translation Studies*, *13*(2), 241–263.

Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of consumer research*, *7*(3), 234–248.

Biesanz, J. C., Neuberg, S. L., Judice, T. N., & Smith, D. M. (1999). When interviewers desire accurate impressions: The effects of notetaking on the influence of expectations 1. *Journal of Applied Social Psychology*, *29*(12), 2529–2549.

Binning, J. F., Goldstein, M. A., Garcia, M. F., & Scattaregia, J. H. (1988). Effects of preinterview impressions on questioning strategies in same-and opposite-sex employment interviews. *Journal of applied Psychology*, *73*(1), 30.

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

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE transactions on professional communication*, *43*(3), 261–278.

Bragger, J. D., Kutcher, E., Morgan, J., & Firth, P. (2002). The effects of the structured interview on reducing biases against pregnant job applicants. *Sex Roles*, *46*, 215–226.

Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters. *Library & information science research*, *22*(4), 371–392.

Bryce, C., Dowling, M., & Lucey, B. (2020). The journal quality perception gap. *Research Policy*, *49*(5), 103957.

Buckley, M. R., Christine Norris, A., & Wiese, D. S. (2000). A brief history of the selection interview: May the next 100 years be more fruitful. *Journal of Management History*, *6*(3), 113–126.

Buijsrogge, A., Derous, E., & Duyck, W. (2016). Often biased but rarely in doubt: How initial reactions to stigmatized applicants affect interviewer confidence. *Human Performance*, *29*(4), 275–290.

Buijsrogge, A., Duyck, W., & Derous, E. (2021). Initial impression formation during the job interview: Anchors that drive biased decision-making against stigmatized applicants. *European Journal of Work and Organizational Psychology*, *30*(2), 305–318.

Burnett, J. R., Fan, C., Motowidlo, S. J., & Degroot, T. (1998). Interview notes and validity. *Personnel Psychology*, *51*(2), 375–396.

Burnett, J. R., & Motowidlo, S. J. (1998). Relations between different sources of information in the structured selection interview. *Personnel Psychology*, *51*(4), 963–983.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person–organization fit and organizational selection decisions. *Journal of Applied psychology*, *82*(4), 546.

Cambon, L., & Steiner, D. D. (2015). When rating format induces different rating processes: The effects of descriptive and evaluative rating modes on discriminability and accuracy. *Journal of Business and Psychology*, *30*, 795–812.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel psychology*, *50*(3), 655–702.

Carnes, A. M., Knotts, K. G., Munyon, T. P., Heames, J. T., & Houghton, J. D. (2019). Think fast: The role of thin slices of behavior in employee selection decisions. *International Journal of Selection and Assessment*, *27*(4), 357–370.

Cesare, S. J., Dalessio, A., & Tannenbaum, R. J. (1988). Contrast effects for black, white, male, and female interviewees 1. *Journal of Applied Social Psychology*, *18*(15), 1261–1273.

Chan, C.-S. (1990). Cognitive processes in architectural design problem solving. *Design Studies*, *11*(2), 60–80.

Chapman, D. S., & Rowe, P. M. (2001). The impact of videoconference technology, interview structure, and interviewer gender on interviewer evaluations in the employment interview: A field experiment. *Journal of Occupational and Organizational Psychology*, *74*(3), 279–298.

Chapman, D. S., & Webster, J. (2001). Rater correction processes in applicant selection using videoconference technology: The role of attributions. *Journal of Applied Social Psychology*.

Chapman, D. S., & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology*, *58*(3), 673–702.

Chapman, G. B., & Elstein, A. S. (2000). Cognitive processes and biases in medical decision making.

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, *12*(2).

Chauhan, R. S. (2022). Unstructured interviews: are they really all that bad? *Human Resource Development International*, *25*(4), 474–487.

Chen, C.-C., & Lin, M.-M. (2014). The effect of applicant impression management tactics on hiring recommendations: Cognitive and affective processes. *Applied Psychology*, *63*(4), 698–724.

Chen, C.-C., Wen-Fen Yang, I., & Lin, W.-C. (2010). Applicant impression management in job interview: The moderating role of interviewer affectivity. *Journal of Occupational and Organizational Psychology*, *83*(3), 739–757.

Chen, Y.-C., Tsai, W.-C., & Hu, C. (2008). The influences of interviewer-related and situational factors on interviewer reactions to high structured job interviews. *The International Journal of Human Resource Management*, *19*(6), 1056–1071.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, *6*(3), 271–315.

Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439–477.

Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, *32*(2), 301–341.

Cole, M. S., Feild, H. S., & Giles, W. F. (2003). Using recruiter assessments of applicants' resume content to predict applicant mental ability and big five personality dimensions. *International Journal of Selection and Assessment*, *11*(1), 78–88.

Connelly, L. M. (2016). Trustworthiness in qualitative research. *Medsurg nursing*, *25*(6), 435.

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

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

Cooke, L. (2010). Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication*, *53*(3), 202–215.

Cooper-Martin, E. (1994). Measures of cognitive effort. *Marketing Letters*, *5*, 43–56.

Corey, S. M. (1933). The interview in teacher selection. *The Journal of Educational Research*, *26*(7), 525–531.

Corner, P. D., Kinicki, A. J., & Keats, B. W. (1994). Integrating organizational and individual information processing perspectives on choice. *Organization Science*, *5*(3), 294–308.

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology*, *37*(1), 45–54.

Cowgill, B. (2018). Bias and productivity in humans and algorithms: Theory and evidence from resume screening. *Columbia Business School, Columbia University*, *29*.

Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (1999). Vicarious learning from dialogue and discourse: A controlled comparison. *Instructional science*, *27*, 431–458.

Crissy, W., & Regan, J. J. (1951). Halo in the employment interview. *Journal of Applied Psychology*, *35*(5), 338.

Crozier, R., & Ranyard, R. (2002). Cognitive process models and explanations of decision making. In *Decision making* (pp. 19–34). Routledge.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*, *5*(5), 241–241.

Culbertson, S. S., Weyhrauch, W. S., & Huffcutt, A. I. (2017). A tale of two formats: Direct comparison of matching situational and behavior description interview questions. *Human Resource Management Review*, *27*(1), 167–177.

Culbertson, S. S., Weyhrauch, W. S., & Waples, C. J. (2016). Behavioral cues as indicators of deception in structured employment interviews. *International Journal of Selection and Assessment*, *24*(2), 119–131.

Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating esl/efl writing tasks: A descriptive framework. *The Modern Language Journal*, *86*(1), 67–96.

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of management review*, *32*(1), 33–54.

Davison, G. C., Vogel, R. S., & Coffman, S. G. (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *Journal of consulting and clinical psychology*, *65*(6), 950.

Day, A. L., & Carroll, S. A. (2003). Situational and patterned behavior description interviews: A comparison of their validity, correlates, and perceived fairness. *Human Performance*, *16*(1), 25–47.

Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes behaviorally anchored rating scales so perennial as an appraisal method? *International Journal of Business and Social Science*, *6*(2).

De Faria, J. R. P., & Yoder, C. Y. (1997). The effects of interviewers' attributional and reasoning styles on hiring decisions. *Journal of Applied Social Psychology*, *27*(13), 1195–1208.

Deffner, G. (1990). Verbal protocols as a research tool in human factors: Symposium abstract. In *Proceedings of the human factors society annual meeting* (Vol. 34, pp. 1263–1264).

DeGroot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of business and psychology*, *24*, 179–192.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986.

De Kock, F. S., Lievens, F., & Born, M. P. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance*, *28*(3), 199–221.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception 1. *Journal of Applied Social Psychology*, *16*(3), 249–267.

Deprez-Sims, A.-S., & Morris, S. B. (2013). The effect of non-native accents on the evaluation of applicants during an employment interview: The development of a path model. *International Journal of Selection and Assessment*, *21*(4), 355–367.

Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review*, *26*(2), 90–111.

Dhar, R., & Sherman, S. J. (1996). The effect of common and unique features in consumer choice. *Journal of Consumer Research*, *23*(3), 193–203.

Dhar, R., & Simonson, I. (1992). The effect of the focus of comparison on consumer preferences. *Journal of Marketing Research*, *29*(4), 430–440.

Di Milia, L., & Gorodecki, M. (1997). Some factors explaining the reliability of a structured interview system at a work site. *International Journal of Selection and Assessment*, *5*(4), 193–199.

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

Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. South-Western Pub.

Dipboye, R. L. (2017a). The selection/recruitment interview: Core processes and contexts. *The Blackwell handbook of personnel selection*, 119–142.

Dipboye, R. L. (2017b). The selection/recruitment interview: Core processes and contexts. *The Blackwell handbook of personnel selection*, 119–142.

Dipboye, R. L., Fromkin, H. L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology*, *60*(1), 39.

Dipboye, R. L., & Jackson, S. L. (1999). Interviewer experience and expertise effects. *The employment interview handbook*, 259–278.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of applied psychology*, *71*(1), 9.

Dougherty, T. W., Turban, D. B., & Callender, J. C. (1994). Confirming first impressions in the employment interview: A field study of interviewer behavior. *Journal of applied psychology*, *79*(5), 659.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. *International journal of information management*, *48*, 63–71.

Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological monographs*, *58*(5), i.

Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do i use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531.

Edwards, J. S., Duan, Y., & Robins, P. C. (2000). An analysis of expert systems for business decision making at different levels and in different roles. *European Journal of Information Systems*, *9*(1), 36–46.

Ellis, A. P., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of applied psychology*, *87*(6), 1200.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. *The Cambridge handbook of expertise and expert performance*, 223–241.

Ericsson, K. A., & Crutcher, R. J. (1991). Introspection and verbal reports on cognitive processes—two approaches to the study of thinking: A response to howe.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. the MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186.

Eveland Jr, W. P., & Dunwoody, S. (2000). Examining information processing on the world wide web using think aloud protocols. *Media Psychology*, *2*(3), 219–244.

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

Fatt, J. P. T. (2000). Attractiveness and outcomes of job interviews. *Management Research Review*, *23*(1), 11.

Fay, C. H., & Latham, G. P. (1982). Effects of training and rating scales on rating errors. *Personnel Psychology*, *35*(1), 105–116.

Feiler, A. R., & Powell, D. M. (2016). Behavioral expression of job interview anxiety. *Journal of Business and Psychology*, *31*, 155–171.

Ferris, G. R., Berkson, H. M., & Harris, M. M. (2002). The recruitment interview process: Persuasion and organization reputation promotion in competitive labor markets. *Human Resource Management Review*, *12*(3), 359–375.

Fiedler, K., & Beier, S. (2014). Affect and cognitive processes in educational contexts. In *International handbook of emotions in education* (pp. 46–65). Routledge.

Florea, L., Valcea, S., Hamdani, M. R., & Dougherty, T. W. (2018). From first impressions to selection decisions: The role of dispositional cognitive motivations in the employment interview. *Personnel review*, *48*(1), 249–272.

Fontana, A., & Frey, J. H. (2005). The interview. *The Sage handbook of qualitative research*, *3*, 695–727.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative health research*, *3*(4), 430–441.

Forsberg, E., Ziegert, K., Hult, H., & Fors, U. (2014). Clinical reasoning in nursing, a think-aloud study using virtual patients–a base for an innovative assessment. *Nurse Education Today*, *34*(4), 538–542.

Fox, S., & Spector, P. E. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just 'g'. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *21*(2), 203–220.

Frank, L. L., & Hackman, J. R. (1975). Effects of interviewer-interviewee similarity on interviewer objectivity in college admissions interviews. *Journal of Applied Psychology*, *60*(3), 356.

French, K. E., Nevett, M. E., Spurgeon, J. H., Graham, K. C., Rink, J. E., & McPherson, S. L. (1996). Knowledge representation and problem solution in expert and novice youth baseball players. *Research quarterly for exercise and sport*, *67*(4), 386–395.

Frieder, R. E., Van Iddekinge, C. H., & Raymark, P. H. (2016). How quickly do interviewers reach decisions? an examination of interviewers' decision-making time across applicants. *Journal of Occupational and Organizational Psychology*, *89*(2), 223–248.

Funkesson, K. H., Anbäcken, E.-M., & Ek, A.-C. (2007). Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. a think-aloud study. *International journal of nursing studies*, *44*(7), 1109–1119.

Gallois, C., Callan, V. J., & PALMER, J.-A. M. (1992). The influence of applicant communication style and interviewer characteristics on hiring decisions 1. *Journal of Applied Social Psychology*, *22*(13), 1041–1060.

García, M. F., Posthuma, R. A., & Colella, A. (2008). Fit perceptions in the employment interview: The role of similarity, liking, and expectations. *Journal of occupational and organizational psychology*, *81*(2), 173–189.

Garg, V., Srivastav, S., & Gupta, A. (2018). Application of artificial intelligence for sustaining green human resource management. In *2018 international conference on automation and computational engineering (icace)* (pp. 113–116).

Ge, X., Chen, C.-H., & Davis, K. A. (2005). Scaffolding novice instructional designers' problem-solving processes using question prompts in a web-based learning environment. *Journal of educational computing research*, *33*(2), 219–248.

Gero, J. S., & Tang, H.-H. (2001). The differences between retrospective and concurrent protocols in revealing the process-oriented aspects of the design process. *Design studies*, *22*(3), 283–295.

Gifford, R., Ng, C. F., & Wilkinson, M. (1985). Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Journal of applied psychology*, *70*(4), 729.

Gilmore, D. C., & Ferris, G. R. (1989). The effects of applicant impression management tactics on interviewer judgments. *Journal of management*, *15*(4), 557–564.

Giordano, G., George, J., Marett, K., & Keane, B. (2011). Reviewers and the detection of deceptive information in recorded interviews. *Journal of Applied Social Psychology*, *41*(2), 252–269.

Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. *Academy of management review*, *11*(1), 191–207.

Graves, L. M. (1993). Sources of individual differences in interviewer effectiveness: A model and implications for future research. *Journal of Organizational Behavior*, *14*(4), 349–370.

Graves, L. M., & Karren, R. J. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology*, *45*(2), 313–340.

Graves, L. M., & Powell, G. N. (1995). The effect of sex similarity on recruiters'evaluations of actual applicants: A test of the similarity-attraction paradigm. *Personnel Psychology*, *48*(1), 85–98.

Graves, L. M., & Powell, G. N. (1996). Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational and Organizational Psychology*, *69*(3), 243–261.

Green, P. C., Alter, P., & Carr, A. F. (1993). Development of standard anchors for scoring generic past-behaviour questions in structured interviews. *International Journal of Selection and Assessment*, *1*(4), 203–212.

Guba, E. G. (1981). Eric/ectj annual review paper: Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational communication and technology*, 75–91.

Hackett, R. D., Lapierre, L. M., & Gardiner, H. P. (2004). A review of canadian human rights cases involving the employment interview. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *21*(3), 215–228.

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), 443.

Han, T. (2017). Scores assigned by inexpert efl raters to different quality efl compositions, and the raters' decision-making behaviors. *International Journal of Progressive Education*, *13*(1), 136–152.

Han, X., Zhou, M., Turner, M. J., & Yeh, T. (2021). Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging. In *Proceedings of the 2021 chi conference on human factors in computing systems* (pp. 1–15).

Harel, G. H., Arditi-Vogel, A., & Janz, T. (2003). Comparing the validity and utility of behavior description interview versus assessment center ratings. *Journal of Managerial Psychology*, *18*(2), 94–104.

Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology*, *42*(4), 691–726.

Harte, J. M., Westenberg, M. R., & van Someren, M. (1994). Process models of decision making. *Acta Psychologica*, *87*(2-3), 95–120.

Harvey, N. (1997). Confidence in judgment. *Trends in cognitive sciences*, *1*(2), 78–82.

Hayes, T. L., & Macan, T. H. (1997). Comparison of the factors influencing interviewer hiring decisions for applicants with and those without disabilities. *Journal of Business and Psychology*, *11*, 357–371.

Hebl, M. R., & Kleck, R. E. (2002). Acknowledging one's stigma in the interview setting: Effective strategy or liability? 1. *Journal of Applied Social Psychology*, *32*(2), 223–249.

Hebl, M. R., & Skorinko, J. L. (2005). Acknowledging one's physical disability in the interview: Does "when" make a difference? *Journal of Applied Social Psychology*, *35*(12), 2477–2492.

Heerkens, H., & Van Der Heijden, B. (2005). On a tool for analysing cognitive processes using exploratory think-aloud experiments. *International journal of human resources development and management*, *5*(3), 240–283.

Heimann, A. L., Ingold, P. V., Debus, M. E., & Kleinmann, M. (2021). Who will go the extra mile? selecting organizational citizens with a personality-based structured job interview. *Journal of Business and Psychology*, *36*(6), 985–1007.

Heintze, C., Metz, U., Hahn, D., Niewöhner, J., Schwantes, U., Wiesner, J., & Braun, V. (2010). Counseling overweight in primary care: an analysis of patient–physician encounters. *Patient education and counseling*, *80*(1), 71–75.

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

Herriot, P., & Rothwell, C. (1983). Expectations and impressions in the graduate selection interview. *Journal of Occupational Psychology*, *56*(4), 303–314.

Hess, K. P. (2013). Investigation of nonverbal discrimination against women in simulated initial job interviews. *Journal of Applied Social Psychology*, *43*(3), 544–555.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, *1*(3), 333–342.

Highhouse, S., & Bottrill, K. V. (1995). The influence of social (mis) information on memory for behavior in an employment interview. *Organizational Behavior and Human Decision Processes*, *62*(2), 220–229.

Highhouse, S., Nye, C. D., & Zhang, D. C. (2019). Dark motives and elective use of brainteaser interview questions. *Applied Psychology*, *68*(2), 311–340.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1–55.

Hollingworth, H. (1922). The personal interview.

Hollmann, T. D. (1972). Employment interviewers' errors in processing positive and negative information. *Journal of Applied Psychology*, *56*(2), 130.

Howard, J. L., & Ferris, G. R. (1996). The employment interview context: Social and situational influences on interviewer decisions 1. *Journal of applied social psychology*, *26*(2), 112–136.

Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, *19*(1), 62–81.

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, *21*(3), 264–276.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*(3), 619–644.

Huffcutt, A. I., & Youngcourt, S. S. (2007). Employment interviews. *Applied measurement: Industrial psychology in human resource management*, 181–199.

Hunt, R. G., Krzystofiak, F. J., Meindl, J. R., & Yousry, A. M. (1989). Cognitive style and decision making. *Organizational behavior and human decision processes*, *44*(3), 436–453.

Jaaskelainen, R. (2000). Focus on methodology in think-aloud studies on translating. *Benjamins Translation Library*, *37*, 71–82.

Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel psychology*, *33*(3), 595–640.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, *67*(5), 577.

Jaspers, M. W., Steen, T., Van Den Bos, C., & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International journal of medical informatics*, *73*(11-12), 781–795.

Johns, G. (1975). Effects of information order and frequency of applicant evaluation upon linear information-processing competence of interviewers. *Journal of Applied Psychology*, *60*(4), 427.

Johnson, K. E. (1992). Cognitive strategies and second language writers: A re-evaluation of sentence combining. *Journal of Second Language Writing*, *1*(1), 61–75.

Johnson, R. D., Stone, D. L., & Lukaszewski, K. M. (2020). The benefits of ehrm and ai for talent acquisition. *Journal of Tourism Futures*, *7*(1), 40–52.

Johnson, S. K., Sitzmann, T., & Nguyen, A. T. (2014). Don't hate me because i'm beautiful: Acknowledging appearance mitigates the "beauty is beastly" effect. *Organizational Behavior and Human Decision Processes*, *125*(2), 184–192.

Jones, S. L., & Kelly, R. (2018). Dealing with information overload in multifaceted personal informatics systems. *Human–Computer Interaction*, *33*(1), 1–48.

Joseph, G.-M., & Patel, V. L. (1990). Domain knowledge and hypothesis genenation in diagnostic reasoning. *Medical Decision Making*, *10*(1), 31–44.

Judge, T. A., Cable, D. M., & Higgins, C. A. (2000). The employment interview: A review of recent research and recommendations for future research. *Human resource management review*, *10*(4), 383–406.

Karsenty, L. (2001). Adapting verbal protocol methods to investigate speech systems use. *Applied Ergonomics*, *32*(1), 15–22.

Katalin, E. (2000). Please, keep talking": The 'think-aloud'method in second language reading research. *NovELTy VII*, *3*.

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27–44.

Keenan, A. (1977). Some relationships between interviewers' personal feelings about candidates and their general evaluation of them. *Journal of Occupational Psychology*, *50*(4), 275–283.

Keenan, A., & Wedderburn, A. (1980). Putting the boot on the other foot: Candidates' descriptions of interviewers. *Journal of Occupational Psychology*, *53*(1), 81–89.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of consumer research*, *14*(2), 200–213.

Kelley, T. R. (2008). Cognitive processes of students participating in engineering-focused design instruction. *Journal of Technology Education*, *19*(2), 50–64.

Kennedy, R. B. (1994). The employment interview. *Journal of Employment Counseling*.

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (bars) and other rating formats 1. *Personnel Psychology*, *34*(2), 263–289.

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

Kith, S., Cassidy, S. A., & Powell, D. M. (2022). The structured interview's resistance to gender discrimination under cognitive load. *International Journal of Selection and Assessment*, *30*(2), 281–301.

Kivetz, R., & Simonson, I. (2000). The effects of incomplete information on consumer choice. *Journal of marketing research*, *37*(4), 427–448.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational behavior and human decision processes*, *79*(3), 216–247.

Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance*, *21*(2), 107–137.

Klehe, U.-C., & Latham, G. (2006). What would you do—really or ideally? constructs underlying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human performance*, *19*(4), 357–382.

Kopelman, M. D. (1975). The contrast effect in the selection interview. *British Journal of Educational Psychology*, *45*(3), 333–336.

Kossek, E. E., & Pichler, S. (2006). Eeo and the management of diversity. *Handbook of human resource management*, 251–272.

Kristof-Brown, A., Barrick, M. R., & Franke, M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management*, *28*(1), 27–46.

Kristof-Brown, A. L. (2000). Perceived applicant fit: Distinguishing between recruiters'perceptions of person-job and person-organization fit. *Personnel psychology*, *53*(3), 643–671.

Kroll, E., & Ziegler, M. (2016). Discrimination due to ethnicity and gender: How susceptible are video-based job interviews? *International Journal of Selection and Assessment*, *24*(2), 161–171.

Kuipers, B., & Kassirer, J. P. (1984). Causal reasoning in medicine: analysis of a protocol. *Cognitive Science*, *8*(4), 363–385.

Kussmaul, P., & Tirkkonen-Condit, S. (1995). Think-aloud protocol analysis in translation studies. *TTR: traduction, terminologie, rédaction*, *8*(1), 177–199.

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American journal of psychology*, *113*(3), 387–404.

Landy, F. J., & Bates, F. (1973). Another look at contrast effects in the employment interview. *Journal of Applied Psychology*, *58*(1), 141.

Latham, G. P., & Saari, L. M. (1984). Do people do what they say? further studies on the situational interview. *Journal of applied Psychology*, *69*(4), 569.

Lazar, A., Kravetz, S., & Zinger, A. (2004). Moderating effects of rater personality on the relation between candidate self-monitoring and selection interview ratings. *International Journal of Selection and Assessment*, *12*(4), 321–326.

Leonard, N. H., Scholl, R. W., & Kowalski, K. B. (1999). Information processing style and decision making. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *20*(3), 407–420.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: development and validation of an interview faking behavior scale. *Journal of applied psychology*, *92*(6), 1638.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293.

Levine, S. P., & Feldman, R. S. (2002). Women and men's nonverbal behavior and self-monitoring in a job interview setting. *Applied HRM Research*, *7*(1), 1–14.

Lewis, C. (1980). Investigating the employment interview: A consideration of counselling skills. *Journal of Occupational Psychology*, *53*(2), 111–116.

Lewis, C. (1982). *Using the" thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center Yorktown Heights, NY.

Li, D. (2004). Trustworthiness of think-aloud protocols in the study of translation processes. *International Journal of Applied Linguistics*, *14*(3), 301–313.

Li, H., & He, L. (2015). A comparison of efl raters' essay-rating processes across two types of rating scales. *Language Assessment Quarterly*, *12*(2), 178–212.

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *25*(1), 29–46.

Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, *88*(3), 476.

Lievens, F., Highhouse, S., & De Corte, W. (2005). The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment. *Journal of Occupational and Organizational Psychology*, *78*(3), 453–470.

Lloyd, P., Lawson, B., & Scott, P. (1995). Can concurrent verbalization reveal design cognition? *Design Studies*, *16*(2), 237–259.

Lu, J. (2014). Regulating employment discrimination in china: a discussion from the socio-legal perspective. *Mich. St. Int'l L. Rev.*, *23*, 437.

Lubbe, D., & Nitsche, A. (2019). Reducing assimilation and contrast effects on selection interview ratings using behaviorally anchored rating scales. *International Journal of Selection and Assessment*, *27*(1), 43–53.

Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, *19*(3), 246–276.

Lundgrén-Laine, H., & Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative health research*, *20*(4), 565–575.

Lunenburg, F. C. (2010). The decision making process. In *National forum of educational administration & supervision journal* (Vol. 27).

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, *19*(3), 203–218.

Macan, T. H., & Dipboye, R. L. (1988). The effects of interviewers' initial impressions on information gathering. *Organizational Behavior and Human Decision Processes*, *42*(3), 364–387.

Macan, T. H., & Dipboye, R. L. (1990). The relationship of interviewers' preinterview impressions to selection and recruitment outcomes. *Personnel Psychology*, *43*(4), 745–768.

Macan, T. H., & Dipboye, R. L. (1994). The effects of the application on processing of information from the employment interview. 1. *Journal of Applied Social Psychology*, *24*(14), 1291–1314.

Madera, J. M., & Hebl, M. R. (2012). Discrimination against facially stigmatized applicants in interviews: an eye-tracking and face-to-face investigation. *Journal of Applied Psychology*, *97*(2), 317.

Majumder, S., & Mondal, A. (2021). Are chatbots really useful for human resource management? *International Journal of Speech Technology*, 1–9.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: police officers' ability to detect suspects' lies. *Journal of applied psychology*, *89*(1), 137.

Manroop, L., Boekhorst, J. A., & Harrison, J. A. (2013). The influence of cross-cultural differences on job interview selection decisions. *The International Journal of Human Resource Management*, *24*(18), 3512–3533.

Martin, S. L., & Klimoski, R. J. (1990). Use of verbal protocols to trace cognitions associated with self-and supervisor evaluations of performance. *Organizational Behavior and Human Decision Processes*, *46*(1), 135–154.

Maurer, S. D. (2002). A practitioner-based analysis of interviewer job expertise and scale format as contextual factors in situational interviews. *Personnel Psychology*, *55*(2), 307–327.

Mayfield, E. C. (1964). The selection interview—a re-evaluation of published research 1. *Personnel Psychology*, *17*(3), 239–260.

Mayfield, E. C., Brown, S. H., & Hamstra, B. W. (1980). Selection interviewing in the life insurance industry: An update of research and practice. *Personnel Psychology*, *33*(4), 725–739.

Mayfield, E. C., & Carlson, R. E. (1966). Selection interview decisions: First results from a long-term research project. *Personnel Psychology*, *19*(1), 41–53.

McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, *63*(2), 325–359.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, *79*(4), 599.

McDonald, T., & Hakel, M. D. (1985). Effects of applicant race, sex, suitability, and answers on interviewer's questioning strategy and ratings. *Personnel Psychology*, *38*(2), 321–334.

Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? a comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers'rating quality. *Personnel Psychology*, *64*(1), 53–87.

Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology*, *87*(2), 293.

Mihaylov, D. (2019). A way to accelerate the process of gathering information for decision-making. *International Journal on Information Technologies and Security*, *4*(11), 39–50.

Miller, A. L., & Dumford, A. D. (2016). Creative cognitive processes in higher education. *The Journal of Creative Behavior*, *50*(4), 282–293.

Moore, C., Lee, S. Y., Kim, K., & Cable, D. M. (2017). The advantage of being oneself: The role of applicant self-verification in organizational hiring decisions. *Journal of Applied Psychology*, *102*(11), 1493.

Morgeson, F. P., Reider, M. H., Campion, M. A., & Bull, R. A. (2008). Review of research on age discrimination in the employment interview. *Journal of Business and Psychology*, *22*, 223–232.

Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment*, *8*(4), 237–247.

Moscoso, S., & Salgado, J. F. (2001). Psychometric properties of a structured behavioral interview to hire private security personnel. *Journal of Business and Psychology*, *16*, 51–59.

Moskowitz, D. S. (1986). Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *Journal of personality*, *54*(1), 294–317.

Motowidlo, S. J., & Burnett, J. R. (1995). Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes*, *61*(3), 239–249.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of applied psychology*, *77*(5), 571.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological bulletin*, *126*(2), 247.

Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? the case of analogical problem-solving. *British Journal of Educational Psychology*, *68*(1), 15–24.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-hall Englewood Cliffs, NJ.

Nguyen, L. S., & Gatica-Perez, D. (2015). I would hire you in a minute: Thin slices of nonverbal behavior in job interviews. In *Proceedings of the 2015 acm on international conference on multimodal interaction* (pp. 51–58).

Nicholls, A. R., & Polman, R. C. (2008). Think aloud: Acute stress and coping strategies during golf performances. *Anxiety, Stress, & Coping*, *21*(3), 283–294.

Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads? reflections on the think-aloud technique. In *Proceedings of the second nordic conference on human-computer interaction* (pp. 101–110).

Nikolaou, I. (2011). Core processes and applicant reactions to the employment interview: An exploratory study in greece. *The International Journal of Human Resource Management*, *22*(10), 2185–2201.

Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment*, *22*(2), 179–189.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

Nolan, K. P., & Highhouse, S. (2014). Need for autonomy and resistance to standardized employee selection practices. *Human Performance*, *27*(4), 328–346.

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2381–2390).

Pälli, P., & Lehtinen, E. (2014). Making objectives common in performance appraisal interviews. *Language & Communication*, *39*, 92–108.

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

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science*, *5*(5), 241–248.

Payne, J. W., Braunstein, M. L., & Carroll, J. S. (1978). Exploring predecisional behavior: An alternative approach to decision research. *Organizational Behavior and Human Performance*, *22*(1), 17–44.

Perry, G. T., & Krippendorff, K. (2013). On the reliability of identifying design moves in protocol analysis. *Design Studies*, *34*(5), 612–635.

Perry, N. E., VandeKamp, K. O., Mercer, L. K., & Nordby, C. J. (2023). Investigating teacher—student interactions that foster self-regulated learning. , 5–15.

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

Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Mishra, P. (2011). Effects of organizational citizenship behaviors on selection decisions in employment interviews. *Journal of applied psychology*, *96*(2), 310.

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, *55*(1), 1–81.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel psychology*, *48*(2), 289–308.

Purkiss, S. L. S., Perrewé, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, *101*(2), 152–167.

Ramsay, S., Gallois, C., & Callan, V. J. (1997). Social rules and attributions in the personnel selection interview. *Journal of occupational and organizational psychology*, *70*(2), 189–203.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological review*, *120*(3), 697.

Rayner, M., Boaz, A., & Higginson, C. (2001). Consumer use of health-related endorsements on food labels in the united kingdom and australia. *Journal of Nutrition Education*, *33*(1), 24–30.

Raza, S. M., & Carpenter, B. N. (1987). A model of hiring decisions in real employment interviews. *Journal of applied psychology*, *72*(4), 596.

Reicks, M., Smith, C., Henry, H., Reimer, K., Atwell, J., & Thomas, R. (2003). Use of the think aloud method to examine fruit and vegetable purchasing behaviors among low-income african american women. *Journal of Nutrition Education and Behavior*, *35*(3), 154–160.

Reinhard, M.-A., Scharmach, M., & Müller, P. (2013). It's not what you are, it's what you know: Experience, beliefs, and the detection of deception in employment interviews. *Journal of Applied Social Psychology*, *43*(3), 467–479.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive science*, *21*(1), 1–29.

Rice, L., & Barth, J. M. (2016). Hiring decisions: The effect of evaluator gender and gender stereotype characteristics on the evaluation of job applicants. *Gender Issues*, *33*, 1–21.

Rice, L., & Barth, J. M. (2017). A tale of two gender roles: The effects of implicit and explicit gender role traditionalism and occupational stereotype on hiring decisions. *Gender Issues*, *34*, 86–102.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, *14*(4), 423–441.

Roberts, S., Celia; Campbell. (2006). *Talk on trial: Job interviews, language and ethnicity* (No. 344). Corporate Document Services.

Roberts, V. L., & Fels, D. I. (2006). Methods for inclusion: Employing think aloud protocols in software usability studies with individuals who are deaf. *International Journal of Human-Computer Studies*, *64*(6), 489–501.

Rost, K. (1989). The influence of patient participation on satisfaction and compliance. *The Diabetes Educator*, *15*(2), 139–143.

Rothstein, M., & Jackson, D. N. (1980a). Decision making in the employment interview: An experimental approach. *Journal of Applied Psychology*, *65*(3), 271.

Rothstein, M., & Jackson, D. N. (1980b). *Decision making in the employment interview: An experimental approach.* (Vol. 65) (No. 3). American Psychological Association.

Roulin, N., Bangerter, A., & Levashina, J. (2014). Interviewers' perceptions of impression management in employment interviews. *Journal of Managerial Psychology*, *29*(2), 141–163.

Roulin, N., Bangerter, A., & Levashina, J. (2015). Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? *Personnel Psychology*, *68*(2), 395–444.

Rowe, P. M. (1963). Individual differences in selection decisions. *Journal of Applied Psychology*, *47*(5), 304.

Rowe, P. M. (1967). Order effects in assessment decisions. *Journal of Applied Psychology*, *51*(2), 170.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & cognition*, *17*, 759–769.

Rynes, S., & Gerhart, B. (1990). Interviewer assessments of applicant "fit": An exploratory investigation. *Personnel psychology*, *43*(1), 13–35.

Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology*, *88*(5), 852.

Sackett, P. R. (1982). The interviewer as hypothesis tester: The effects of impressions of an applicant on interviewer questioning strategy. *Personnel Psychology*, *35*(4), 789–804.

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of management*, *36*(4), 941–973.

Sanchez-Burks, J., Blount, S., & Bartel, C. A. (2006). Fluidity and performance in intercultural workplace interactions: the role of behavioral mirroring and relational attunement. *Ross School of Business Paper*(1039).

Sandlund, E., Olin-Scheller, C., Nyroos, L., Jakobsen, L., & Nahnfeldt, C. (2011). The performance appraisal interview–an arena for the reinforcement of norms for employeeship. *Nordic journal of working life studies*, *1*(2), 59–75.

Sarkar, P., & Chakrabarti, A. (2013). A support for protocol analysis for design research. *Design Issues*, *29*(4), 70–81.

Savolainen, R. (2007). Filtering and withdrawing: strategies for coping with information overload in everyday contexts. *Journal of information Science*, *33*(5), 611–621.

Saxena, D., & Lamest, M. (2018). Information overload and coping strategies in the big data context: Evidence from the hospitality sector. *Journal of Information Science*, *44*(3), 287–297.

Schilpzand, P., Restubog, S. L. D., & Chen, Y. (2022). Sick on the day of the interview? effects of presenteeism on selection outcomes and the moderating role of raters' perspective-taking. *Journal of Organizational Behavior*.

Schleicher, D. J., Venkataramani, V., Morgeson, F. P., & Campion, M. A. (2006). So you didn't get the job... now what do you think? examining opportunity-to-perform fairness perceptions. *Personnel Psychology*, *59*(3), 559–590.

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International journal of selection and assessment*, *19*(2), 198–208.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel psychology*.

Schmitt, N., Klimoski, R. J., & Rowland, K. M. (1991). *Research methods in human resources management*. South-Western Pub.

Schneier, C. E. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology*, *62*(5), 541.

Schuh, A. J. (1978). Contrast effect in the interview. *Bulletin of the Psychonomic Society*, *11*(3), 195–196.

Schwab, D. P., Heneman III, H., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. In *Academy of management proceedings* (Vol. 1975, pp. 222–224).

Scott, W. D., Bingham, W., & Whipple, G. (1915). The scientific selection of salesmen. *Advertising and Selling*, *25*(5), 5–6.

Sears, G. J., & Rowe, P. M. (2003). A personality-based similar-to-me effect in the employment interview: Conscientiousness, affect-versus competence-mediated interpretations, and the role of job relevance. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *35*(1), 13.

Sepp, S., Howard, S. J., Tindall-Ford, S., Agostinho, S., & Paas, F. (2019). Cognitive load theory and human movement: Towards an integrated model of working memory. *Educational Psychology Review*, *31*, 293–317.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, *22*(2), 63–75.

Siddique, C. (2004). Job analysis: a strategic human resource management practice. *The International Journal of Human Resource Management*, *15*(1), 219–244.

Silvester, J., & Anderson, N. (2003). Technology and discourse: A comparison of face-to-face and telephone employment interviews. *International Journal of Selection and Assessment*, *11*(2-3), 206–214.

Simplicio, J. S. (2007). A closer look at the truth behind the hiring process: How colleges really hire. *Education*, *128*(2).

Sinclair, M., & Ashkanasy, N. M. (2005). Intuition: Myth or a decision-making tool? *Management learning*, *36*(3), 353–370.

Siswanto, J., Suakanto, S., Andriani, M., Hardiyanti, M., & Kusumasari, T. F. (2022). Interview bot development with natural language processing and machine learning. *International Journal of Technology*, *13*(2), 274–285.

Snow, A. J. (1924). An experiment in the validity of judging human ability. *Journal of Applied Psychology*, *8*(3), 339–346.

Speer, A. B., Christiansen, N. D., & Laginess, A. J. (2019). Social intelligence and interview accuracy: Individual differences in the ability to construct interviews and rate accurately. *International Journal of Selection and Assessment*, *27*(2), 104–128.

Speer, A. B., Tenbrink, A. P., Schwendeman, M. G., & Wegmeyer, L. J. (2020). Individual differences in effective interview design: Factors affecting question choice. *International Journal of Selection and Assessment*, *28*(3), 310–321.

Speer, A. B., Wegmeyer, L. J., & Delacruz, A. Y. (2022). Factors leading to interview question decisions: Introducing the model of interviewer question preferences. *International Journal of Selection and Assessment*, *30*(3), 392–410.

Springbett, B. (1958). Factors affecting the final decision in the employment interview. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *12*(1), 13.

Stevens, C. K. (1998). Antecedents of interview interactions, interviewers'ratings, and applicants'reactions. *Personnel Psychology*, *51*(1), 55–85.

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

Stewart, G. L., Dustin, S. L., Barrick, M. R., & Darnold, T. C. (2008). Exploring the handshake in employment interviews. *Journal of Applied Psychology*, *93*(5), 1139.

Stiegler, M. P., & Tung, A. (2014). Cognitive processes in anesthesiology decision making. *Anesthesiology*, *120*(1), 204–217.

Strawn, M. W., & Thorsteinson, T. J. (2015). Influence of response mode on order effects in the interview. *Human Performance*, *28*(3), 183–198.

Stumpf, S. A., & London, M. (1981). Management promotions: Individual and organizational factors influencing the decision process. *Academy of Management Review*, *6*(4), 539–549.

Suen, H. K., & Ary, D. (2014). *Analyzing quantitative behavioral observation data*. psychology press.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational behavior and human decision processes*, *65*(3), 252–267.

Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology*, *101*(5), 625.

Sydiaha, D. (1959). On the equivalence of clinical and statistical methods. *Journal of Applied Psychology*, *43*(6), 395.

Sydiaha, D. (1961). Bales' interaction process analysis of personnel selection interviews. *Journal of Applied Psychology*, *45*(6), 393.

Sydiaha, D. (1962). Interviewer consistency in the use of empathic models in personnel selection. *Journal of Applied Psychology*, *46*(5), 344.

Sylva, H., & Mol, S. T. (2009). E-recruitment: A study into applicant perceptions of an online application system. *International Journal of Selection and assessment*, *17*(3), 311–323.

Tamler, H. (1998). How (much) to intervene in a usability testing session. *Common ground*, *8*(3), 11–15.

Taylor, M. S., & Sniezek, J. A. (1984). The college recruitment interview: Topical content and applicant reactions. *Journal of Occupational Psychology*, *57*(2), 157–168.

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of occupational and organizational psychology*, *75*(3), 277–294.

Thomas, O., & Reimann, O. (2023). The bias blind spot among hr employees in hiring decisions. *German Journal of Human Resource Management*, *37*(1), 5–22.

Thompson, K. W., Sikora, D. M., Perrewé, P. L., & Ferris, G. R. (2015). Employment qualifications, person–job fit, underemployment attributions, and hiring recommendations: A three-study investigation. *International Journal of Selection and Assessment*, *23*(3), 247–262.

Topor, D. J., Colarelli, S. M., & Han, K. (2007). Influences of traits and assessment methods on human resource practitioners' evaluations of job applicants. *Journal of Business and Psychology*, *21*, 361–376.

Tordesillas, R. S., & Chaiken, S. (1999). Thinking too much or too little? the effects of introspection on the decision-making process. *Personality and Social Psychology Bulletin*, *25*(5), 625–631.

Tsai, W.-C., HsinHung Chen, F., Chen, H.-Y., & Tseng, K.-Y. (2016). When will interviewers be willing to use high-structured job interviews? the role of personality. *International Journal of Selection and Assessment*, *24*(1), 92–105.

Tsai, W.-C., Huang, T.-C., & Yu, H.-H. (2012). Investigating the unique predictability and boundary conditions of applicant physical attractiveness and non-verbal behaviours on interviewer evaluations in job interviews. *Journal of Occupational and Organizational Psychology*, *85*(1), 60–79.

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

Tullar, W. L. (1989). Relational control in the employment interview. *Journal of Applied Psychology*, *74*(6), 971.

Tullar, W. L., Mullins, T. W., & Caldwell, S. A. (1979). Effects of interview length and applicant quality on interview decision time. *Journal of Applied Psychology*, *64*(6), 669.

Turnley, W. H., & Bolino, M. C. (2001). Achieving desired images while avoiding undesired images: exploring the role of self-monitoring in impression management. *Journal of applied psychology*, *86*(2), 351.

Tziner, A., Joanis, C., & Murphy, K. R. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and ratee satisfaction. *Group & Organization Management*, *25*(2), 175–190.

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological bulletin*, *63*(2), 100.

Ungson, G. R., Braunstein, D. N., & Hall, P. D. (1981). Managerial information processing: A research review. *Administrative Science Quarterly*, 116–134.

Upadhyay, A. K., & Khandelwal, K. (2018). Applying artificial intelligence: implications for recruitment. *Strategic HR Review*, *17*(5), 255–258.

Valenzi, E., & Andrews, I. (1973). Individual differences in the decision process of employment interviewers. *Journal of Applied Psychology*, *58*(1), 49.

Van Dam, K. (2003). Trait perception in the employment interview: A five–factor model perspective. *International Journal of Selection and Assessment*, *11*(1), 43–55.

Van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, *87*(1), 176.

Van Gog, T., Paas, F., Van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, *11*(4), 237.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied psychology*, *90*(3), 536.

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 147–177.

Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress*, *11*, 29–41.

Varghese, L., Lindeman, M. I. H., & Finkelstein, L. (2018). Dodging the double bind: The role of warmth and competence on the relationship between interview communication styles and perceptions of women's hirability. *European Journal of Work and Organizational Psychology*, *27*(4), 418–429.

Vianen, N. E. V., & Willemsen, T. M. (1992). The employment interview: The role of sex stereotypes in the evaluation of male and female job applicants in the netherlands 1. *Journal of Applied Social Psychology*, *22*(6), 471–491.

Walker, J. T., Fenton, E., Salter, A., & Salandra, R. (2019). What influences business academics' use of the association of business schools (abs) list? evidence from a survey of uk academics. *British Journal of Management*, *30*(3), 730–747.

Wanberg, C. R., Ali, A. A., & Csillag, B. (2020). Job seeking: The process and experience of looking for a job. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*, 315–337.

Wang, Y., & Ruhe, G. (2007). The cognitive process of decision making. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *1*(2), 73–85.

Ward, P., Hodges, N. J., & Williams, A. M. (2004). Deliberate practice and expert performance: Defining the path to excellence. In *Skill acquisition in sport* (pp. 255–282). Routledge.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological review*, *20*(2), 158.

Watson, J. B. (1920). Is thinking merely the action of language mechanisms? *British Journal of Psychology*, *11*, 87–104.

Weichselbaumer, D. (2004). Is it sex or personality? the impact of sex stereotypes on discrimination in applicant selection. *Eastern Economic Journal*, *30*(2), 159–186.

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

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., ... Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and learning*, *4*, 63–85.

Wilhelmy, A., Kleinmann, M., König, C. J., Melchers, K. G., & Truxillo, D. M. (2016). How and why do interviewers try to make impressions on applicants? a qualitative study. *Journal of Applied Psychology*, *101*(3), 313.

Wilhelmy, A., Roulin, N., & Wingate, T. G. (2021). Does it take two to tango? examining how applicants and interviewers adapt their impression management to each other. *Journal of Business and Psychology*, *36*, 1053–1076.

Williams, E. N., & Morrow, S. L. (2009). Achieving trustworthiness in qualitative research: A pan-paradigmatic perspective. *Psychotherapy research*, *19*(4-5), 576–582.

Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology*, *82*(6), 900.

Wolfe, E. W., Kao, C.-W., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written Communication*, *15*(4), 465–492.

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64–77.

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of occupational psychology*, *62*(3), 191–199.

Xiao, Z., Zhou, M. X., Chen, W., Yang, H., & Chi, C. (2020). If i hear you correctly: Building and evaluating interview chatbots with active listening skills. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–14).

Xun, G., & Land, S. M. (2004). A conceptual framework for scaffolding iii-structured problem-solving processes using question prompts and peer interactions. *Educational technology research and development*, *52*(2), 5–22.

Yakkundi, S., Vanjare, A., Wavhal, V., & Patankar, S. (2019). Interactive interview chatbot. *International Research Journal of Engineering and Technology*, *6*(4), 2746–2748.

Young, A. M., & Kacmar, K. M. (1998). Abcs of the interview: The role of affective, behavioral, and cognitive responses by applicants in the employment interview. *International Journal of Selection and Assessment*, *6*(4), 211–221.

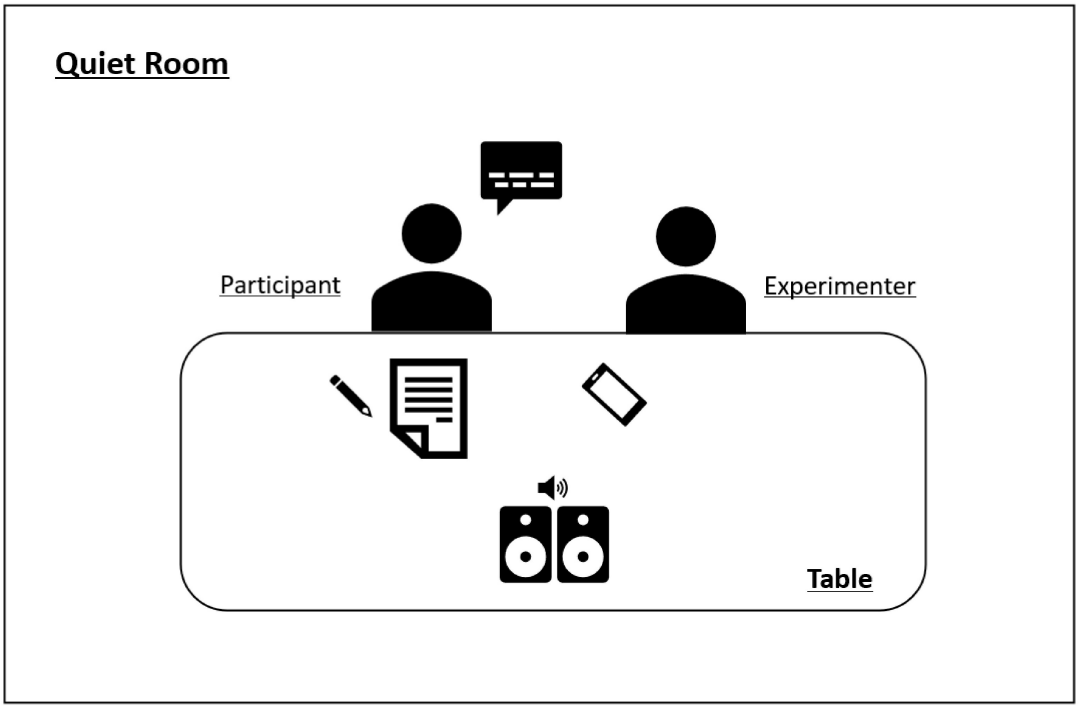
Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational behavior and human performance*, *18*(2), 269–294.

Zedeck, S., Tziner, A., & Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology*, *36*(2), 355–370.

Zysberg, L., & Nevo, B. (2004). "the smarts that counts?": Psychologists' decision-making in personnel selection. *Journal of Business and Psychology*, *19*, 117–124.

Appendix

A. The Environment of Data Collection



Audio interview records

Experimental materials

Recording equipment

B. Consent Form

UNIVERSITY OF STRATHCLYDE

CONSENT FORM FOR RESEARCH PARTICIPATION

Study Title

Developing a Think-Aloud-Based Procedure to Explore the Cognitive Process of Employment Interviewers

I am a PhD student at the Department of DMEM, University of Strathclyde. At my current research stage, I need your valuable inputs. This form contains important information about the research motivation, instructions about your participation, and the procedure of processing your inputs.

Overview

Participants are invited to take part in a scenario-based experiment simulating an employment interview.

The purpose of this experiment is to investigate how participants, acting as interviewers, process information during employment interview to make hiring decisions. At the end of the experiment, a more detailed explanation will be provided about the research hypotheses and implications.

Instructions

You will be asked to act as an interviewer and perform oral evaluation over two "candidates".

Duration: 60-90 mins

As one of my research requirements, audio-record will be made during the experiment. The record will only be used for research purpose and confidentiality is ensured. No sensitive and personal information will be gathered.

Your inputs may be quoted in any presentations or articles resulting from the experiment analysis. A pseudonym will be used to protect your identity, unless you specifically request that you be identified by your true name.

What are the possible risks or discomforts?

To the best of our knowledge, this experiment contains no more risk or harm than you would experience in everyday life.

However, you might feel unfamiliar with the experiment or the associated research activities. In such cases, we highly respect your feelings and preference to continue or discontinue the experiment.

What are the possible benefits?

This experiment is designed to learn more about problem solving and decision making under a job interview setting. The research results derived from this experiment may help explain how information processing takes place during a job interview. This may enhance interview design and process for interviewers as well as preparation for interviewees.

Data protection

Results of this experiment may be made publicly available through publications and presentations. The inputs you provide will be handled as confidentially as possible and individual names and other sensitive/personal information will not be collected/used.

Financial Information

Participation in this study will involve no cost to you. You will not be paid for participating in this study.

Your rights

Your participation in this study is voluntary. You can choose to or not to answer any questions during the experiment. At any time, you can choose to discontinue or pause or withdraw from this experiment. All of your inputs will be processed only with your consent.

Contact information

If you have any questions about this experiment, the research can be contacted through the following means.

Name: Shiyun Zhao (Renee)
Department: DMEM, University of Strathclyde, Glasgow
Mobile: (+44) 07762206028
Email: shiyun.zhao@strath.ac.uk

Consent

I have read this form and understood all procedures associated with this experiment. I have been given the opportunity to ask questions and my questions have been answered. If I have additional questions, I have been told whom to contact. I agree to participate in the experiment described in this form. A copy of this consent form will be provided if needed.

Participant's Name (printed):

Participant's Signature: Date:

C. Pre-Experiment Survey

1. Please indicate your age:

   ☐ 18-30 ☐ 31-40 ☐ 41-50 ☐ 51-60 ☐ 61 or above

2. Please indicate your gender:

   ☐ Male ☐ Female ☐ Prefer not to say

3. Have you ever worked as a job interviewer?

   ☐ Yes (turn to Q4) ☐ No (turn to Q6)

4. How long have you worked as a job interviewer?

5. How often do you conduct job interviews?

6. Have you ever attended a behavioural-based interview?

7. Please indicate the region of your interview experience:

   ☐ UK ☐ China ☐ Other: _____

D. Warm-Up Session Instruction

The following content was read by the experimenter to the participant:

You will soon receive a mathematical problem. Please work on the problem while concurrently speaking out aloud all the thoughts that come into your mind as if talking to yourself. The purpose of this task is to practice TAM, and your performance on the problem will not be judged.

The following content was presented on a piece of paper to the participant:

There is a family of four, including Father, Mother, Daughter and Son.
They are together 150 years old.
Father is twice as old as Daughter.
Mother is as old as Father.
Son is six years younger than Daughter.
How old is Son?

E. Mock Interview Task

Scenario Experiment: Mock Employment Interview

Background

ABC Co.,Ltd is a medium-sized enterprise that provides marketing programmes to both top brands and start-ups. The company is now looking for a Project Manager who is expected to collaborate with cross-functional teams to deliver high-quality results through well-managed, on-time and on-budget projects.

Main Responsibilities of the Project Manager

- Ensure the smooth delivering of each project within budget, within scope, on time and to the customer's satisfaction
- Compile the Statement of Work; ensure the scope of the project is documented and milestones and priorities are clearly understood and agreed by both the team and stakeholders
- Drive risk management activities and processes, including risk identification, analysis, planned response, monitoring, reporting and control, to ensure project risks and issues are dealt with promptly
- Plan and balance priorities whilst managing multiple projects and tasks or when faced with changes in the scope
- Monitor and report project progress to key stakeholders
- Responsible for the management of a multifaceted teams assigned to projects, identify and set clearly defined roles and responsibilities for the team members to ensure seamless communication and coordination
- Overall responsibility for effective communication processes, meetings and activities within the project
- Provide the social interaction needed across the whole organisation, to connect with people and facilitate the optimal flow of information

(continued)

Key Attributes of the Project Manager

- Analytical skills - can clearly and quickly analysis project data and produce updates and reports
- Adaptability - can manage multiple projects and is able to respond positively to change and lead others through change
- Good lateral thinking and problem-solving skills
- Effective leadership and collaborative team player
- Excellent written, verbal and presentation skills
- Strong time management skills
- Ability to manage pressure

After the CV screening and technical tests, **several** candidates have been shortlisted. The next step is to conduct a face-to-face job interview with each of them to determine if they possess certain **soft skills** to fit the position.

Please assume that you are one of the two hiring managers responsible for the interview. While the other one leads the conversation with the candidates, you are expected to:

(1) listen carefully to the conversation,

(2) take notes when necessary, and

(3) make comments on the skill qualification of the candidates.

For ease of experiment, we will only 'interview' **two** candidates and assess **four** dimensions of their soft skills (see *Rating Scale*) based on the important questions as follows:

Interview Questions

Q1: How did you deal with the situation where you had to handle multiple tasks or projects and the time was tight?

(continued)

Q2: Have you ever encountered oppositions from others to your ideas or to your work arrangements?
Q3: Tell us about a situation at work where something important did not go according to your plan.
Q4: Can you tell us something about the most stressful task or situation in your work experience?

Experiment Task

You are about to hear a conversation extracted from a job interview, please do the following:

- listen carefully to each conversation and take notes when necessary to as the record will only be played once *(All the notes and ratings are for your own use in future discussion and can be amended.)*
- after each Q&A, you are expected to verbally express all the thoughts that come to your mind during the process, including but not limited to raising prompting questions, analysing information, and making evaluations
- interactions with the facilitator will be minimised during the process; however, if necessary, the facilitator will intervene
- complete the rating scale and make the final decision (for future discussion only)
- there is no time limit for verbal report

F. Behavioural-Anchored Rating Scale

Planning & Organisation:

Attribute \ Scale	5 (excellent)	3 (average)	1 (unsatisfied)
Perceptiveness & Initiative	Anticipates needs and steps required to complete tasks and prepares for future assignments	Determines specific work tasks and available resources to ensure timely completion of core work assignments	Focuses exclusively on the task at the current moment, does not schedule beyond the immediacy
Work Scheduling	Consistently develops comprehensive and integrated project plans that help achieve the strategic needs of the organisation	Effectively prioritises tasks based on their importance and urgency	Struggle to prioritise actions, needs a great deal of guidance around what to do and when
Implementation & Monitoring	Excels in implementing project plans, effectively manages multiple projects to completion	Effectively implements project plans and monitors the progress of team members	Does not manage everyday tasks or multiple tasks well, does not effectively monitor own progress and the work activities of team members
Change & Risk Management	Anticipates the need for alternate course of actions; successfully implements contingencies in response to frequently changing demands; knows project status and risks at all times	Modifies plans in time to address changing priorities and needs when necessary	Has no effective measure to deal with changes in tasks and demands
Productivity Improvement	Consistently completes projects utilising minimal or optimal resources within or ahead of established timeframes	Delivers results on time and within budget; does not over-promise or under-deliver	Does not consistently deliver results on time or within budget

Notes	**Rating (1-5)**

Problem-solving:

Attribute \ Scale	5	3	1
Problem Perception & Identification	Has instinctive skill in recognising problem situations before they develop	Effectively recognises and defines problems associated with work	Struggle to recognise and define problems
Assessment of Impact	Is extremely skilled in identifying and evaluating the impact of problem situations	Aware of the possible impact caused by problem situations	Lack of awareness and abilities to evaluate problematic situations
Development of Solutions	Is highly creative in the development of problem-solving techniques, seeks out and attempts to solve the root causes of problems	Weights advantages and disadvantages of proposed solutions	Requires assistance in weighing advantages and disadvantages of potential solutions
Gathering & Analysing Info.	Possesses great skill in gathering and analysing multi-faceted data or information specific to problem situations	can obtain and analyse factual data or information in relation to problem situations	Possesses insufficient awareness and skills in gathering and analysing data or information specific to problem situations
Relevant Knowledge	Have a thorough understanding of all relevant procedures, principles and/or policies, and can effectively apply them to problem solving	Can understand and apply some of the relevant procedures, principles and/or policies to problem solving	Unable to correctly understand and apply any relevant procedures, principles and/or policies to problem solving

Notes	**Rating (1-5)**

(continued)

Leadership & Conflict Management:

Attribute \ Scale	5	3	1
Respect Individual Differences	Treats all people with dignity, respect, and fairness, inspires team members do so as well	Treats everyone with dignity, respect and fairness	Does not treat all people with the same level of respect, dignity or fairness
Handling Conflicts	Resolves conflict constructively, regards conflict as an opportunity for improvement rather than just a problem	Resolves conflicts constructively and professionally, seldom requires external assistance	Seldom attempts to resolve conflicts, let problems fester and escalate, often requires third party intervention
Open to Different Ideas	Actively solicits and encourages ideas from a wide variety of individuals, is always careful to ensure every side is heard and considered before reaching a conclusion	Listens to and carefully considers ideas from key people, ensures all sides are heard before reaching a conclusion	Does not show interest in listening to ideas from others, has difficulty respecting ideas when different from own, tends to reach conclusions before listening to all sides
Sharing Resources & Being Helpful	Anticipates and acts upon opportunities for helping others succeed, enthusiastically shares time, resources and knowledge with others	Feel obligated to spend time, resources and knowledge to help others and the team succeed	Misses opportunities to share time, resources and knowledge with others and provide necessary assistance
Promoting Teamwork	Visibly and proactively encourages teamwork, recognises contributions from all team members	Encourages teamwork, promotes respect among all team members	Does not value teamwork, neglects the importance of respect among all team members

Notes	**Rating (1-5)**

Stress Management:

Attribute \ Scale	5	3	1
Stressor Monitoring	Knows well about the trigger and consequence of pressure (i.e. stressors) over time	Can identify the trigger and consequence of pressure in time	Not aware of the trigger and consequence of the pressure
Reaction to Stress	Responses positively to pressure with practical actions, can be motivated by reasonable pressure to produce better work	Keeps calm and clear-minded when faced with pressure, manage to ensure the work progresses under pressure	Reacts to pressure in a negative way, allow pressure to impede or degrade the quality of work
Decompressing	Can independently deal with pressure in most cases, seek for help when necessary	Successfully overcome stressful situations with external assistance	Cannot deal with pressure properly and get overwhelmed
Aspiration for Improving Stress Management Skills	Regularly introspects the stressors, reactions and outcomes for the improvement of pressure management	Attempts to improve the ability of pressure management	Shows no aspiration to improve the ability of pressure management

Notes	**Rating (1-5)**

Overall Comments	**Accept**	**Accept with reservations**	**Do not accept**

G. Open Grid Rating Scale

Planning & Organisation:

Attribute \ Scale	5 (excellent)	3 (average)	1 (unsatisfied)
Perceptiveness & Initiative			
Work Scheduling			
Implementation & Monitoring			
Change & Risk Management			
Productivity Improvement			

Notes	Rating (1-5)

Problem-solving:

Attribute \ Scale	5	3	1
Problem Perception & Identification			
Assessment of Impact			
Development of Solutions			
Gathering & Analysing Info.			
Relevant Knowledge			

Notes	Rating (1-5)

(continued)

Leadership & Conflict Management:

Attribute \ Scale	5	3	1
Respect Individual Differences			
Handling Conflicts			
Open to Different Ideas			
Sharing Resources & Being Helpful			
Promoting Teamwork			

Notes	Rating (1-5)

Stress Management:

Attribute \ Scale	5	3	1
Stressor Monitoring			
Reaction to Stress			
Decompressing			
Aspiration for Improving Stress Management Skills			

Notes	Rating (1-5)

Overall Comments	Accept	Accept with reservations	Do not accept

H. Post-Experiment Survey

1. To what extent do you think the scenario experiment simulates a real interview?

   ☐ 1 (poorly simulated) ☐ 2 ☐ 3 (modest) ☐ 3 ☐ 5 (well simulated)

2. Do you think the length of the whole experiment is appropriate?

   ☐ 1 (inappropriate) ☐ 2 ☐ 3 (modest) ☐ 3 ☐ 5 (appropriate)

3. Do you think the task difficulty is appropriate? Is there anything you feel hard to understand during the process (e.g., language, candidates' responses)?

   ☐ 1 (inappropriate) ☐ 2 ☐ 3 (modest) ☐ 3 ☐ 5 (appropriate)

4. To what extent do you feel confident about your evaluations and decisions?

   ☐ 1 (unconfident) ☐ 2 ☐ 3 (modest) ☐ 3 ☐ 5 (confident)

5. Do you think TAM affects your decision making? If yes, how and to what extent do you think TAM has affected your decision making?

   ☐ -5 (strongly negative) ☐ -5 ☐ 0 (no impact) ☐ 3 ☐ 5 (strongly positive)

6. Do you have any further advice on this experiment?

I. Sample of the Raw Protocols

Okay. Um, yeah, I find his style, his talking style, a little bit difficult because he does talk as if he's reading. Um, so um, his personality is not coming through that much when he's giving his answers. So, I'm, I'm kind of not getting a sense of, of how he would be to work with. It was a very confusing start. I wasn't quite sure where he was going to go with his answer to this question. He talked about, um, people having different approaches and different understandings. Um, it was good that he said that, to argue his point and to deal with opposition, he would, um, conduct a test to prove a theory. And he then went on to give the specific example of where he did that. So, it took him a little while to get round to answering the question. But he did so in the end. Um, he, he went on to talk about not wanting to ask colleagues for help or advice, because it might be, um, an imposition on their work schedule, or it may have an impact on their performance appraisal. So, he will go directly to his manager, which, um, I felt moved away from the question, but also, um, showed a lack of ability to deal with things independently. To go to your manager for something like that would not be what I would expect people to do. So that can have raised a little bit of a red flag in terms of his ability to, um, show initiative and be self-motivated. The example he gave on the social app, um, I thought that was quite a good example.

(continued)

In the end, he came up with a solution to a particular problem. Um, he put forward his proposal. It wasn't accepted by his manager or members of his team. But when he tested the idea, he got a good response. And it's now a fundamental part of the app. So, I thought that was a good example. It would have been fantastic if you'd started with that. Um, it would have been less confusing in terms of trying to again, offer a hypothetical, um, situations. I think, um, with this candidate, I think I would be focusing on, before I ask the question, saying "please give a very specific example to back up your answer" so that I would got more of an idea of how he actually performed in a situation. Um, his tone is very distant as it's not very conversational. Um, so it's quite difficult to get sense of, um, how he communicates with people in the workplace, because it's quite stilted. It's quite unnatural. Um, but we'll see how he gets on with the next question.

J. Sample of Protocol Translation

也让我想到自己的第一份工作，	*It also reminds me of my first job.*
那个时候就是要做那种开学典礼，那些活动，然后也是要直播，	*I was responsible for organising an opening ceremony that would go live.*
然后也是设备上面会有很多问题，可能调试都是对的，但是到正式的时候就是会出问题，	*Everything was fine in the rehearsal, but you know, unexpected situations always happen.*
整个人的状态都很紧张，整个人都是很暴躁的。	*And I was really nervous at that time and could hardly control my temper.*
所以我觉得这个人面试，我觉得他还不是很像一个项目经理，就是不太像一个 leader 的那种角色，没有感觉到他很有做经理的能力。	*So, I think this candidate, according to his story, I don't think he has performed like a manager or a leader, I didn't quite get that.*
所以他之前那个题目中罗列出来的那些比较标准的流程，其实在他这感觉没有应用得特别的好。	*So, for those standard procedures he stated in the previous question, I don't think he has effectively applied them in real practice.*
在这么短短四个问题中他举了两个例子，每次举例子都是和其他公司合作，然后失败了。	*Throughout these four questions, he gave two examples where he cooperated with another company and the projects failed.*
然后失败的原因全都是因为对方，自己没有任何问题，自己公司没有任何问题，全都是对方的事。	*It seems that he tends to attribute the failure to the other party, that he and his company have done nothing wrong.*

K. Coding Form and Sample of Coded Protocols

No.	Segment	Performance	Response	Other
22	Yeah, I think this answer was less successful than the first one.		J6-J1	
23	He tried to give a specific example at the start when he was talking about the sort of opposition that he had come across.		I4	
24	He talked about scheduling and method and he said that some people had wanted a bit more flexibility,		I2	
25	but he didn't go into specific example.		I3	
26	I would like to ask for a specific example and say could you clarify what you mean by that and talk into a bit more detail.		I5	
27	It also raised the question about his own leadership style.	J1		
28	If people were asking for more flexibility,	I2		
29	was he being a little bit too dictatorial in terms of what you want people to do?	I4		
30	He then went into a hypothetical situation where he said he would want people to submit a detailed proposal,	I2		
31	which again tells me something about his leadership style in his approach, which is very formal.	I4		
32	There was no mention of using interpersonal skills to encourage people to talk and explain what they wanted, or to bring people around to his way of thinking, or that he was willing to listen and adopt their suggestions.		I3	
33	It was very much submitting something specific and he would think about it.	I4		
34	Um, he also said he would offer them the opportunity to work independently.	I2		
35	I think as a project manager that you absolutely have to give people the flexibility and the opportunity to work independently.	J5		
36	So that raised a little bit of concern with me.	J1		

L. Sample of Protocol Modification

No.	Segment	Performance	Response	Other
124	Aspiration for improving stress management skills.			I1
125	He didn't really talk about improving his ability to manage pressure.		I3	
126	So he seemed to be quite actively engaged in understanding what pressure he was under.	I4		
127	Yeah, he could identify these stress triggers.	J1		
128	But we didn't talk about long term impact of stress.		I3	
129	*(A score of 3 was given here)*	J1		

List of Publications, Participated Conferences & Working Papers

1) Zhao, S. (2021, December). Exploring the Cognitive Process of Interviewer During Employment Interview Using Think Aloud Method. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE. (pp.69)

2) December 2019 - London, UK (*MBAcademy International Business Conference*) (oral)

3) Beyond Decision Analysis - A Systematic Review of Process Research on Employment Interview. Working Paper.

4) From First Impressions to Selection Decisions: Tracking Interviewers' Decision Making at a Cognitive Level. Working Paper.