

**The Situational Judgment Test:
Cognition, Constructs and Criterion Validity**

Sally Mumford

Management School

University of Sheffield

PhD Thesis

July 2015

Acknowledgements

I would very much like to thank Professor Rod Nicolson, Professor Fiona Patterson and Professor Malcolm Patterson for the supervision of this work. I am very grateful for their continued advice, help and support during the preparation and finalisation of this thesis, and particular thanks to Professor Malcolm Patterson for his help after joining the supervisory team at such a late stage in the process.

I would also like to extend my thanks to Nigel Bax, Philip Chan and the Work Psychology Group for their input and help over the studies.

Finally, I would like to thank both my external and internal examiners, Dr Iain Coyne and Dr Kamal Birdi respectively, for their time and effort taken in examining this thesis and furthermore for their suggestions relating to alterations and clarifications throughout the work; I feel that their comments have very much improved the quality of the thesis and the studies therein.

Abstract

The Situational Judgment Test (SJT) is a personnel selection test in which a hypothetical scenario is provided, and the applicant asked to choose one or rate multiple answers. Unlike conventional knowledge tests, performance does not rely upon typical academic intelligence. A key requirement of research is to fully understand the construct validity, criterion validity and SJT theory.

This thesis is a set of 3 studies that investigates firstly, the construct validity of an SJT in medicine, secondly, establishes a new construct of interest to the SJT (formal operational thought, FOT) and thirdly, investigates FOT in relation to both the construct and criterion validity of an SJT in medicine.

In study 1 a current model of the construct validity of the SJT was examined alongside two measures assessing individual differences that are not currently included in SJT theory ('Need for Cognition' and 'Occupational Self-Efficacy') with the aim of increasing the amount of variance explained in SJT scores. Neither existing theory nor the additional variables significantly explained SJT performance.

Inspection of the literature revealed aspects of intelligence previously unexplored in relation to the SJT; fluid intelligence and FOT. A new measure of FOT was validated and FOT was conceptualised as a higher level ability for complex reasoning, independent to that used for primary cognitions.

Study 3 then assessed FOT, SJT performance, academic attainment and job performance within medical students. FOT scores did not significantly explain variance in the SJT scores. However, they had incremental validity over the SJT in explaining actual job performance.

The studies demonstrate that introducing FOT yielded additional unique variance over the SJT when explaining job performance and that this concept is a higher level thought store, independent of procedural and declarative knowledge stores. These ideas are presented in a theoretical model integrating existing intelligence and SJT theory.

Table of Contents

Acknowledgements	2
Abstract	3
CHAPTER 1: INTRODUCTION	10
Development and Evolution of the Thesis	13
Structure of the Thesis.....	14
Research Questions	16
Research Question 1:	16
Research Question 2:	16
Research Question 3:	16
Research Question 4:	17
Research Question 5:	17
Research Methods and Epistemology	17
Terminology	19
CHAPTER 2: SELECTION LITERATURE REVIEW	20
Human Resource Management and Personnel Selection Overview	20
Personnel Selection Methods.....	21
Typical Performance: Personality Measures	23
Maximal Performance: Cognitive Ability	29
Intelligence	29
History of intelligence theory	29
History of intelligence testing.....	31
Contemporary views on intelligence	32
High Stakes Testing in Personnel Selection.....	41
CHAPTER 3: SITUATIONAL JUDGMENT TEST & MEDICINE LITERATURE REVIEW	43
Situational Judgment Tests	43
SJT Development	45
History of the SJT.....	47
The SJT: Theoretical Ideas and Elements	48
The SJT: The situational element	49
The SJT: The judgment element.....	52

The SJT: Cognitive Psychology Theory.....	54
Implicit Trait Policies.....	58
ITPs and personality	58
ITPs and General Experience	60
ITPs and Cognition.....	62
The SJT and other methods of selection.....	63
Medical Selection and Assessment	66
Individual Differences.....	67
Structure of Medical Training and Selection	68
Assessment in Medical Education and Training	72
Medicine and the SJT	73
Criterion Validity of the SJT: Job Performance	75
Construct Validity of the SJT	79
Gaps in the Literature: Research Directions	90
Construct and criterion validity	90
Theory and Modelling the SJT	91
Medical Selection	92
Summary.....	93
Summary of Aims.....	94
Research Questions	94
Research Question 1:	94
Research Question 2:	94
Research Question 3:	95
Research Question 4:	95
Research Question 5:	95
CHAPTER 4: CONSTRUCT VALIDITY OF THE SJT	96
Study 1: Situational Judgment Test Performance and Demographic, Personality and Non-cognitively Orientated Variables in Medical Students.....	96
Acknowledgements and Role in Study	96
Introduction to Study.....	97
Theoretical Basis for Study	98
Hypotheses	103
Hypothesis 1: Personality and SJT performance	103
Hypothesis 2: Need for Cognition and SJT performance.....	104

Hypothesis 3: Occupational Self-Efficacy and SJT performance	104
Method	105
Study Design.....	105
Participants	105
General Procedure	106
Task Procedures.....	106
Results	110
Questionnaire Measures and SJT	110
Questionnaire measures and SJT domains	113
Personality, Need for Cognition, Occupational Self-Efficacy and SJT performance	115
Discussion.	117
Hypotheses Summary	117
SJT Performance and Personality	117
Introduction of New Constructs: Need for Cognition and Occupational Self-Efficacy.	119
Limitations.....	120
Summary and Conclusions	121
CHAPTER 5: PIAGETIAN THOUGHT.....	124
Study 2: Formal Operational Thought Proficiency and Cognitive Profiles in Control and Dyslexic Samples of Young Adults	124
Introduction to Study.....	124
Theoretical Basis for Study	126
Study Design	130
Sample Justification	130
Dyslexia	132
Study 2 Literature.....	133
The SJT and Intelligence.....	133
Piagetian Thought.....	135
Breaking Formal Operations Down.....	138
Consequences and relevance of formal operational thought	142
Formal Operational Thought and Contemporary Intelligence	145
Assessing Formal Operational Thought	146
Field Specific Tests: Medicine	146
Hypotheses	148
Hypothesis 1: Dyslexia and DST-S performance.....	148

Hypothesis 2: Dyslexia and nonverbal reasoning performance	148
Hypothesis 3: Dyslexia and formal operational thought	148
Hypothesis 4: Formal operational thought and lower cognition	149
Method	149
Study Design.....	149
Participants	149
Materials and Procedure	150
Formal Operational Thought	156
General Procedure	164
Results	164
DST-S General Performance	164
Formal Operational Thought	167
Discussion	170
Hypotheses Summary	170
Formal Operational Thought	171
The Formal Operational Thought Measure	173
Dyslexia.....	173
Limitations.....	174
Conclusion	175
CHAPTER 6: FORMAL OPERATIONAL THOUGHT AND THE SJT.....	177
Study 3: Situational Judgment Test performance, Formal Operational Thought, University Attainment and Job Performance in Final Year Medical Students	177
Introduction	177
Theoretical Basis for Study	178
Study Design	180
Hypotheses	181
Hypothesis 1: Cognitive ability and the SJT	181
Hypothesis 2: Cognitive ability and formal operational thought	182
Hypothesis 3: Formal operational thought and SJT construct validity	182
Hypothesis 4: SJT criterion validity	182
Hypothesis 5: Formal operational thought and job performance	183
Hypothesis 6: Formal operational thought, cognitive ability and job performance	184
Method	184
Study Design.....	184

Participants	184
Materials	185
Procedures	185
Summary of methods	192
Results	193
Sample Size	193
Cognitive Ability	194
Job Performance	194
Formal Operational Thought	195
Situational Judgment Test General Performance	196
Linear Correlations	198
Regression Analyses	198
Discussion	202
Hypotheses Summary	202
General Results Summary	202
Formal Operational Thought Discussion	205
Conclusions	210
Limitations	211
Future Research	212
CHAPTER 7: DISCUSSION	214
Summary of Thesis	214
Final Discussion	217
Practical Implications for Personnel Selection	218
Consequential Themes and Theory of Research	219
Future Research and Practice Recommendations	225
Final Comment	230
References	232
Appendix	263
Appendix I. Explanation of thesis terminology.	263
Appendix II. Breakdown of five factors of personality	265
Appendix III: Need for Cognition Scale (18 items)	267
Appendix IV: Short form version of Schyns & von Collani Occupational Self-Efficacy Scale	268
Appendix V. Diagram of balance tester for postural stability test.	269

Appendix VI. Phonemic segmentation stimuli.	270
Appendix VII. Spoonerisms stimuli.	271
Appendix VIII. Backwards span stimuli.	272
Appendix VIII. Formal operational thought test questions for study 2.	273
Appendix X. Information and consent form for study 2.	281
Appendix XI. Information letter and consent form for study 3.	283

CHAPTER 1: INTRODUCTION

This thesis focused upon the Situational Judgment Test (SJT) as a selection measure within the context of medical selection and assessment. Medical selection and assessment is undoubtedly an important field due to the large human and financial costs, both to the individuals involved directly in clinical care and the general population that fund the healthcare system in the United Kingdom and have interests in holding faith within the system. Selecting the correct individuals is therefore an area of organisational research that generates a lot of interest and is both relevant to society and complex to investigate. In this dynamic area of selection, new measures and ways of choosing doctors are often developed and introduced with the aim of improving the selection and assessment processes used.

The SJT is one of these newer selection methods that has relatively recently been used for selection of General Practice (GP) speciality posts (from 2007), for ranking medical students for their training years following graduation (from December 2012) and now even for entry into some medical schools (from course entry for Autumn 2013). It is therefore clearly important that this measure is well understood and utilised. Reflecting this idea, the primary and general aim of this thesis is to explore the SJT and to question what exactly is being measured by this test in the studies. More specifically, this thesis investigates the SJT in relation to the existing theory surrounding the construct validity of the SJT and then attempts to investigate constructs that may be assessed by the measure but have not yet been examined alongside the SJT. The SJTs used in the studies within this research project are designed for and completed by medical students who are either applying for their Foundation training years in the final year of their university degree or have assumed the role of a Foundation year 1 trainee. The focus of the thesis develops with emphasis upon an abstract area of intelligence known as formal operational thought (FOT). The theory used in this thesis allows for a multi-dimensional approach to understanding the SJT using ideas from organisational, cognitive and developmental psychology. The thesis incorporates ideas from these three different areas of psychology and in particular uses well established theory from both intelligence literature and the theory concerning the developmental disorder of dyslexia, as aids to further explain SJT performance and the construct validity of the measure.

The SJT is a low fidelity selection measure used within the context of medical selection and assessment. A high fidelity measure presents a replica of a job task (for example, an in tray task within an assessment centre which is highly similar to an actual job task) whereas a low

fidelity measure (such as an SJT) uses less realistic stimulus and responses are less exact approximations of actual job behaviours (Motowidlo, Dunnette & Carter, 1990).

An SJT is a test that provides a hypothetical scenario asking the applicant to rate answers or to choose one answer from multiple options. SJTs are a low fidelity simulation used in personnel selection for many fields, domains or purposes. They can typically be described as an explanation of a hypothetical work related scenario that then asks applicants to rate a number of responses upon how likely they are to respond in such a way and which are the most effective responses in the situation described. Responses are either in a ranking format or a forced option format (i.e. choose one answer from the options). Applicant responses are usually marked in comparison with those responses of experts within the specific field and a scoring key is developed prior to testing.

SJT are thought to predict job behaviour as based upon behavioural consistency ideas i.e. past behaviour is the best predictor of future behaviours (Wernimont & Campbell, 1968). Whereas high fidelity simulations allow for an actual scenario and real demonstration of procedural knowledge and skills, low fidelity simulations, such as the SJT, do not. Low fidelity simulations assess responses regarding procedural knowledge and skills, and these are seen as precursors for effective and later job behaviour (Motowidlo, Hooper & Jackson, 2006) in that they capture behavioural intentions rather than actual behaviour. In high fidelity simulations, the behavioural consistency ideas mentioned before are much more straightforward and theoretically apparent (Alon et al., 2009).

Weekley & Ployhart (2006) provide a comprehensive summary of the SJT as a method. They note that SJTs can be developed in different ways (critical incidents vs. job analysis), with different instruction formats (what would you do as a behavioural tendency instruction vs. what should you do as a knowledge tendency instruction) and response options (single best answer vs. ranking options) and these measures can be used for numerous purposes or fields (e.g. medicine, military professions or teaching job entry). This versatile measure has become increasingly popular since the early 1990's and the measure shows positive and promising findings. For example, SJTs have been shown to predict job performance (Lievens & Patterson, 2011). There is much research surrounding performance upon the SJT and the breadth of different topics investigated. However, there are theoretical gaps in the explanation of what underlies the SJT and a lack of 'identification of an underlying theme or convergence

towards a coherent theory' (Weekley & Ployhart, 2006, p6). Hence, it is not clear through existing theory what this measure is assessing specifically.

SJTs are typically acknowledged as differing as measurement methods and it is accepted that they can measure numerous and many different constructs (McDaniel et al., 2006). This is a consequence of the design of the measures for different fields, topics and/or purposes. For example, different SJTs will differ in the strength of correlations with personality variables or cognitive ability measures according to what they were designed to assess. Although, in broad theoretical terms there are links between the personality factors and cognitive ability constructs and SJT performance. Hence, similarly, the theory presented in this thesis concerning SJTs should be taken as an approximation and represents 'constructs assessed on average' (McDaniel et al., 2006).

As noted, the SJT has become increasingly involved at different stages within medical selection over recent years. Medical selection using SJTs is a high volume, high cost, high stakes, and high pressure operation taking place in most countries, with over 8000 candidates annually in the UK for just 3, 250 GP jobs alone (Patterson, 2009). The costs involved are illustrated by an example of training a General Practitioner in 2006; the annual training costs for a GP were approximately £87,000, which across the three year training programme totals £261,000 per trainee (Patterson, Lane & Carr, 2009).

Medical selection and assessment is an almost continuous process from medical degree entry to consultancy achievement, with yearly assessments following this point as well. The methods can vary, including interviews and academic screening for course entry, or within university options of measures include Multiple Choice Questionnaires (MCQ's), oral examinations, poster presentations, and clinical assessments, such as Observed Structured Clinical Exams (OSCEs). Due to being continually assessed and monitored medical professionals need to show both strong academic and non-academic qualities to both begin and to continue on a medical career path. The methods used for the assessing and selection are therefore highly important. Furthermore, it is important that the methods chosen can both deliver and identify not only the academically capable students, but the individuals who have the non-cognitive abilities to be a 'good' doctor, for example empathy, communication and organisational skills.

This introduction comments on the evolution of this thesis over time and gives explanation for how the direction and themes of research developed over the research project. The broad

ideas and concepts surrounding the SJT and an overview of medical selection have been described in this introductory chapter as a foundation of knowledge to facilitate the reading of the thesis. These issues are then covered in further detail in the following chapters. Chapter 1 concludes with a discussion of the epistemological stance of the researcher and methods used therein.

Development and Evolution of the Thesis

The work presented in this thesis represents four years of efforts to understand the SJT as a means of high-stakes personnel selection, and in particular to further explain beyond existing theory what the SJT is measuring. The researcher had the benefit of direct access to the work of a co-supervisor, Professor Fiona Patterson, on medical selection. Professor Patterson had recently completed a 10 year study from designing a medical selection SJT to the evaluation of it's' effectiveness in predicting subsequent medical performance. The results were very promising, confirming the value of the SJT as a component of the high-stakes selection process, but indicated that there was unexplained variance in subsequent performance and the theory behind the measure.

This thesis consists of three studies. The initial attempt to investigate the SJT involved assessment of the incremental validity (i.e. validity over and above another measure) of two measures assessing other individual differences from Murphy (2012) that are not currently included in current SJT theory, the 'Need for Cognition' and 'Occupational Self-Efficacy' variables, over personality factors in explaining variance in SJT performance. Although all SJTs are somewhat different due to the numerous purposes and designs of the measures, broadly speaking there are established links shown between cognitive ability, personality factors and SJT performance (McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006). The analyses indicated that the additional variables of Need for Cognition and Occupational Self-Efficacy did not account for any further variance in study 1. Therefore, further variables that might be expected to be involved in medical jobs or SJT performance were investigated. In particular, it was considered that job performance involves learning from experience on-the-job, and that a key variable not investigated in the SJT, or other components of medical selection, was the fluid intelligence required to benefit fully from post-appointment work experiences. A component of fluid intelligence known to be important in high level cognitive performance is the Piagetian construct of 'formal operational thought' (FOT), and consequently the researcher developed a measure of FOT to evaluate its effectiveness. Study 2 investigated the FOT measure in a sample of university students, contrasting dyslexic

students and typically achieving students, and demonstrated the sensitivity of the measure whilst confirming a theoretical conceptualisation of FOT as an ability. The use of the established dyslexia theory and previous research was useful to conceptualise a cognitive location and placing of FOT, i.e. FOT is a higher level ability whereas dyslexia is a consequence of impaired lower level abilities. This therefore allowed for the novel measure of FOT used to be assumed valid and used in study 3 alongside another SJT designed for medical selection.

In study 3, the FOT measure was administered to a cohort of graduating medical students for whom the researcher had access to data on academic performance, SJT performance and actual job performance. Interestingly, the FOT measure provided no additional variance in the SJT but it did provide additional unique contribution to variance in the on-the-job performance. This result provided a promising lead for further development of medical selection tests to augment a standard SJT with an FOT measure, albeit results were interpreted with caution as this key finding only remained significant at the .1 level when cognitive ability was introduced into the analytic models. Furthermore, cognitive ability was the strongest predictor of job performance and fully explained the relationship between the SJT and job performance.

The studies presented in this thesis provide information regarding SJTs and discuss the general literature surrounding this measure. However, the empirical results may be seen only as applicable to the specific SJTs used in the research. Results are discussed as generalisable (i.e. in a broad theoretical model) although it is noted that this may not necessarily be the case and each individual SJT may differ from another. Results should therefore be interpreted for the SJTs used in the research and then as a guideline for investigation into other specific SJTs using the ideas that follow in this thesis.

Following the description of the thesis structure below, the specific research questions will be presented with a discussion of epistemological stance, before the body of the thesis.

Structure of the Thesis

Chapter 2 builds upon the foundations of the literature described in this introduction with a review. The three studies are framed firstly, by a literature review of personnel selection. Chapter 3 continues the introductory sections with a literature review of SJTs in selection and medicine. In order to conform to the conventional Ph.D. format I have provided the literature review for intelligence and the literature review for the SJT at the start of this thesis along

with inspection of the SJT literature and an overview of the medical professional system/career path. This literature is relevant when discussing the SJT due to the established links with cognitive ability, although within the thesis these ideas are drawn upon from study 2 onwards.

Chapter 4 reports the first study of the thesis, investigating an existing model of construct validity for the SJT from McDaniel et al., (2006) with the inclusion of two new non-cognitive variables (Need for Cognition and Occupational Self-Efficacy) and their relationships with the measure. A detailed account of the measures used, the method and the analysis can be found, followed by a discussion of the results. The first study investigated the current modelling of the SJT (McDaniel et al., 2006), the constructs of interest and their validity described in this model. The results were somewhat disappointing and the model as described was not supported in this study for this SJT. The new variables did not add incremental validity over the existing model's variables in SJT performance. Consequently, the challenge evolved to further explain the SJT scores and investigate other constructs of potential interest and relevance.

Chapter 5 reports the second study of the thesis. This study required further exploration of the literature into cognitive psychology and in particular, the Piagetian reasoning ability of FOT, which is an abstract area of intelligence. Investigation of the literature led to inspection of the concept of higher level thought through Piagetian reasoning from the intelligence literature, and this was suggested as an abstract area of intelligence that may affect performance upon the SJT. This area of intelligence has not been previously investigated in relation to the SJT. A measure was developed and piloted using samples differing in cognitive abilities across basic cognitive tasks (e.g. rudimentary reading and writing skills). The two samples allowed for validation of the measure through group differences in lower level cognition performance (e.g. reading or writing) and similarities in higher level cognition performance (FOT). This novel construct (FOT) may be a new construct of interest when attempting to define what the SJT measures. This FOT literature only becomes relevant in study 2 and hence, is presented within the introduction to this second study along with a brief discussion of the developmental disorder dyslexia and a justification for using a sample of participants with this disorder in the study. This already well established and extensively researched topic allowed for further conceptualisation of FOT and to validate the novel measure used. Following the literature, a detailed account of the measures used, the method and the analysis can be found, followed by a discussion of the results.

Chapter 6 reports the third study of this thesis which investigates SJT performance, academic performance, actual job performance and FOT ability in final year medical students. A detailed account of the measures used, the method and the analysis can be found, followed by a discussion of the results.

This final study is followed by a summary of all three studies and the concluding discussions for this thesis in chapter 7. Conclusions were drawn by representing the results within an integrative model of existing theory and these findings, containing intelligence modelling, SJT construct/criterion validity and job performance relationships. The possibility of using Piagetian reasoning as a concept in personnel selection, as well as the directions for future research, are discussed.

This thesis tackles five main research questions and these are noted below.

Research Questions

Research Question 1:

To what extent can variables from existing SJT theory explain the construct validity of the SJT within a medical selection context?

Research Question 1 is tackled in study 1 (chapter 4) and study 3 (chapter 6).

Research Question 2:

Can the inclusion of additional non-cognitive variables assessing individual differences further explain the construct validity of the SJT over and above variables from existing SJT theory?

Research Question 2 is tackled in study 1 (chapter 4).

Research Question 3:

Can the inclusion of an abstract intelligence variable assessing higher level thought further explain the construct validity of the SJT in a medical selection context over and above variables from existing SJT theory?

Research Question 3 is tackled in study 3 (chapter 6) with the development of a measure for the research in study 2 (chapter 5).

Research Question 4:

Can the inclusion of an abstract intelligence variable assessing higher level thought further explain medical job performance independent of the SJT within a medical selection context?

Research Question 4 is tackled in study 3 (chapter 6).

Research Question 5:

How can existing theoretical literature surrounding the SJT's construct and criterion validity be expanded and linked to develop an integrative conceptual model of the SJT?

Research Question 5 is tackled throughout the thesis and in particular in the final discussion (chapter 7).

Before the body of this thesis, the epistemological stance of the researcher and the methods are discussed below.

Research Methods and Epistemology

There is a general domination of quantitative and empirical research within the social sciences (Martin, Carlson & Buskitt, 2007). Regardless of this fact, these methods are still in fact the most suitable for the studies in question. Qualitative approaches are less suitable for research concerning the validity of personnel selection techniques, as it is necessary to rank people in terms of capabilities and suitability for a job role, especially when looking at medical selection and assessment methods.

Quantitative methods allow researchers to reduce behaviours to quantities or to a state where they are quantifiable and therefore easily ranked and distinguishable based upon performance (e.g. by a score upon a questionnaire). The typical assessment of academic achievement and progress through careers in medicine is done through objective and quantifiable measurements that allow for defensible and meaningful decisions to be made based upon candidate's performances against other applicants. It is logical that investigating such measures is done using a similar methodology that can be defended in the same manner.

Qualitative methods move the focus of research from quantifying results to the exploration of depth within the material and the investigation of 'meanings, context and a holistic approach to material' (Hayes, 1997). They are driven by the principal of 'verstehen' which states that behaviours and actions vary according to social circumstance or a researcher's understanding of them.

It is possible to see an instance where using qualitative methods such as interviews may be of benefit though. For example, interviewing those who fail to pass selection tests or even those who are selected for a job role and then appear to be struggling or failing to deal with the job may allow for descriptive exploration and analysis of the factors and rich information regarding individual differences involved. However, at present it is necessary to build tools and to both objectively and fully understand the theory behind the selection process through ranking and quantitative objective-led methods of separating individuals according to levels of the noted variables.

The philosophical basis for this research stems back to Poppers' positivist methodologies and the assessing of hypotheses to give falsification and protect against 'fanciful theorizing in management research' (Donaldson, 1996, p.164). The ideas can be named interchangeably as positivism, empiricism or modernism. The key idea for this movement is that observation of the empirical world gives the only possible foundation for knowledge, and furthermore that this foundation is therefore neutral, value-free and objective. Other characteristics of positivism include that all theoretical statements must be capable of empirical testing and falsification. Following on from this point it is therefore logical that the subjective realms of the world (that are non-falsifiable or observable) are held to be beyond the scope of scientific investigation. Positivists also support the idea of a constant conjunction; this idea gives basis for the establishment of cause and effect relationships in that, if an effect follows a cause in a predictable and regular manner they can be linked through a causal relationship.

These research methods that have been decided upon and described are purely quantitative. Whereas within qualitative research the researchers themselves are seen as central to the collection, analysis and interpretation of data (i.e. subjective views and researcher involvement) this is not the case for quantitative research. Researchers are asking for the causes of certain variables and looking for relationships enabling *erklaren*.

Erklaren describes the full explanation of behaviour from the observation and examination of causal variables resulting in those behaviours. These deductive research methods allow for phenomena to be explained by providing a deterministic account of the external causal variables which brought about the phenomena in question. There is no subjective interpretation on the part of the researcher or their influences upon the world or cultures to affect this research. The data collected will be objective measures of the participant's

cognitive capabilities and the potential bias involved in subjective qualitative data collection, analysis and interpretation can be avoided.

The epistemology (i.e. the questioning of whether the knowledge gained through research is justified) for this thesis therefore falls into seeing the status of human behaviour as fairly determined. The observations and reporting of this from a personal point of view can be objective; the real world is out there and awaiting discovery regardless of one's perceptions, and the lack of qualitative methods in this research allows for no researcher interference with either the behaviours of others, the study results or inferences made from these things.

Terminology

Throughout the thesis there are references to specific terminology regarding the explanation of relationships between variables and the validity of measures. Whilst these are explained when first mentioned, so that these can be referred to by the reader throughout, these terms are noted with a short explanation of each in Appendix I.

CHAPTER 2: SELECTION LITERATURE REVIEW

Human Resource Management and Personnel Selection Overview

Human Resource Management (HRM) can be defined as ‘a strategic, integrated and coherent approach to the employment, development and well-being of the people working in organisations’ (Armstrong, 2009, p.4).

Dyer and Holder (1998) further specify HRM practice as working towards a number of goals falling into four main headings; contribution (e.g. what kind of employee behaviour is required?), composition (e.g. what headcount is required and specifying a staffing ratio and skill mix), competence (e.g. what is the level of behaviour required?) and commitment (e.g. how do employees need to feel attached and be identified in the workplace?)

From a psychological point of view, HRM includes many branches that have to consider numerous aspects of the work system, people and environments, for example, recruitment, management, relationships, the organisational setting, learning and development, rewarding people, employee relationships, health and safety as well as policies and procedures in the workplace (Armstrong, 2009).

Due to the diversity of occupations and organisations that HRM practices deal with it is often found that there are different approaches to the practice. For example, Storey (1989) makes a distinction between a ‘hard’ and a ‘soft’ approach to HRM. Some companies may deal with people as a ‘tool’ to achieve a business and economic advantage for the company over other organisations. This ‘hard’ view of HRM is a business strategic stance looking for competitive advantage. In contrast, a ‘soft’ view of HRM looks more to employees as progressive and responsible beings, seeing job productivity, and ultimately company productivity, as a result of humanistic factors such as job satisfaction.

In fact, selection itself, whilst a broad field with extensive publications and research within and surrounding the topics, is very much still a developing area and has important questions left unanswered (Ryan & Ployhart, 2013). In their review of personnel selection, Ryan & Ployhart demonstrate the range of possibility and scope for development in their review concerning designing selection systems, the outcomes of selection, construct choice, methods and evaluation, to name but a few of the areas of selection as a field.

Within HRM, the term personnel selection covers the process of assessing and selecting those who are most suitable and who will perform to the highest and most productive of standards within a certain job role. This may be classified as the objective of people resourcing strategy, which sees applicants as ways to achieve competitive advantage and the highest levels of performance and productivity. People are rewarded and maintained within the organisation through benefits, psychological contracts and mutual trust (Armstrong, 2009). The strategic HRM approach to resourcing places importance upon organisational capability and therefore searches for people whose attitudes and behaviour are likely to fit into the company. It requires a 'systematic approach, starting with human resource planning and proceeding through recruitment, selection and induction, followed by performance management, learning and development, recognition and reward' (Armstrong, 2009, p.481).

This process begins with a job analysis to establish a clear job description and a clear person specification, including what competencies are necessary for the particular role. Certain selection criteria can then be based upon these job specifications and a selection method chosen according to the criteria in question as well as the cost, fairness, validity, reliability and legality of the measures under consideration. Candidates will then be recruited through advertising to the selection process. Finally, evaluations and judgements will be made objectively about the applicants resulting in one or more candidates being hired for the role.

Personnel Selection Methods

Vinchur (2007) notes that there are some generalisations that can be made about psychological interactions as well as effect and role within personnel selection over approximately the last 100 years. He states that intelligence testing, interviews, biographical information and job skill tests were firstly popular and have remained so, but gained vast improvements and advancements over the years. Personality assessment and situational methods (i.e. those that describe/demonstrate scenarios and require responses) were introduced later on whilst the criteria of reliability and validity of measures/scales has remained fairly stable, robust and durable throughout the introduction of new angles and methods of selection. Vinchur (2007) also notes that most topics of measurement (e.g. personality or intelligence) and practices have remained the same with improvements and research surrounding them (e.g. new forms of administration using technology), as opposed to new predictors themselves being introduced (e.g. social skills). It is apparent therefore that with such robust and long standing practices within personnel selection, it is consistently

important that we have a full and up to date understanding of these measures, and especially of the more recent introductions, e.g. the situational tests.

There are many personnel selection assessment methods in use and available today. These include interviews, psychometric tests, references, bio data, work sample tests and assessment centres. These selection methods can give evidence of an applicant's abilities, skills, weaknesses, characteristics and attainments (Armstrong, 2009). In general, selection methods can be split into two categories; maximal performance (e.g. cognitive ability, achievement, psychomotor and physical measures) vs. typical performance (e.g. a personality, interest, motivations and values measures). Broadly speaking these cover tests that rank people on how well they perform and in which applicants should do the best they can (maximal performance), and tests that do not have correct/incorrect answers (typical performance).

There is extensive literature examining the important issues about these methods, for example, the accuracy of the method in terms of validity, the reactions to the method and the application of the method (what it is used for and within which settings is it commonly used) (Patterson, Lane & Carr, 2009). Methods may be chosen based on a variety of criteria by job analysts and recruiters. For example, the reliability and validity of a method, the fairness, the legality, the cost, the practicality, candidate reaction or the generality of a method may all be considered. In general, the validity and reliability of a method are key to assess. The reliability of a measure characterises whether a measure is consistent across conditions. This can be looked at across time (test- retest reliability), in parallel forms (e.g. similar items of equal difficulty), internal reliability (i.e. within the scale itself, such as Cronbach's alpha reliability coefficient; Cronbach, 1929) and/or inter-rater reliability (i.e. how much two raters agree using the same instrument to assess the same applicant).

Validity is slightly more complex and can be broken down into different aspects. For example, face validity (the apparent plausibility of the test), construct validity (the test measures what it is supposed to), criterion validity (how well one variable predicts performance upon another variable) and synthetic validity (the use of valid tests to produce a battery of tests for assessment).

Selection methods all vary in their objective validity and reliability strengths, and as one might expect, certain methods are more suitable for certain job roles or companies, and consequently, validities/reliabilities are different according to the job role or company the

measure is administered for. Two of the most researched and established methods used in personnel selection are the psychometric measures of personality tests (typical performance) and cognitive ability tests (maximal performance).

Personality and cognitive ability are constructs of great interest and involved in current modelling of the Situational Judgment Test (SJT; McDaniel et al., 2006) which is the focus of this thesis. Existing theory suggests that personality and cognitive ability are two key constructs assessed by the SJT (McDaniel et al., 2006) and hence these areas of selection will be discussed in more detail below. Both are constructs that are later discussed in this thesis and are both seen in the general SJT theory. Furthermore, researchers are encouraged to use the surrounding theory of these concepts when considering, investigating or developing SJTs (Campion, Ployhart & MacKenzie, 2014).

Firstly, personality as a ‘typical performance’ variable is considered and then cognitive ability as a ‘maximal performance’ variable.

Following the discussion of personality measures and assessment, the literature review discusses cognitive ability (intelligence). The cognitive ability discussion below considers intelligence as a concept, the history of intelligence, the testing of intelligence and contemporary views of intelligence. This thesis develops to introduce new constructs of interest related to the SJT, alongside the constructs of personality and cognitive ability within intelligence. It is therefore important to give a clear explanation and background of these well established and researched constructs.

Typical Performance: Personality Measures

Personality can be defined as a particular pattern of behaviour and thinking that prevails across time and situations and differentiates from one person to another (Martin, Carlson & Buskist, 2007). Henceforth it logically follows that the investigation and explanation of one’s personality will help explain individual differences between subjects. There are many personality tests and there is an underlying assumption in the literature that personality traits can be measured. Frank (1939) notes that the difficulty with measuring personality is in fact ‘a lack of any clear-cut conception of what is to be studied’ (p.389, cited in Murphy, 2012). The difficulty is also that a theory of personality on which to build assessment techniques must consider the uniqueness of individuals, the different traits shown across different situations, and also the fact that there is a large amount of commonality in human behaviour regardless of personal uniqueness (Murphy, 2012).

Hogan (2006) explains personality characteristics as being directed through one of two routes; getting ahead (i.e. for personal achievement or gains) or getting along (i.e. social interaction). Murphy (2012) notes that numerous personality traits may in fact be relevant to both orientations. For example, if one is thoughtful to friends then one is likely to be thoughtful in the workplace and may do extra work for others and receive reward or be a better employee.

Personality theories

Different researchers view personality in different ways. For example, personality may be viewed in terms of personality types, i.e. people fall into categories of one type of personality or another. Some researchers also view personality from a purely situational position or a purely trait-oriented position (Murphy, 2012). Pervin (1980) suggested that an in-between stance was preferable and incorporated these two extremes. Pervin (1980) defined personality as a set of characteristics of a person (or people) that accounts for consistent patterns of response to situations.

The more dominant view in the modern literature is the trait theory point of view. Trait theorists investigate personality by looking at the extent that a subject expresses a certain personality trait, for example, agreeableness. Traits are seen as lasting across time and situations; a concept that one takes with themselves from situation to situation. Trait theorists describe the regular patterns of behaviour that people exhibit and do not claim to show an all-encompassing account of behaviour (Goldberg, 1993).

Allport (1897-1967) was an early trait theorist and tirelessly identified all the words (18,000) in a dictionary that described aspects of a personality (Allport & Odbert, 1936). After eliminating temporary states of emotion and synonyms he was left with over 4,000 words still. Hence, the complexity and difficulty of this task of measuring such a large and open concept is evident. Nevertheless, research has moved on extensively and researchers tend to agree that there are between 3 and 16 personality factors in the recent literature.

Cattell (born 1905) used Allport & Odbert's (1936) list of adjectives and the statistical process of factor analysis to identify sixteen personality traits. Cattell referred to the factors as source traits; the cornerstones on which personality was built. Some of the most important traits identified included warm-cool, concrete thinking-abstract thinking, affected by feelings-emotionally stable, submissive-dominant and sober-enthusiastic.

Hans Eysenck (1916-1997) was another trait theorist who used factor analysis. Eysenck identified three main factors and these were extraversion, neuroticism and psychoticism. This theory has extensive support and the three factors have been found to have the highest validity of all the proposed personality factors (Kline, 1993).

The most dominant and widely used theory of personality in the modern day is the five-factor model (McCrae and Costa, 1985; 1987; 1990). The model is regarded as robust (Magai and McFadden, 1995). The model was built upon further word analysis, similar to Allport, by Tupes and Christal (1961) and Norman (1963). The five factors are neuroticism, extraversion, openness, agreeableness and conscientiousness. These can be measured using the Neuroticism, Extraversion and Openness Personality Inventory (NEO-PI, McCrae & Costa, 1990) which consists of 181 items that are descriptive statements. The Big Five Personality Scale (Goldberg, 1992) also assesses the big five model of personality.

There is extensive endorsement and use of the 5 factor personality idea within personality research and it appears to be a valid and reliable measure with the five noted variables being present consistently within research (e.g. Goldberg, 1992; Norman, 1967; Bond, 1979). The Cronbach's alpha reliability coefficients were noted as .81, .72, .78, .85 and .79 in the original study (Goldberg, 1992). The model allowed for researchers to categorise numerous traits into the five categories and make progress in understanding personality and occupational performance difference relationships across the early 1990's (e.g. Tett, Jackson & Rothstein, 1991; Barrick & Mount, 1991).

Although the 5 factor model is widely accepted within the literature, there is still some debate about the structure and universal applicability, and in particular based upon inter correlations between factors, the existence of higher order factors and/or lower order facets of the model (Mount & Barrick, 2012.) The literature appears to support the idea of two higher order factors; firstly, a 'factor α ' (also referred to as 'stability' or 'communion') consisting of conscientiousness, agreeableness and emotional stability, and 'factor β ' (also referred to as 'plasticity' or 'agency') consisting of extraversion and openness to experience (Mount, Barrick, Scullen & Rounds, 2005; De Young, 2006; Digman, 1997; Markon, Krueger & Watson, 2005; Mount et al., 2005). Factor α (stability/ communion) has been defined as socialization processes, conformity and the extent to which one is consistent in motivation, mood and social interactions (Mount & Barrick, 2012). Factor β (plasticity/ agency) has been defined as personal growth, enlargement of the self, nonconformity and the extent to which

an individual searches for new experiences, particular those of an intellectual and social nature (Mount & Barrick, 2012).

Researchers have also suggested that there are numerous lower order facets within the main five factors that may provide more predictive power, but research has failed to produce a well-accepted taxonomy of these facets (Mount & Barrick, 2012). For example, rugged individualism (Hough, 1992), masculinity-femininity (Costa, Zonderman, Williams & McCrae, 1985) or social adroitness, competence and insight (Ashton, Jackson, Jeles & Paunonen, 1998). There are suggestions of six factor models as improvements to the five factor model. For example, the HEXACO including an honesty-humility factor (Ashton & Lee, 2005; Ashton et al., 2004). However, McCrae & Costa (2008) counter this suggestion with the argument of honesty-humility being encapsulated by the agreeableness domain in the five factor model. Mount & Barrick (2012) note the importance of researchers choosing the appropriate level of granularity in their assessment of personality to fit their particular research questions and variables.

Personality tests
Following the consideration of personality theories, the testing and assessment of personality will now be discussed.

Personality tests can broadly be separated into objective or projective measures. Objective personality tests are similar to classroom tests and generally contain multiple choice format questions asking the applicant to indicate how much they agree or disagree with a statement. The questions are clear and scoring is therefore determined before testing. An example of this type of test is the Minnesota Multiphasic Personality Inventory which was first published by Hathaway and McKinley in 1939. The scale was made up of true/false items based on subscales of lying, defensiveness, frequency and cannot answer scales. There is some caution that should be taken with marking scales such as this, for example McCrae & Costa (1990) warn against the rational thought people may have whilst participating that may then result in wrongful answering and/or wrongful interpretation. They noted the example of an item reading 'Before voting, I thoroughly research all the qualifications of all the candidates' and if they answer true to this it would be interpreted a lie (the obvious assumption is that nobody does this extensive research before an election). However, in practice, one might say to oneself 'surely they did not mean to say that, and they meant to ask me if I take voting seriously and know about the candidate' and therefore answer with a 'true' reply. A correction factor should therefore be applied during marking, and instructions given clearly before test administration.

In contrast to this, there are measures of personality derived from psychodynamic personality ideas and these are called projective tests of personality. Psychoanalysts hold that behaviour is built upon unconscious processes and that through the revelations of these then a personality will be revealed. Tests are highly ambiguous compared with objective ones, and there are no specific right or wrong answers. This may reveal a richer and more descriptive nature of answers and henceforth, more detail than objective measures may be able to give. Furthermore, it is unlikely the candidate would know which answers could be more desirable than others and so social desirability may be lower. However, they are highly liable to interpreter bias as this is a subjective process and validity may be low on such a measure (Groth- Marnat, 1997) as well as reliability (Lilienfeld, Wood & Garb, 2001). An example of a projective test of personality is the Rorschach Inkblot Test, which was first published in 1921, where subjects are asked to determine what they see on a number of cards with black and white or colour symmetrical ink blots.

Personality in selection

Mount & Barrick (2012) discuss controversy surrounding the use of using personality tests in personnel selection based upon the belief that personality tests have low validity in predicting work outcomes. Contrary to this belief, the body of research tackling the personality and job performance/outcomes has grown rapidly and there are now thousands of empirical studies into this relationship making up a solid and extensive base of literature (e.g. Barrick & Mount, 2005; Morgeson et al., 2007; Ones, Dilchert, Viswesvaran & Judge, 2007). Studies have shown relationships and meaningful findings that show merit in using personality tests in selection. For example, personality in high schools predicted career success over fifty years later (Judge, Higgins, Thoresen & Barrick, 1999). Used alongside measures of cognitive ability, personality tests can help to predict both work performance and behaviours under motivational control (e.g. counterproductive work and citizenship behaviours) respectively (Mount & Barrick, 2012).

Whilst the research base grows, the relationships between personality and job performance, which is arguably the most important criterion to predict within personnel selection, remain moderate at best for predicting behaviour (Morgeson et al., 2007). Personality is not as strong as cognitive ability when predicting job performance (correlation values typically <.30), although Mount & Barrick (2012) note the lack of meta-analytic research on personality predicting behaviours influenced primarily by motivation at work which may be a factor to

consider. Nonetheless, the Judge et al., (1999) study remains strong evidence for personality as a predictive measure. Furthermore, including personality in job selection has been seen to predict outcomes such as highly motivated employees (Judge & Iles, 2002), those willing to go above and beyond the call of duty (Hurtz & Donovan, 2000), those less likely to engage in theft of time, resources and money (Berry, Ones & Sackett, 2007), team member effectiveness (Bell, 2007) and leadership effectiveness (Judge, Bono, Iles & Gerhardt, 2002). Although correlations are rarely reported as $>.30$, it is argued that as employers will find all of these criteria useful as predictions. When all aspects taken together, the value of personality as a predictor measure may be somewhat invaluable (Mount & Barrick, 2012).

Furthermore, when relating personality to other individual differences, Judge, Jackson, Shaw, Scott & Rich (2007) conducted a meta-analysis of the relationship between cognitive ability and the five factor model. They found that only the openness to experience domain correlated with cognitive ability above $.10$ ($r=.22$), with an average correlation value across all the five domains of only $.07$, suggesting that the personality traits and cognitive ability measures assess very different and independent constructs. Similarly, the correlations between personality factors and interest attributes are generally reported as low in meta-analytic research (Barrick et al., 2003; Larson, Rottinghaus & Borgen, 2002; Mount et al., 2005). Hence, it is important to include personality in selection as it will give greater predictive value when used along these other independent individual difference predictors.

Meta-analyses over the last fifteen years suggests that specific personality traits predict successful performance and behaviour at work (Mount & Barrick, 2012) and correlations of the five traits for numerous criteria are noteworthy (Ones, Dilchert, Viswesvaran & Judge, 2007), e.g. the job outcomes of overall job performance ($r=.23-.27$), task performance ($r=.15$), training performance ($r=.40$), and certain types of performance such as, teamwork ($r=.37$), entrepreneurship ($r=.36$) and leadership ($r=.45$). Validities appear to be stronger for behaviours that are more discretionary and volitional (e.g. citizenship behaviour and training success) as opposed to more closely monitored or structured behaviours (e.g. task or overall job performance; Barrick & Mount, 2012).

Following this discussion of typical performance assessment, maximal performance assessment will now be described. This includes the history of intelligence, the testing of intelligence and contemporary ideas and modelling of intelligence which goes on to inform the theory and study choices in this thesis. Both personality and cognitive ability measures

and theory are used within the studies in this research project and are discussed in relation to the SJT in the following literature.

Maximal Performance: Cognitive Ability

It is necessary to discuss the background in the literature to intelligence before considering the cognitive ability tests themselves. The intelligence theories and literature inform the measures used within personnel selection and the widely used and established cognitive ability assessment measures. Furthermore, when questioning new methods of personnel selection and the theory behind them, it is both useful and necessary to target past empirical research and literature surrounding intelligence, in order to establish what aspects of cognitive ability are being specifically assessed by the measures. The theoretical understanding and conceptualisation of intelligence and the aspects within it are key when designing, implementing and evaluating personnel selection methods that assess both typical cognitive ability and other aspects of intelligence.

Intelligence

Cognitive ability tests are measures of intelligence that assess a range of mental abilities and in particular, learning in a specific area, through a selection of questions, problems or tasks. Neisser et al., (1996a) describes intelligence as a person's ability to learn and remember information, to recognise concepts and their relations, and to apply the information to their own behaviour in an adaptive way. Tests of intelligence may be called 'general ability tests' as they are typically concerned with the establishment and measurement of 'general intelligence'. The concept of general intelligence is widely known as 'g' which was termed by Spearman (1927), who proposed one of the first theories of intelligence.

History of intelligence theory

Spearman developed the two-factor theory of intelligence, explaining intelligence as a product of firstly, the 'g' factor (a general factor) and the 's' factor (a factor specific to the test). Instead of calling the g factor 'intelligence' Spearman defined 'g' as a combination of three qualitative principles of cognition; the apprehension of experience (i.e. ability to understand and perceive what one experiences), the inferring of relations (i.e. making sense of relations between concepts) and inferring of correlates (i.e. making sense of a rule and the ability to apply it to another situation). In his research, Spearman concluded that the general factor 'g' was responsible for the correlations he found among different tests of abilities.

In the 1940's, Cattell conceptualised intelligence using the Gf-Gc theory (fluid-crystallised theory) and this work was based upon his own and Thurstone's factor-analytical work in the 1930's. Fluid intelligence concerned inductive and deductive reasoning that originated and were affected by biological, neurological and environmental interaction factors, whereas crystallised intelligence consisted of knowledge from the influences of acculturation (Cattell, 1957). This original theoretical stance on intelligence was further developed by Horn in 1965 to include four other abilities. These four other abilities were visual perception/processing (Gv), short-term acquisition and retrieval (Gsm), long-term storage (Glr) and processing speed (Gs). Over the next 30 years, factors were added following research and as a result of Horn's work, along with colleagues, there then existed a 10-factor Cattell-Horn Gf-Gc theory (Horn, 1991). This included the original abilities, as well as reaction time and decision speed (Gt), quantitative abilities (Gq), reading/writing ability (Grw) and auditory processing ability (Ga).

In the meantime, Sternberg (1985) went on to develop a triarchic theory of intelligence concerning the structure of intelligence. Hence, there are three aspects of intelligence involved; componential intelligence (i.e. mental mechanisms used to plan to execute a task), experiential intelligence (i.e. ability to deal with new situations and solve problems that have already been dealt with before) and contextual intelligence (i.e. intelligence reflecting behaviours developed from natural selection). He described metacomponents, performance components and knowledge acquisition components.

Metacomponents decide the nature of a problem in hand. For example, a skilled reader will decide how long to spend reading a passage dependent on how much information they want to acquire from it (Wagner and Sternberg, 1985). Performance components are the processes used to perform the task in hand and knowledge acquisition components are used to gain new knowledge from sifting out the unnecessary information. Stemler and Sternberg (2006) also describe creative, analytical and practical thought; creative thought is invoked when tasks are novel, analytical thought is invoked when familiar problems that are fairly abstracted away from life are encountered and practical intelligence is invoked when components are applied to everyday life experiences in order to adapt to, shape and select environments. Thus, the same components applied in different situations consequently invoke different types of thought.

Carroll (1993) also conceptualised the structure of intelligence and in his large scale review of cognitive abilities worldwide suggested a three stratum theory. Modern day views upon intelligence now generally accept such a hierarchical model. Such views describe a general mental ability 'g' as an upper stratum level consisting of the capacity to process, store and retrieve information. The middle stratum consists of fluid intelligence (ability to think logically), crystallised intelligence (ability to use skills and experience), general memory and learning, visual perception, ideational fluency, knowledge, achievement and perceptual speed. The lower stratum is made up of approximately seventy specialised abilities, for example, discriminate musical pitches and lexical knowledge. Further discussion of contemporary views of intelligence follows a summary of the history of intelligence testing that preceded and provided the foundations for modern ideas.

History of intelligence testing

Binet and Simon (1905-1911) were asked to develop a way of evaluating the developmental status of children and in particular, a way to identify those children who would need extra schooling and would not benefit from mainstream school admission. They developed a battery of tests assessing maximal performance on a range of tasks focusing on higher mental processes such as attention, thinking and memory. For scoring, they introduced the concept of mental age (MA) against that of chronological age (CA) to give an intelligence quotient ($MA / CA \times 100 = IQ$). Using a standardised set of scores with mean =100, 1 sd =15, and assuming that IQ falls across the population in the pattern of a normal distribution, it is therefore clear which people fall into the lowest and highest ranges of intelligence. However, it should be noted that this is an artificial representation of the real distribution of ability, and in fact, intelligence may not even be a singular concept (Andrade & May, 2004). In 1960, this version of the Stanford-Binet test was released using the deviation IQ, instead of a ratio IQ, which resulted in children's scores being compared to others of the same age.

Another renowned early researcher in the intelligence field was Wechsler. Wechsler developed the Wechsler Intelligence Scale for Children (WISC, 1949; 1974) and the Wechsler Adult Intelligence Scale (WAIS, 1955). The newest version of the WAIS is known as the WAIS IV (2008) and is a range of many individual tests validated on 2450 individuals from ages 16-89 across numerous cultures. The subtests included vocabulary, block design, comprehension and object assembly. Each of the major scales then consists of further subscales. For example, within the verbal scale there would also be sub scales of comprehension, information and similarities.

In terms of adult intelligence testing, Galton (1883-1928) is reported as the first investigator to perform large scale testing. However, Ackerman & Heggestad (1997) noted that the criteria assessed (e.g. hearing) did not meet current measures of intelligence. They noted numerous steps towards achieving maximal performance and complete testing in adult intelligence (for example, Cattell & Farrand, 1896; Sharp, 1899; Wissler, 1901; Yerkes, Bridges & Hardwick, 1915) but this goal does not appear to have been achieved until the nonverbal Army Beta Examination (as seen in Yoakum & Yerkes, 1920). They note progression after World War 1 to using Thurstone's (1919) Intelligence Test IV and the use of flippant instructions for use as 'optional' entry requirements to institutions.

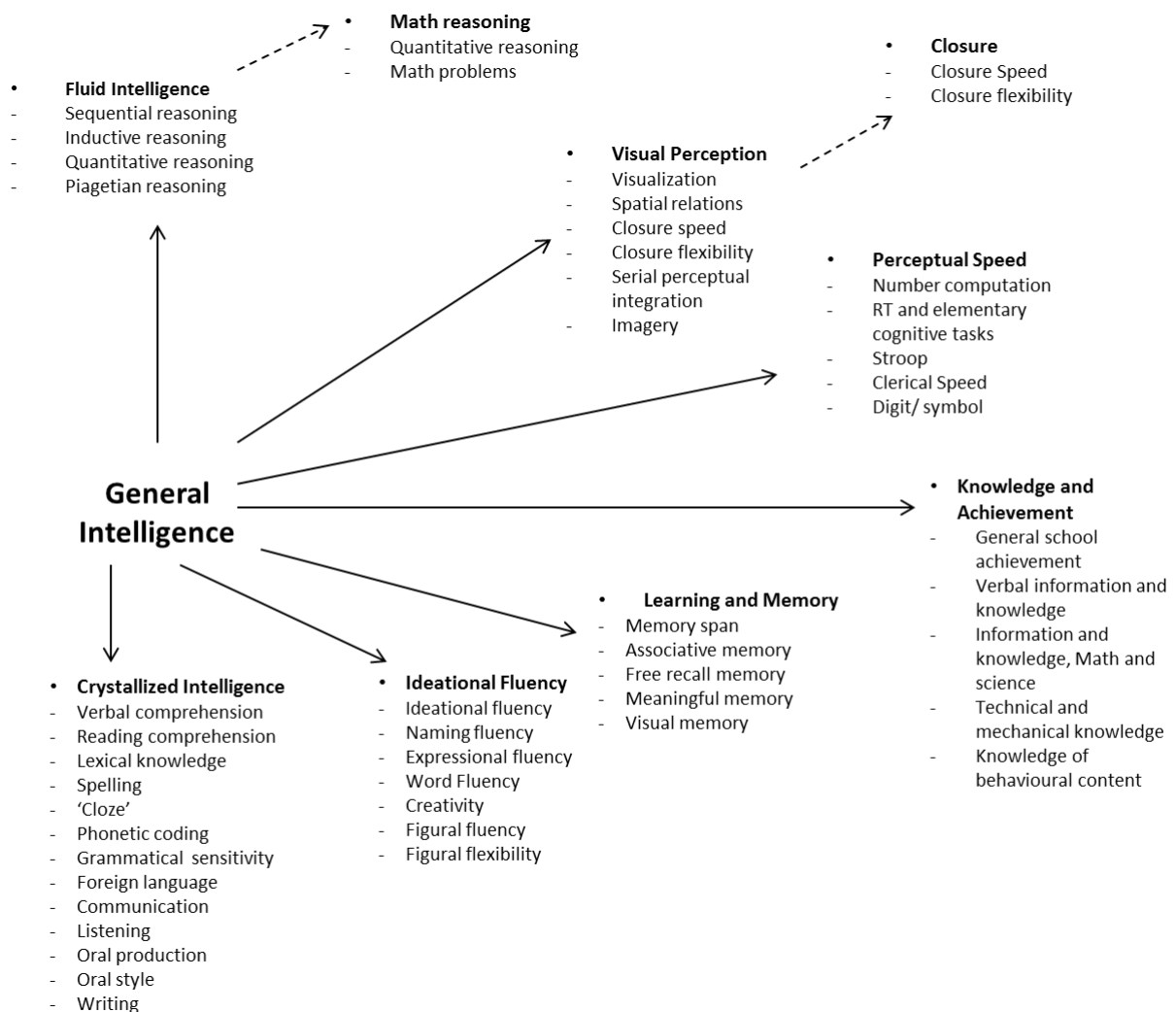
The WAIS is now the most widely used, established and validated test of adult intelligence and has the advantage of scoring verbal and performance abilities separately (Martin, Carlson & Buskist, 2007). Modern measures of the WAIS have Cronbach's alpha reliability coefficients ranging from .92-.98 for the domains assessed, hence internal reliability is very high (Sattler & Ryan, 2009). This is noted as especially useful for neuropsychological assessment and the examination of brain damage. It is accepted as a useful and reliable test of intelligence and assumed that as well as tapping into the smaller intelligence aspects of the middle and lower striatums, it generally assesses 'g' due to the fact that sub scale scores tend to correlate well with each other (i.e. someone tends to score well across all tests, or poorly across all tests). Interpreting such overlaps between subtests has been controversial (Andrade & May, 2004) as it is dependent on viewing 'g' as a real measure of ability (i.e. g is a result of environment and experience interaction giving neural development and is a key part of intelligence) vs. 'g' as a statistical artefact from the correlational analyses (i.e. g is a test answering ability and only once this is discarded, is there then rich intelligence information to be investigated).

Contemporary views on intelligence

Modern views on intelligence now tend to integrate intelligence theories and reflect the synthesis of recent research and in particular, the factor analytical studies and meta-analyses (Flanagan, Ortiz & Alfonso, 2013). The mostly recently noted theory is a combination of Carroll's three-stratum theory, and Cattell and Horn's Gf-Gc theory. This is known as Cattell-Horn-Carroll (CHC) theory. This is noted as the most empirically supported and well-validated theory of the structure of cognitive abilities (Flanagan, Ortiz & Alfonso, 2013).

Before the conceptualisation of the CHC theory, Ackerman & Heggedad (1997) noted a similar representation to the CHC theory and combined ideas from both the original models. As noted, most theories nowadays are in agreement that there is a hierarchical structure representing intelligence. There is also some agreement about how scales relate to each other, with a general level of 'g' followed by a broad group of abilities and then a narrow group of factors in the lower stratum. They adapt a list and structure of abilities from Carroll (1993) as shown diagrammatically below in Figure 1. Carroll's (1993) theory of intelligence is widely accepted as the most comprehensive account of intelligence (Andrade & May, 2004) as he based the theory upon results from nearly five hundred experiments carried out across cultures over a period of 60 years.

Figure 1. A list and structure of ability constructs within intelligence (from Ackerman & Heggedad, 1997, based on information from Carroll 1993).



*(Third order construct = General intelligence, g; second order constructs shown with solid lines; first order constructs shown with dashed lines).

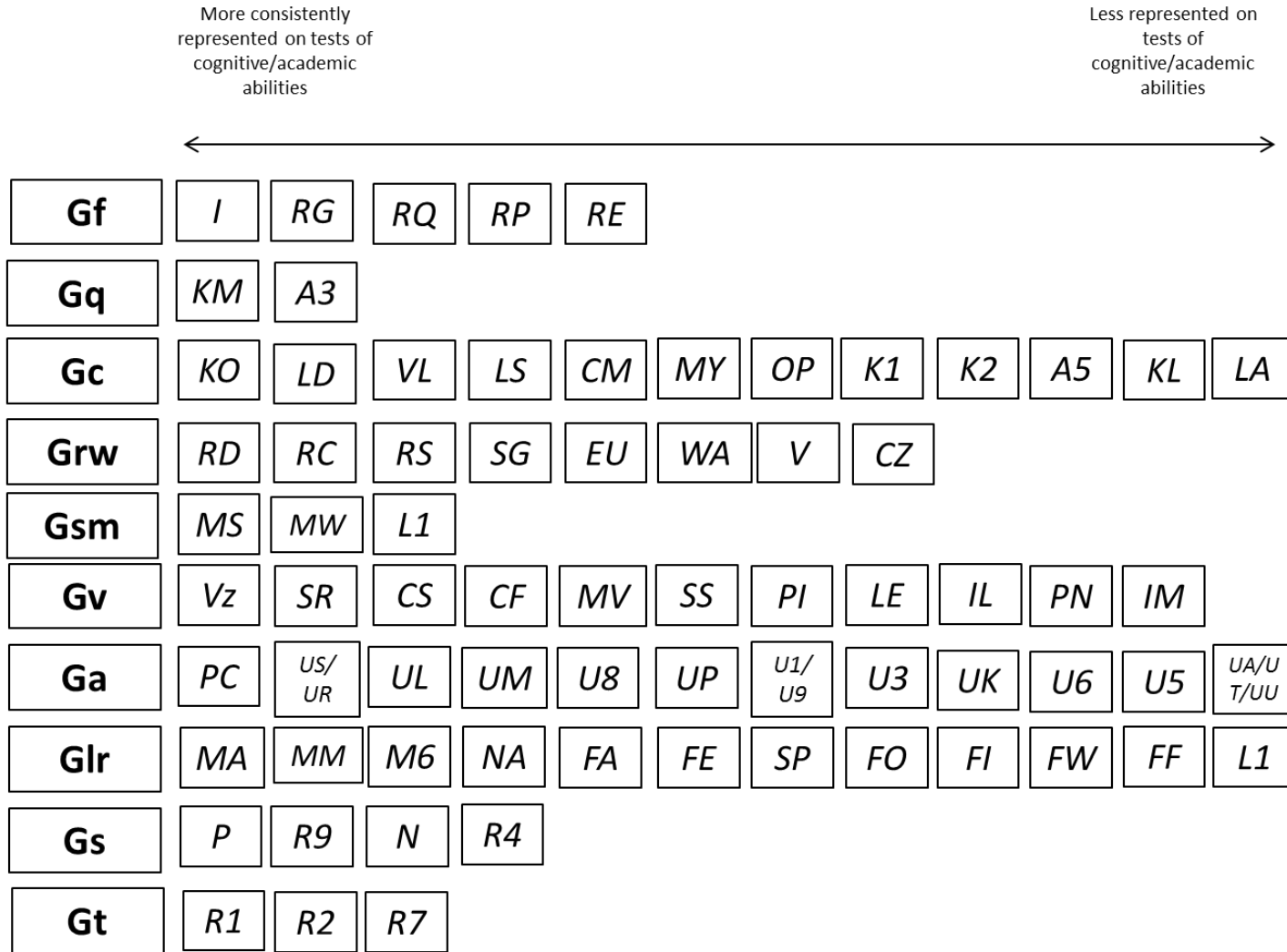
Carroll established three levels of intelligence and defined the lowest level as being specific to performance of a particular elementary cognitive task. He found all of these specific factors to be correlated, but argued that they should be represented separately as the second order abilities seen here, for example ‘learning and memory.’ Alongside these, he represented two abstract factors of fluid and crystallised intelligence which give aspects of dealing with novel situations and applying learnt skills, respectively. Carroll found all the general levels of ability to be correlated as well, leading to the common and widely accepted measure of ‘g’. Cattell-Horn’s Gf-Gc concepts are included and elaborated upon in the above model.

Whilst the Cattell-Horn and Carroll theories separately are indeed similar, there are key differences. For example, Carroll assumes a general level of intelligence (‘g’) whereas Cattell-Horn does not. Carroll includes the abilities of ‘quantitative reasoning’ and ‘reading/writing ability’ within the fluid and crystallised intelligence concepts respectively, whereas for Cattell-Horn these abilities are distinct as themselves. Cattell-Horn theory also separates the short term memory ability from other memory abilities, whereas Carroll does not. In an attempt to marry these noted differences McGrew (1997) produced an integrated theory that became known as the CHC model.

The CHC model presents ten broad cognitive factors with more than 70 narrow abilities, without the presence of ‘g’. As noted by Kaufman, Kaufman & Plucker in the Oxford Handbook of Cognitive Psychology (2013) this model has been influential in the development of IQ tests, for example, the 5th edition Stanford-Binet (Roid, 2003), the Kaufman Assessment Battery for Children, 2nd edition (KABC-II: Kaufman & Kaufman, 2004) and the Woodcock-Johnson, 3rd edition (WJ III: Woodcock, McGrew & Mather, 2001).

A model of the Cattell-Horn-Carroll theory of cognitive abilities guiding intelligence testing can be seen below in Figure 2.

Figure 2. CHC Theory model of Cognitive Abilities (from Flanagan et al., 2000 and McGrew (1997); cited in Flanagan, Ortiz & Alfonso (2013).



(*Broad abilities shown in bold. Narrow abilities shown in italics).

The twelve broad abilities noted are fluid reasoning (Gf), quantitative knowledge (Gq), crystallised intelligence (Gc), reading and writing (Grw), short-term memory (Gsm), visual processing (Gv), auditory processing (Ga), long-term memory retrieval (Glr), processing speed (Gs) and reaction and decision speed (Gt). The narrow abilities can be seen listed below the broad abilities in Table 1. Those abilities shown in *italic* are noted as more commonly assessed in cognitive and academic intelligence tests than those shown in standard text.

Table 1. Broad and narrow abilities defined from CHC theory.

Broad ability	Narrow abilities included
Fluid reasoning:	<i>Induction (I)</i> <i>General sequential reasoning (RG)</i> <i>Quantitative reasoning (RQ)</i> Piagetian reasoning (RP) Reasoning speed (RE)
Quantitative knowledge:	<i>Mathematical knowledge (KM)</i> <i>Mathematical reasoning (A3)</i>
Crystallised intelligence:	<i>General verbal information (KO)</i> <i>Language development (LD)</i> <i>Lexical knowledge (VL)</i> <i>Listening ability (LS)</i> <i>Communication ability (CM)</i> <i>Grammatical sensitivity (MY)</i> Oral production and fluency (OP) General science information (K1) Information about culture (K2) Geography achievement (A5) Foreign language proficiency (KL) Foreign language aptitude (LA)
Reading and writing:	<i>Reading decoding (RD)</i>

	<i>Reading comprehension (RC)</i>
	<i>Reading speed (RS)</i>
	<i>Spelling ability (SG)</i>
	<i>English usage knowledge (EU)</i>
	<i>Writing ability (WA)</i>
	<i>Verbal language comprehension (V)</i>
	<i>Cloze ability (CZ)</i>
Short-term memory:	<i>Memory span (MS)</i>
	<i>Working memory (MW)</i>
	<i>Learning abilities (L1)</i>
Visual processing:	<i>Visualization (Vz)</i>
	<i>Speeded rotation (SR)</i>
	<i>Closure speed (CS)</i>
	<i>Flexibility of closure (CF)</i>
	<i>Visual memory (MV)</i>
	<i>Spatial scanning (SS)</i>
	<i>Serial perceptual integration (PI)</i>
	<i>Length estimation (LE)</i>
	<i>Perceptual illusions (IL)</i>
	<i>Perceptual animations (PN)</i>
	<i>Imagery (IM)</i>
Auditory processing:	<i>Phonetic coding (PC)</i>
	<i>Speech sound discrimination (US)</i>
	<i>Resistance to auditory stimulus discrimination (UR)</i>
	<i>Sound localization (UL)</i>
	<i>Memory for sound patterns (UM)</i>
	<i>Maintaining and judging rhythm (U8)</i>
	<i>Absolute pitch (UP)</i>
	<i>Musical discrimination and judgment (U1/U9)</i>
	<i>General sound discrimination (U3)</i>

	Temporal tracking (UK)
	Sound intensity discrimination (U6)
	Sound-frequency discrimination (U5)
	Hearing and speech threshold (UA/UT/UU)
Long term memory retrieval:	<i>Associative memory (MA)</i>
	<i>Meaningful memory (MM)</i>
	<i>Free recall memory (M6)</i>
	<i>Naming facility (NA)</i>
	<i>Associational fluency (FA)</i>
	<i>Expressional fluency (FE)</i>
	<i>Sensitivity to problem (SP)</i>
	<i>Originality/creativity (FO)</i>
	<i>Ideational fluency (FI)</i>
	<i>Word fluency (FW)</i>
	<i>Figural fluency (FF)</i>
	Learning abilities (L1)
Processing speed:	<i>Perceptual speed (P)</i>
	<i>Rate of test-taking (R9)</i>
	<i>Number facility (N)</i>
	Semantic processing speed (R4)
Decision speed:	<i>Simple reaction time (R1)</i>
	<i>Choice reaction time (R2)</i>
	<i>Mental comparison speed (R7)</i>

The refinement of intelligence theory has come about through researchers appreciating the importance of the psychological theory and practical assessment working in harmony. Furthermore, researchers now consider it is important to ensure assessment rigorously avoids construct overrepresentation, construct underrepresentation and construct irrelevance variance. New methods of assessment such as cross battery assessment based upon contemporary theories of intelligence allow for systematic, reliable and theory-based

interpretations of ability across cognitive, achievement and neuropsychological assessments (Flanagan, Ortiz & Alfonso, 2013).

In fact, in 2012, Schneider and McGrew reviewed the CHC model and made various omissions and alterations. Each of the ten broad factors were altered, apart from Gq; quantitative knowledge, (i.e. the narrow abilities involved in these broad factors were reduced or refined). Furthermore, they included six more broad abilities which are psychomotor speed (Gps), domain specific knowledge (Gkn), olfactory abilities (Go), tactile abilities (Gh), kinaesthetic abilities (Gk) and psychomotor abilities (Gp). Although these changes do not inform this thesis due to the timing of writing and publication of these improvements, they are noted here for completeness in the literature discussion of intelligence.

The CHC theory appears to be the most extensively developed and consequently is the most widely used intelligence model in IQ testing (Kaufman, Kaufman & Plucker, 2013). Kaufman, Kaufman & Plucker (2013) also note that assessments using this theory are incorporating the latest research on cognitive mechanisms relating to 'g'. However, it is relevant to note the other stances on intelligence that are present in the literature for completeness of this literature review.

With regards to neuropsychological modelling of intelligence, Luria (1966; 1970; 1973) considered processing of information as a three block model. The first block targets focused attention and sustaining this, whilst the second receives the information and stores it, allowing for the third to use all information together to give either a complete view of the picture/information (simultaneous processing) or a fragmented view of individual parts (successive processing). This neuropsychological model puts emphasis on cognitive processing as opposed to the elements of intelligence itself. Nonetheless, it has been used in the development of various cognitive tests more recently (e.g. the Cognitive Assessment System, Naglieri & Das, 1997).

With regards to incorporating further intelligences and abilities into the typical intelligence view, another theory comes from Gardner's Theory of Multiple Intelligences (1983) which establishes eight different intelligences; linguistic, logical-mathematical, spatial, bodily-kinaesthetic, music, intrapersonal, interpersonal and naturalistic. Gardner noted an over emphasis, that may still be very present in education systems nowadays, upon mathematical and linguistic knowledge (Gardner, 1993). Due to various issues (for example, philosophical, conceptual and cognitive) and the varying validity of tests associated with the theory, the idea

has come under some criticism (e.g. White, 2008; Lohman, 2001; Plucker, 2000). Gardner has consistently defended his theory (1995, 1998) holding that the misapplication of a theory is not evidence enough to deem a theory weak.

Furthermore, to note another aspect of intelligence, since the 1990's, emotional intelligence has become a concept used in personnel selection and cognitive testing. Emotional intelligence (EI) theories are based upon the observation that individual differences exist in the extent to which individuals can reason about and use emotions to enhance thought (Salovey & Mayer, 1990). Mayer, Salovey & Caruso (2008) have offered the most influential model of EI in a four branch model explaining EI as involving the ability to: (a) perceive emotions in oneself and others accurately, (b) use emotions to facilitate thinking, (c) understand emotions, emotional language, and the signals conveyed by emotions, and (d) manage emotions so as to attain specific goals (cited in Kaufman, Kaufman & Plucker, 2013). Whilst in assessments, relationships with important outcomes such as social competence and quality of relationships hold after controlling for general cognitive ability and personality, the theory has received criticism over predictive validity and whether the knowledge of emotions is utilised as well as present (e.g. Brody, 2004; Oatley, 2004; Zeidner, Roberts & Matthews, 2004).

Finally, with regards to involving neuroscience and cognition research into intelligence modelling, Anderson (1992, 1995) developed the theory of minimal cognitive architecture. Knowledge is held in this theory to be acquired through two different routes; (a) a thoughtful problem solving route consisting of equal yet unique verbal and spatial processors, and (b) a route involving information processing modules that consists of modules that develop over ones lifespan and include abilities such as perception of three dimensional space, phonological encoding and theory of mind. Anderson (2008) holds that the theory can explain the discrepancy between IQ and extraordinary abilities and developmental disorders and average/high IQ presence.

Following the overview of the intelligence literature, it is now necessary to introduce personnel selection within HRM in a high stakes testing setting. The SJT is used as a selection method within high stakes testing in personnel selection and hence this following section will provide a background to high stakes testing and offer a foundation to then introduce and frame the specific method of the SJT and the medical field that this thesis focuses upon.

High Stakes Testing in Personnel Selection

High stakes testing is where a selection test results play a critical role for individuals in gaining access to employment or education (Sackett, Schmitt, Ellingson & Kabin, 2001). Many different selection techniques may be used and these include knowledge tests, simulation tests, interviews, bio data or references. This literature review will now focus upon high stakes testing and the key selection measure of interest to this thesis that is the simulation method of an SJT

Simulation tests typically mimic psychological and/or physical aspects of a job (Lievens & De Soete, 2012). They are all designed to sample job behaviour (Motowidlo, Dunnette & Carter, 1990) and take this as a prediction that through behavioural consistency past behaviour will predict future behaviour (Wernimont & Campbell, 1968). Simulations are conceptualised as multidimensional ‘methods’ (Lievens & Se Soete, 2012) as they measure a variety of performance dimensions as opposed to more typical measures or inventories assessing cognitive ability or personality traits. The content measured in simulations can therefore vary dramatically from one simulation to another.

Simulations can also be further classified into high fidelity or low fidelity simulations. High fidelity simulations actually present a situation to an applicant in more or less ideal replicas of on-the-job-tasks (Roth, Bobko & McFarland, 2005) and therefore require an actual behavioural response (Thornton & Rupp, 2006). For example, an assessment centre where role play is involved and applicants undertake exercises such as in-tray tasks. There is extensive support for their validity (Arthur, Day, McNelly & Edens, 2003).

Motowidlo, Dunnette & Carter (1990) introduced the term ‘low fidelity simulations’ and these have become a popular alternative to other simulation tests. These low fidelity simulations are less realistic but aim to still capture the behavioural tendencies of an applicant. SJTs are low fidelity simulations as they use less realistic stimulus materials and responses are less exact approximations of an actual behavioural job response (Motowidlo, Dunnette & Carter, 1990). Important research therefore lies around how much fidelity can be lost from a simulation before it becomes a useless measure. In fact, their criterion validity appears to be unaffected and in research have proved a useful alternative to expensive and time consuming high fidelity simulations (e.g. Christian, Edwards & Bradley, 2010).

Simulation methods, such as the SJT, can be used across many fields and for many career gateways or assessments. This thesis focuses upon the SJT within a medical career

framework in particular, and therefore this literature review will now discuss the importance of the selection processes within medicine and identifying individual differences in medics, the structure of the training and selection within medicine, and the methods used within the field for these purposes.

Following this broad discussion of personnel selection, the literature concerning the SJT and medical selection that is more specific to this thesis is described in chapter 3 below. Chapter 3 begins by discussing the SJT in general terms before relating it to the medical field.

CHAPTER 3: SITUATIONAL JUDGMENT TEST & MEDICINE

LITERATURE REVIEW

Situational Judgment Tests

This literature review has already noted that one of the major issues within Human Resource development and personnel selection methods is the ability to identify the strengths and weaknesses of an applicant in terms of job requirements, i.e. tap into individual differences in ability, and especially cognitive ability. This has been shown to be increasingly true and of utmost importance when looking at medical selection and assessment in particular.

Personnel selection within the medical field has introduced and focused upon the low fidelity assessment method of an SJT for an element of late and post- University selection procedures (Foundation and speciality training as noted). This SJT method will now be described and discussed in more detail.

An SJT is a test that provides a hypothetical scenario, asking the applicant to rate answers/choose one answer from multiple options. SJTs are low fidelity simulations used in personnel selection for many fields. They can typically be described as an explanation of a hypothetical work related scenario that then asks applicants to rate a number of responses upon how likely they are to respond in such a way and which are the most effective responses in the situation described. Responses are either in a ranking format (as the example below) or a forced option format (i.e. choose one answer from the options). Applicant responses are usually marked in comparison with those responses of experts within the specific field and a scoring key is developed prior to testing.

An example of a written SJT question is shown below:

- One of your colleagues has gone on holiday and his father, who has become ill, has just been admitted to your hospital. Your colleague phones you from his holiday and asks if you can find out more about what is happening with his father.
Rank the following actions, using 1 for the action that you would most likely take, etc. down to 5 for the action that you are least likely to take:
A. You simply refuse to answer your colleague's question
B. You speak to your consultant, explain the situation and ask for the information that your colleague wants

- C. You politely tell your colleague that you cannot do what he is asking you to do because it is not Trust policy
- D. You ask the father if he gives permission for you to ask the consultant about him
- E. You retrieve the fathers notes and give the information to the colleague

(<http://www.medical-interviews.co.uk/example-sjt.aspx>, Jan 2011).

The SJTs are thought to predict job behaviour as based upon behavioural consistency ideas i.e. past behaviour is the best predictor of future behaviours. Whereas high fidelity simulations allow for an actual scenario and real demonstration of procedural knowledge and skills, low fidelity simulations, such as the SJT, do not. Low fidelity simulations assess responses regarding procedural knowledge and skills, and these are seen as precursors for effective and later job behaviour (Motowidlo, Hooper & Jackson, 2006) in that they capture behavioural intentions rather than actual behaviour. In high fidelity simulations, the behavioural consistency ideas mentioned before are much more straightforward and theoretically apparent (Alon et al., 2009). Hence, although an applicant may know the best course of behaviour and have an intention to do this (as shown in an SJT response) they may respond differently when they actually have to perform the behaviour (a high fidelity simulation could show this). Although, it should be noted there is extensive social psychology research endorsing ideas that behavioural intentions make that intention much more likely to happen (Ajzen & Fishbein, 1980).

There are distinctions made between handwritten SJTs and those that are video-based or multimedia SJTs (Lievens, Peeters & Schollaert, 2008). For full clarity therefore, the SJTs are low fidelity simulations that can also differ in the fidelity of their own stimulus. The written tests have low stimulus fidelity, and the video based scenario and multimedia SJTs having a higher fidelity. For the latter, questions and responses following the video material are presented visually with an accompanying narration. The portrayal of information is therefore richer and more realistic; hence their stimulus fidelity and their validity are increased (Funke and Schuler, 1998). Such video- based or multimedia SJTs also hold the potential to be 'branched' (also referred to as nested or interactive; Olson-Buchanan et al., 1998). They can be programmed to allow for the next question choice to depend upon a previous questions answer; hence, they are standardised, yet dynamic and allow for a closer representation and mimicker of real life encounters and situations (Lievens, Peeters and

Schollaert, 2008). Medical selection SJTs are typically written stimuli and written responses, either handwritten or using a computer.

To summarise, the SJT is a potential measure that can assess domain independent and important aspects of professional behaviour, as noted above when considering the assessment of competency as opposed to what is assumed to be typically intelligence. This will be further examined following discussion of SJT development and the history of the SJT.

Following this description of the SJT, the general development and background, current modelling and the theory behind the SJT are discussed along with comparisons of the measure against others selection measures. Then, more specifically, the SJT in medicine, and the criterion and construct validity of the method are discussed to give a complete picture of the measure and surrounding ideas.

SJT Development

There are two main ways to consider the development of an SJT. Firstly, a group of experts (relevant to the situations/job role in question) are used to produce examples of exceptionally good or bad behaviour at work; 'critical incidents' and events (Flanagan, 1954). This can be done from memory, job visiting or archival events (e.g. Hunter, 2003). This method is referred to as a 'critical incident method'. Secondly, SJT items may be written to reflect items of an underlying model, e.g. a list of competencies, literature review or attributes identified from job analysis (seen in Stevens and Campion, 1994; 1999; Motowidlo, Dunnette and Carter, 1990). This is referred to as a 'model based method'.

Critical incident methods provide a rich source of information, but can be time consuming and expensive to create (Weekley, Ployhart and Holtz, 2006). Model based methods seem promising and may overcome the time and expense issues, although they may lack theory about work situations in particular, i.e. the situational element of an SJT (Weekley, Ployhart and Holtz, 2006). This debate is still open and very much involved in current day research topics. The method of writing itself may be seen as potential moderator for the effects of SJT upon job performance, or a mediator for possible antecedent variables in predicting performance upon the SJT. Typically, a question stem (i.e. the scenario) is written or chosen, and then the process of writing answers/options follows on. The complexity and fidelity of such an SJT stem can also differ from measure to measure, in terms of difficulty, detail, realism and specificity (McDaniel and Nguyen, 2001).

These developed SJT stems are then grouped and reduced/extended in length and matter to give similarly constructed and built scenarios. Secondly, another group of experts are asked to suggest one or more responses to these given scenarios. These will then make up the response options for the applicants. These experts should be skilled enough to give correct and logical paths of action as well as being able to provide less sensible and less effective behaviours. These experts are often referred to as SMEs (Subject Matter Experts). The response literature is a popular area for research in the field and concerns various dimensions; for example, the origin of the response options, construct-based response options, complexity of response options, fidelity of response options and 'fakability' in the effectiveness of the options presented (Weekley, Ployhart and Holtz, 2006). Furthermore, the response instructions literature is extensive (McDaniel, Hartman and Grubb, 2003; McDaniel and Nguyen, 2001; Nguyen, Biderman and McDaniel, 2005) although still requires more work to establish how response instructions affect differing SJTs.

In most cases, experts are asked to generate alternative answers to a situation, although in some situations the SJT stem developers may write most of the options (e.g. Stevens and Campion, 1999). Answers generated by experts are more likely to be more realistic and numerous in options (e.g. Wagner and Sternberg, 1985). Due to the nature of an SJT and the tendency to produce just one score for the test it is likely that responses will target many constructs giving the multidimensionality of the scale (Motowidlo et al., 2006). Furthermore, options can range in terms of fidelity (e.g. video based or written) and difficulty.

Response instructions are generally classified into either behavioural tendency instructions (what one *would* do) or knowledge based instructions (what one *should* do) (Nguyen et al., 2003). In 2003, Ployhart and Ehrhart examined the reliability and validity of SJT items with regards to their instructions. They found that behavioural instructions were viewed more favourably. Response instructions have been found to impact the constructs measured by the SJT (McDaniel et al., 2003). Knowledge based instructions were more highly correlated with cognitive ability, when compared to behaviour based instructions that correlated more highly with personality constructs. This conclusion was based upon a meta-analysis of 41 knowledge instruction vs. 21 behavioural instruction SJTs. Both types of instructions are equally associated with measures of job performance (.26; McDaniel, Hartman, Whetzel & Grubb, 2007).

Finally, the scoring key is decided upon by subject experts from ranking the answers and scoring the 'best' options with the higher scores and the 'worst' options with the lower scores. This is the typical foundation of scoring for most SJTs (Weekley, Ployhart & Holtz, 2006) although there are alternative options, including scoring keys based upon empirical keying or theory (Krokos, Meade, Cantwell, Pond & Wilson, 2004; MacLane, Barton, Holloway-Lundy & Nickles, 2001; Paullin and Hanson, 2001).

History of the SJT

The early beginnings of the SJT are rooted in civil service examinations in the United States of America. For example, DuBois (1970) asked questions of a situational nature targeting which actions applicants would take after describing a brief scenario when testing for an Examiner of Trade Marks (Patent Office) role. The Binet (1905) scale was later used and this included abstract questions asking what one 'ought' to do in a described situation. These stems of the SJT may in fact be interpreted as a situational interview (Latham & Saari, 1984). This method is similar in form to that of an SJT and requires an interviewer to describe job-related scenarios and then rank the applicant responses. It may loosely be described as a verbal version of the modern and typically written SJT. However, it does come close to matching the SJT in terms of form (Weekley & Gier, 1987) and validity (McDaniel, Whetzel, Schmidt & Maurer, 1994).

Weekley and Ployhart (2006) note that these early assessments did not include ways of handling or interpreting the answers, hence, interpretation and marking was subjective and dependent upon the marker/examiner. McDaniel (2001) identifies the first widely used SJT with closed-ended response options as the George Washington Social Intelligence Test (Moss, 1926). From this point, development is seen through the World War II with little reported about the success of these attempts to measure judgment (Northrop, 1989) before focus moved onto 'supervisory and managerial potential' from approximately 1940 through to the 1980's (McDaniel et al., 2001).

Research upon the SJT has been reported as relatively sparse from this time until the re-introduction of the ideas surrounding the SJT in the late 1990's. The concept of 'tacit knowledge' from Sternberg and colleagues (Sternberg, Wagner & Okagaki, 1993; Wagner, 1987; Wagner & Sternberg, 1985) and 'low fidelity simulations' by Motowidlo, Dunnette and Carter, 1990) reintroduced the idea of the SJT as a tool that could have real use and promise for the future in the HRM and psychological literature. The main conclusions drawn about

the early methods from before this period, are that they tended to measure a form of ‘g’ (general intelligence) regardless of the fact they were designed to measure judgment as a concept (for example, Carrington, 1949, Millard 1952; Thorndike and Stein, 1937).

Weekley & Ployhart (2006) reported that the number of studies published on the topic of the SJT doubled from 1999 to 2004. Furthermore, they justify the increased interest in this selection and assessment method as being due to its promise and positive features, such as the unique capturing of the measure and the measure giving incremental validity over cognitive ability and personality measures (for example, Clevenger, Pereira, Weichmann, Schmitt & Schmidt-Harvey, 2001; Weekley and Ployhart, 2005).

Weekley & Ployhart (2006) also note the extensive development of the SJT and the research surrounding it. However, they also pinpoint the problem of empirical research moving ahead of theory development. There is much research surrounding performance upon the SJT and the breadth of different topics investigated. However, there are theoretical gaps in the explanation of what underlies the SJT and a lack of ‘identification of an underlying theme or convergence towards a coherent theory’ (Weekley & Ployhart, 2006, p6).

In relation to medicine in practice, much work has been done to encourage further understanding of the patient as a person and to see care from a patient’s perspective (Tuckett, Boulton, Olson & Williams, 1985). The focus upon human encounters and the importance placed upon general professional and interpersonal behaviour by doctors that the SJT brings is a welcome addition to the medical field. It may discourage the focus on the disease instead of the patient as a human, reduce the complexity of human encounters and promote the importance of the patient’s values, opinions and experiences. These ideas are very much in line with an SJT assessment that deals with hypothetical encounters and situations in the workplace.

The SJT: Theoretical Ideas and Elements

The SJT will now be discussed primarily from a social and cognitive psychology point of view. This discussion of the SJT will be broken down into two sections covering firstly literature relating to the ‘situational’ element and, secondly, the ‘judgment’ element of the SJT. Following this, the most recent theoretical developments from cognitive psychology will be discussed allowing for a complete picture of the SJT from all strands of psychology relevant to this thesis and the studies included (organisational, social and cognitive

psychology). Where relevant, the psychological literature and theory is related to an SJT in medical selection/ careers in particular.

The SJT: The situational element

Social psychology can be described as the scientific study of the behaviour of human thoughts, feelings and behaviours as affected by the actual, imagined or implied presence of others (Allport, 1954). Within this there is clearly an important aspect to consider, namely the situation that an individual finds themselves in, i.e. assessment of the circumstances and how one makes sense of these circumstances. There is extensive and very well publicised research into the importance and effect of a situation upon human behaviours, for example, Milgram's obedience study (1963) and Zimbardo's prison study (1974) which both infamously found participants acting in extreme, out of character and sometimes unpredictable ways due to the potency of the influences and situations they found themselves in.

Heider's (1958) theory of attribution dynamics is a key idea of how people make sense or meaning out of their situation. The theory implies that humans act as detectives or amateur researchers who identify an 'actor' in a situation and label themselves as an 'observer' of this person. They then take steps to identify what is driving the behaviour of the actor (in the situation, and personal forces) as this behaviour may result in extreme or important changes in the situation. This acquired knowledge allows for predictability to the situation and therefore a tactical advantage.

Gessner & Klimoski (2006) note that in reality, people do not seem to expend such effort or act like a 'detective' as described. Furthermore, this only appears to happen if the situation is seen as important, novel, or even dangerous. The SJT may in fact fit into one, some or all of these cases. The main gateways where the SJT is used for medical selection are during the selection procedure for medical school, progression onto Foundation training or specialisation thereafter. These in fact are extremely important to an applicant and lead to a great deal of pressure with large consequences for error or poor performance. Furthermore the test material is always novel due to the nationally run nature of the assessment or selection procedure, even if the layout and format of the test are somewhat consistent. Lastly, the SJT may in some cases, be viewed as dangerous due to the severe consequences for career progression or job success that can follow from poor performance upon the SJT. This may be a different type of 'danger' to that felt when one's life may be at risk, but it may still

be felt. This is likely to be subjective and relative to the individual and the importance based upon the test itself.

More recently, Fiske & Taylor (1991) have developed a more detailed approach of how individuals make sense of their situation. They hold that actors have automatic or controlled mental processes and that these are dependent upon the situation. Generally, an individual uses automatic responses and this is a way to deal with everyday life. However, if the situation is perceived as novel, important or dangerous, then more mental effort is used to develop plans, strategy and awareness of the situation in hand. This allows the individual to interpret the situation correctly and act accordingly. This is done through an individual's perceptions and actions, along with one's memory and learnt experience, building up filters such as schemas, scripts and heuristics, which have influences over one's behaviour (Kunda, 1999).

Gessner & Klimoski (2006) note two further important theories when considering sense-making. The first of these interaction theories is based upon Barker's (1963) ecological perspective and the modification of this to studying 'group processes' by McGrath (1984). Barker's perspective developed the important concept of 'behaviour setting'. This concept involves a 'standing pattern of behaviour where the behaviour is surrounded by and congruent with the environment' (Gessner & Klimoski, 2006, p18). i.e. the behaviour setting is a product of the personal qualities and the environment. This is held when it is observed that different individuals perform more similarly in the same situation than one individual performs across situations (Barker & Wright, 1954).

To be effective it is necessary for the SJT, as a selection method in high stakes testing settings to identify variability and individual differences between applicants, even though the external situation itself may predict that individuals will perform more similarly as they are in the same situation. It follows that there are different strengths of a situation, and that as these differ the amount that a method reveals about an individual vs. the situation itself will also relatively differ. Gessner & Klimoski (2006) suggest that this situation strength may be a result of the cueing properties of a situation (i.e. cues that elicit experiences from memory, such as saying hello to a colleague), the anticipation of rewards (i.e. working for a common or personal gain), social pressure (e.g. the anticipation of review from others produces a homogenizing effect; Frink & Klimoski, 1998) and/or the ambiguity of the situation (i.e.

people are more likely to conform or look to others for help in an ambiguous situation; Moscovici & Zavalloni, 1969).

McGrath (1984) built upon Barker's ideas and conceptualised 'group interaction' concerning the situation both before the interaction and after, and the forces of the situation that shape communication within group members and their behaviour. Hence, interactions are constantly being reshaped due to the constantly changing environment. Gessner & Klimoski relate this model to the SJT and see the SJT as a replacement for the key 'actor group' in the McGrath model. The objective environmental influences, such as the task environment, individual differences and the organisation, feed into the behaviour setting (i.e. the expectations of the group about the individuals and the characteristics of the task in hand) through the objective task itself and the perceived pattern of relationships between the group. The behaviour setting then leads into the SJT itself and shows how these interactions may affect performance upon what appears to be an objective task.

The second important theory of sense making with regards to a situation identified is that of Rosenthal & Jacobsen (1966). This theory targets the situation whilst actually undertaking the SJT, as opposed to the build up to and around it as discussed in the earlier model by McGrath (1984).

Rosenthal & Jacobsen noted that there are influences from an experimenter or researcher; experimenter effects which concern the participants characteristics relevant to the experiment, and experimenter expectancies, which concern the desired outcome for the experiment. It is held that both of these affect the researcher/ experimenter and the participants. This is therefore important to note with regards to how the SJT is constructed and the participation run, from an administrator/ organiser as the experimenter or researcher, and from an SJT developer view.

Hence, social psychology puts emphasis on the situation before and during the SJT, from both a personal, stakeholder and group stance. Performance can be held to be shaped by the situation and the interactions with this and others around the applicant. Hence, the objectification of the situation and test is of utmost importance. There is also potential for the situation to be used as a sense-making one for applicants, as opposed to a purely uniform experience and set up. The SJT is viewed as encompassed within the behaviour setting and is a result and product, and therefore intertwined, with the situation itself. Determinants affecting performance upon the SJT may then be held as past experience and personality.

Sense making is a product of prior experience of similar and different situations, and the motivations and influences over behaviour from stable personality traits and self-regulation.

This fits logically into the noted model of McDaniel et al., (2006) (see p.73 for model diagram) which holds personality traits (and cognitive ability) as constructs affecting performance upon the SJT through experiences and education (different situations). Hence, one's sense-making ability could be added into the model as a factor that the personality traits and cognitive ability may or may not go through to reach the education/training and experience influences. The sense-making ability is affected by the four main constructs and hence may influence an indirect route of affect upon SJT performance for these traits.

The 'judgment' element of the SJT will now be discussed following this section covering the 'situational' element of the SJT.

The SJT: The judgment element

Hastie & Dawes (2001) define judgment as 'the human ability to infer, estimate and predict the character of events' (p.48). The SJT clearly asks applicants to make a 'judgment' on a described situation and to follow through using their explicit or implicit abilities to pick a suitable answer/rank a set of answers in order of likelihood of action.

Judgment and decision making (JDM) theory is thought to have stemmed from Egon Brunswik's (1952) thoughts on perception. Hammond (1955) interpreted this work and established the importance of perception when considering judgment, i.e. the interpretation and inferences made from incomplete, partially available and fallible clues in an individual's situation. As noted in Heider (1958), people are seen to act as detectives in their environment in order to acknowledge what is happening and why it is the case. People use judgment cues (i.e. what they take in from the world around them) to then overcome indecision (Hogarth, 2001) and reach conclusions.

Hammond's (1955) work went on to establish analytical and intuitive decisions upon a continuum as opposed to earlier ideas that saw these extremes as a dichotomy. He developed the idea of 'quasi-rationality' that holds decisions as commonly using both elements of the logical and methodical analytical thought and thought that has no such step-by-step guideline or explanation behind it. Hammond acknowledges that there are strengths and weaknesses to both extremes upon the scale and proposed that quasi-rationality overcomes these downfalls by mixing the two concepts of thought; a form of common sense.

This in fact is what the SJT may be measuring; it is a method requiring a judgment of quasi-rationality common sense. Brooks & Highhouse (2006) distinguish the SJT from other commonly used ability tests as a measurement of this concept where the other methods do not tap into the element of intuition in thought, only the analytical element. They reviewed the literature upon intuition and pinpoint Hogarth's (2001) definition of an intuitive response as something that is reached with little effort, involving little or no conscious deliberation. Hogarth also suggested that confidence and speed are also stronger and quicker respectively when judgments are intuitive. However, this does not make them correct as often further examination of choices require the acceptance of principals, such as little errors may have to be made in order to avoid larger ones and worse consequences. In fact mechanical procedures that follow logical thought procedures consistently outperform judgments (Camerer & Johnson, 1991) due to the fact of consistency based upon the formula giving consistency over time. Intuition does triumph in prediction sometimes. For example, with professional financial analysts vs novices (Johnson, 1988) and particularly in environments where one needs to be able to predict change (e.g. Blattberg & Hoch, 1990, Whitecotton, Sanders & Norris, 1998). Medicine is indeed one of these latter situations and due to the ever changing and developing job role itself and the situation in hand, e.g. different patients and cases with different symptoms, where the monitoring of environment and case changes is imperative to protect, help and prevent potentially fatal consequences in the long term, could possibly be noted as one of the most important areas in which to be aware of change and development.

In fact, the use of good judgment as the ability to go beyond the information given and to rely on one's broader knowledge and past experiences (Funder, 1987) may be what is important for SJTs designed for medical selection. This may underlie good performance upon the SJT, and as a medical professional this may be a key basis for successful, safe and logical practice. There may be importance placed upon the decision of whether to rely upon intuition or analytical thought. This may be an aspect of good judgment, and research also suggests that those who rely upon a deliberate or methodical reasoning can make poorer decisions than those who trust their intuition (Wilson & Schooler, 1991). For example, decision makers who analyse reasons for consumer choices make worse decisions than those who do not (McMakin & Slovic, 2000). The reliance upon intuition, along with logical and analytical thought, can be summarised into a good judgment foundation in that all aspects should be considered and combined, and the extent of each elements pull or strength is relative to the

situation and environment in hand. Much as described previously the situation plays a part as well as the internal logic of an individual.

The scoring of such judgment can be classified into two categories (Funder, 1987). Firstly, agreements with others, as is used widely in the marking of SJTs against expert panel ratings, and secondly, the ability to predict behaviour.

Brooks & Highhouse (2006) report one study using the ability to predict behaviour as a scoring method, in Zalesny & Highhouse (1992). Student teachers had to predict the actions of a teacher after watching videotapes of an unruly pupil, and the student teachers rated the target teacher's performance upon a range of dimensions. The predictions about the target teacher's behaviour significantly correlated with judgemental accuracy (i.e. measure of agreement with expert performance ratings). In relation to using such a method of scoring with the SJT, it would also be expected that high scorers would be better at predicting what others would do in a work situation (Brooks & Highhouse, 2006).

Brooks and Highhouse (2006) conclude their discussion upon JDM stressing the importance of its involvement in the understanding of the concept behind the SJT. They believe judgment is a key area to consider, and in fact, ironically from the inclusion within the title, may in fact be the area that needs further consideration in order to further theoretical and background understanding.

Following discussion of the current theory surrounding the situational and judgment elements of the SJT, the most recent cognitive psychology theoretical ideas will now be discussed to complete the background of the SJT and the current theory surrounding this measure.

The SJT: Cognitive Psychology Theory

Although SJTs have been in use for many decades, there has been surprisingly little systematic analysis of the underlying constructs and the theoretical terms that may be used to model the SJT methodology. The most recent and developed efforts have come from McDaniel et al., (2006) which are noted later in this literature review in the discussion of construct validity. This theory sees three personality traits (conscientiousness, agreeableness and emotional stability) and cognitive ability as the main constructs the SJT measures (a theory of construct validity of the SJT). This will be further discussed in the sections following.

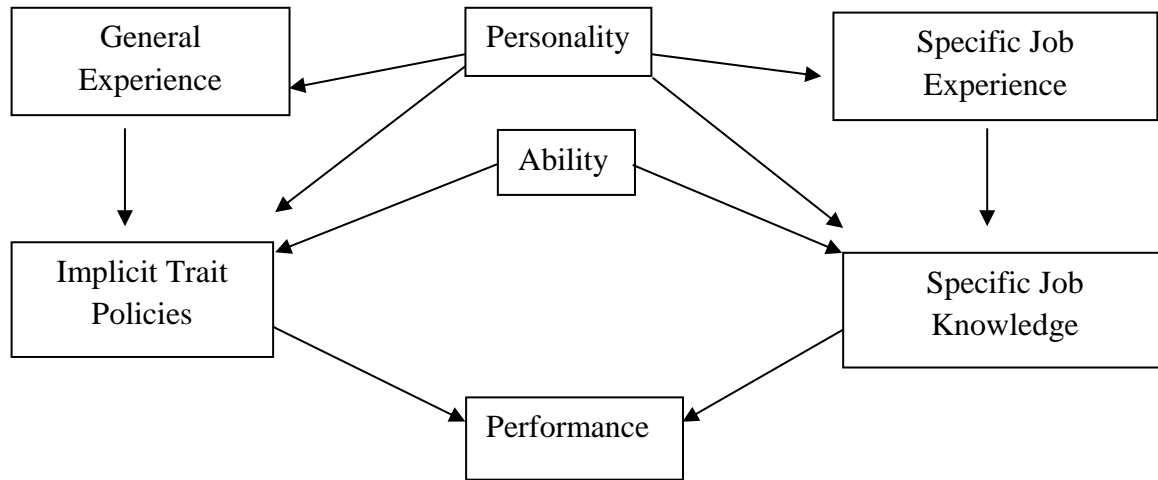
It is considered that what may be an advantage of an SJT (their versatility and construct interchangeability for example) may also be a disadvantage, in that it may lead to confusion and a lack of understanding regarding what they are actually measuring (construct validity). As noted, SJTs differ according to many factors and therefore what they measure may also differ. Theories of the SJT construct validity are described as approximations of what the SJT may measure in broad terms; specific SJTs will likely differ in the amount of correlations shown with the key constructs.

A recent meta-analysis by Christian, Edwards, and Bradley (2010) attempted to shed some light on the SJT literature and classified SJTs into different topics according to what they measured (e.g. interpersonal skills or job knowledge skills) and then considered their administration method (e.g. paper and pen or video based simulations). In general, research was empirically strong and SJTs broadly predicted job performance. However, there were only small amounts of certain categories in the comparison, such as SJTs measuring teamwork skills, showing that the adaptability of the SJT as a measure is still potentially unexploited. They also conclude that it is important to undertake appropriate job analysis and to design SJTs to test the specific corresponding constructs.

It may be seen that this is precisely the strategy that Patterson and her colleagues used to develop the SJTs for medical selection, as discussed in the earlier literature discussion (see Lievens & Patterson, 2011). The SJT showed incremental validity over knowledge tests and a correlation with actual job performance of .37 (Lievens & Patterson, 2011). Although this shows a relationship between the SJT and future job performance it appears that there is a lot more to be understood about the SJT and how this value may be improved to give a clearer picture of job performance in assessment.

In an important paper, Motowidlo and Beier (2010) took steps to tackle the issue of understanding the SJT further and produced a theoretical model of the SJT with the antecedents affecting performance upon the SJT, as shown below in Figure 3.

Figure 3. Theory of knowledge constructs underlying the SJT. (Motowidlo & Beier, 2010, p.322).



At present, in the literature there are no other psychological models beyond this and the McDaniel et al., (2006) model, that have attempted to further define the explanation of the SJT, beyond a measure of general procedural knowledge. Furthermore, even recent reviews of the literature surrounding SJTs (e.g. Campion, Ployhart & Mackenzie, 2014) are calling for theory of the SJT to be investigated and used in the development of SJTs with a distinct lack of empirical examinations of theoretical ideas being reported in the literature.

This model describes a complex set of relationships leading into performance (notably SJT performance, however this may also be interpreted as actual job performance dependent upon outcome measures in a study). ‘Personality’ encompasses the traits held by an individual, and there is a distinction drawn here between the holding of a trait, and the expression of this trait, for example, personality traits may include agreeableness, extraversion or conscientiousness. An individual’s personality profile has a causal effect upon their specific job experience (i.e. past experience of specific job related scenarios), general experience (i.e. life experience from people and situations that are not job related) and specific job knowledge (i.e. fine grained knowledge that has come from job experience or an individual’s cognitive ability) as well as their Implicit Trait Policies (ITPs).

Implicit Trait Policies are described as implicit beliefs about causal relations between personality traits and behaviour (Motowidlo, Hooper & Jackson, 2006). They are the awareness of how effective or ineffective it would be in a certain situation to express a personality trait. Implicit Trait Policies (ITPs) is now a term commonly used within SJT

related literature although it is still relatively new. ITPs appear to capture a dimension of procedural knowledge that is not based upon specific job experience, i.e. measuring them can give us evidence of procedural knowledge that can be gained in many different situations and theoretically by anybody; they have wide application possibilities.

ITPs can work in conjunction with specific job knowledge although they do have an independent effect upon SJT performance, for example, one may have an ITP for conscientiousness and so believe being conscientious in situations is the correct way to behave. However, this ITP's expression may be altered by the influence of knowledge learnt from specific job experience i.e. in some situations, being conscientious may not be the best or most effective way to behave. Hence, the concepts are very much intertwined. This link also allows for ITPs to compensate for a lack of specific job knowledge, for example, if you do not know the most effective behaviour for a situation, ones' ITP from a general experience may give you an answer. This idea also works the other way around and a lack of specific job knowledge may be compensated for by an ITP giving a reasonably effective behavioural response.

'Ability' is another key concept in the model and refers to an individual's cognitive ability, in terms of intelligence. Motowidlo & Beier (2010) explain that cognitive ability has a causal effect upon ITPs and specific job knowledge; an individual who is more intelligent will be expected to have developed more accurate ITPs and views upon trait expression than a less intelligent individual. It is also expected that more intelligent people will learn more quickly in job scenarios and therefore gain more specific job knowledge than those who are less intelligent.

Motowidlo & Beier (2010) discuss the model and develop the most current and coherent attempt at modelling the SJT in terms of cognitive psychology. They acknowledge that there are still gaps in the literature and areas for future researcher to target:

“very little is known about how people acquire procedural knowledge... that is, little is known about antecedents of the type of knowledge that is captured by SJTs.” (p.322, Motowidlo & Beier, 2010).

Furthermore, the research establishes that scores of an SJT by individuals with no specific job knowledge or experience can correlate significantly with job performance equal to the correlations for experts. They argue that there must therefore be another form of knowledge

being used. Motowidlo & Beier (2010) also argue that this knowledge is involved within the ITP concept. The ITP concept is considered in more detail below, in relation to the other concepts included in the Motowidlo & Beier (2010) model and with consideration of other psychological theories and concepts that may interact with the ITP. The ITP is the newest theoretical component to be introduced into modelling the SJT within literature and therefore requires a detailed explanation to ensure completeness in this literature review. Further research is however required into this concept and its relationships within the Motowidlo & Beier (2010) proposed model.

ITPs will be considered in relation to personality, general experiences and finally, cognitive ability.

Implicit Trait Policies

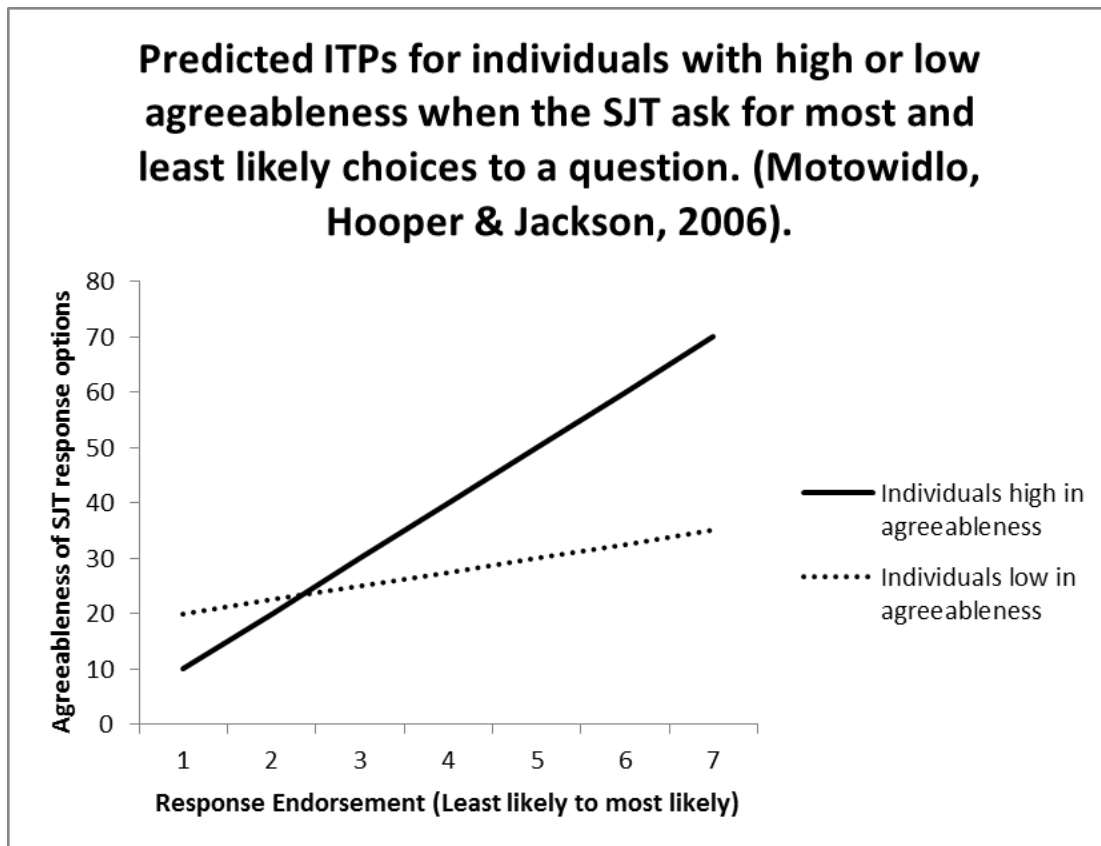
ITPs and personality

Personality traits are often correlated with SJT performance (McDaniel & Nguyen, 2001). Motowidlo & Beier created their model to explain how the expression of personality traits may be shown through SJT performance. ITPs are different from personality traits. McCrae & Costa (1996) distinguish between basic tendencies and characteristic adaptations. Basic tendencies are fundamental abilities, for example, learning abilities and personality traits, whereas characteristic adaptations are the result of basic tendencies interacting with an environment, for example, a skill, habit, preference or attitude (Motowidlo & Beier, 2010). The latter term can be used in parallel with that of ITP.

Personality is seen through the expression of ITPs in a dispositional fit pattern (Motowidlo et al., 2006). That is, people develop beliefs about the effectiveness of a behaviour/trait expression in line with their own traits. For example, if an agreeable action is the most effective behaviour in a situation, and an individual possesses the personality trait of agreeableness, the individual is more likely to see this behaviour as an effective one, and furthermore, be correct in this belief. People believe that the expression of a trait they possess is the most effective response, and if this is truly the case in reality, then this person demonstrates more knowledge about how to behave. Parallels can be drawn between this idea and Tajfel & Wilkes (1963) accentuation hypothesis in that ITP's are accentuation effects for behavioural judgements that are influenced primarily by what you think is effective and secondly, by a personality trait. Motowidlo, Hooper & Jackson (2006) provide a hypothetical model to illustrate the interaction of the ITP and personality when making choices about

actions upon an SJT. They also include how, in the absence of a personality trait to influence the ITP (for example, if an individual is not highly agreeable vs. an individual who is highly agreeable) the former group will also judge the agreeable behaviour as correct, but the reaction will be less extreme. This is shown on the graph in Figure 4 below.

Figure 4. Expected relationship between possession of a personality trait and the interaction with an ITP upon an SJT question response.



Furthermore, beyond this model of fit, which shows how an individual's own personality fits into an ITP expression, in theory there could be an ideal set of ITP's for any profession. Professional bodies can also fit into the expression of an ITP, and in turn how an individual expresses an ITP and their personality within their professional organisation. Whilst this is an idealistic suggestion due to the diversity and individuality of factors affecting ITPs there may be a basic understanding of certain ideas to do with this 'common sense' tool that can be trained or taught.

For example, in paediatrics, empathy has been shown to be an important trait for later job success (Patterson, Ferguson & Thomas, 2008). Just as core competencies can be pinpointed for all specialities, there may be core ITPs important for all professionals. Further job

analysis may reveal specific important ITPs for each speciality within medicine for example. However for cost and initial development it seems sensible to focus more generally upon 'core ITPs'. It should be noted that it is not suggested that a person's personality can be altered but merely an awareness of an important trait put into play.

Social cognitive psychology ideas may affect the expression and/or altering of ITPs and this is a direction for future research to consider when targeting the manipulation of an ITP. For example, an individual's perception of the situation and henceforth a behaviour situation (Barker, 1963), one's self-awareness (for example, private vs public self-awareness and the effects upon behaviour, Carver & Scheier, 1981) and self-efficacy beliefs (i.e. one's expectations of success in a given situation, Bandura, 1982) may also interplay with these ideas and affect the expression of an ITP, whether one chooses to alter their environment or how malleable the ITP concepts may be based around these factors.

ITPs and General Experience

Procedural knowledge is the knowledge of how to do something and knowledge that comes from experiences; yours or others (c.f. modelling theory, Bandura, 1982) and can involve both explicit and implicit memory stores/processes (Martin, Carlson & Buskist, 2007). Procedural knowledge can be assumed to be generally ubiquitous to everyday life and can include tasks such as learning to ride a bike or learning to drive a car.

Drawing upon further literature from Bandura, the expression of an ITP may be reinforced by conditioning and any rewards or positive outcomes received as a result of expressing a certain personality trait in a certain situation. Furthermore, the concept may become extinguished if they are not used regularly. The exact origin or time of establishment of an ITP is unknown and research may consider that they are formed in a potential 'critical period' such as is hypothesised about language acquisition timing (Penfield & Roberts, 1959). Furthermore, after acquisition, the effect on these ITPs and therefore the stability of these concepts is also unknown. The ever-changing cognition of a human suggests that ITPs may continuously develop and evolve and it is likely that they are affected by life events and general experiences as suggested by Motowidlo & Beier (2010).

More importantly, one should ask whether any amount of training or teaching can alter the effects of past experience. Motowidlo & Beier (2010) note the possibility of training and altering ITPs. The ability to mould, condition and alter behaviour and opinions, for example through rehabilitation processes and counselling, suggests that cognitive views and

mechanisms can be altered and manipulated. People can change and work through issues and move past bad memories or experiences that potentially cause or influence later deviant or psychologically abnormal behaviour (e.g. humanistic therapy; Rogers, 1951, and Cognitive Behavioural Therapy; Ellis, 1973). However, this may not always be the case, for example, parents are a great source of learning over the course of development and although, the correlation between parents and children of specific attitudes is only a weak one, for broader issues it is much stronger (Connell, 1972). Kasser, Koestner & Lokes (2002) also found strong links between earlier parental behaviour and the later adult values of the child. Findings like these suggest that parents can influence a child's later values and attitudes, and potentially more so for broader concepts, such as ITPs, as opposed to specific factors, e.g. personality traits. From another stance, parents may also affect their children's development of beliefs and attitudes in a similar way to how many young adults may deliberately take a point of view that is in opposition to that of their parents.

Other authority figures, for example, a teacher may also play a role in the development of an ITP. Authority figures and superiority may be a key factor in the belief that a trait expression or action is effective. This in turn may be mediated by the amount of respect felt for this authority figure. A source of authority respected by an individual may also still hold influence over an ITP at a later date. This may be relevant when considering training and those teaching the courses. A senior doctor who is respected may yield a greater influence over a trainee than a stranger. Such ideas are built upon source credibility concepts from social psychology (Greenwald, 1968).

Indeed, the whole socialisation process may influence ITP development. Interactions with others, culture, lifestyle and religion may alter how effective an individual sees certain behaviours. In turn, an individual's beliefs may alter who they spend time with, what they do and therefore the experiences they undertake. Hence, the cycle exists that ITP development is affected by personality and experience, but experience and personality are affected themselves by ITPs.

ITPs derived from strong, individual and personal values such as religion or culture may be impossible or very difficult to alter. ITPs derived from more abstract factors such as friendships, observations of others or work experiences may be less strongly embedded in a moral personal core and hence may be easier to influence.

Consequently, the measurement not just of ITP content, but also of strength may be of concern. Other learning sources in media and technology have been found to strongly influence peoples' attitudes (Goldberg & Gorn, 1974; McKay & Covell, 1997) and articles, games, pictures and videos increasingly accessible. It is now accepted as normal for children to learn electronically, both in and out of academic institutions (for example, through using downloaded information on various electronic tablets and portable computers). Controlling for exposure to such sources of learning is nearly impossible. An issue to be considered is that if we intend to nurture an ideal set of ITPs then the 'overriding' of the ITP may be able to reoccur. Internet and media exposure are likely routes for this to happen.

ITPs and Cognition

An individual with greater cognitive ability is assumed to be able to facilitate learning, acquisition and retention of knowledge quickly and more easily than others of lesser ability. Therefore, it is expected that they would also develop more correct ITPs about effective relationships between behaviours and situations (Schmidt & Hunter, 1998). In turn, a higher cognitive ability will affect the level of specific job experience they experience and how much knowledge they take from it. This follows the continuum that general knowledge facilitates the acquisition of more specific knowledge (Hambrick, 2003; Hambrick & Engle, 2002).

For Medicine itself, becoming an expert in the field relies heavily upon cognitive ability and the field itself is unique in a number of ways. Primarily, it is a field that is constantly changing. Choudhry, Fletcher and Soumerai (2005) describe 'keeping up' as a hurdle for practitioners and there is a huge interest in how doctors continue learning after initial clinical training. Expertise cannot be namely based upon cognitive ability therefore, and instead relies somewhat on the time since graduation and experience; graduates are seen to hold more expertise than those on the lower rungs of the learning ladder. In recall tests however intermediates (new graduates) tended to recall more information of case studies than later doctors (named experts) or earlier learners (named novices) (Coughlin & Patel, 1987). This may be due to different forms of learning and knowledge organisation; early on learning basic mechanisms at play (Kim & Ahn, 2003) and then using prototype categorisation in semantic memory (Rosch & Mervis, 1976) before finally progressing to exemplar models of categorisation and episodic memory (Medin, Altorn & Murphy, 1984). Regardless, in relation to ITP alteration, a cognitive ability peak may be seen to be at this intermediate point. Perhaps this is the critical period and where people should be targeted to ensure the highest

rate of cognitive ability to allow for ITP progression or change. It may also mean that this is the point where ITPs are most malleable as cognitive capacity allows for increases, big loads and potentially manipulation.

However, there are people who demonstrate good cognitive ability i.e. good declarative knowledge, and yet fail to acquire a good level of procedural skill i.e. they fail to employ sensible decisions or choose effective behaviours in certain situations. This may be due to a lack of ITPs or 'faulty ITPs'. For example, all medical trainees have been screened for university entry and have shown better than average scores across examinations. Hence, those students who fail to become 'good' doctors are generally lacking in the procedural knowledge as opposed to a lack of declarative fact learning. If there were practices in place to help, then a number of potential doctors could possibly become more successful following assistance before university or job entry. Hence, cost, time and effort of assessment as well as later personnel selection is at the forefront of these ideas.

The ITP concept and surrounding psychological theory relating to the development, expression and manipulation of the ITP has been described in detail. This concept has been recently introduced into the literature, and has led to a step forward in the understanding the expression of personality through SJT performance. This current standing upon the SJT is important to note as one of the two main theoretical models of the SJT, before exploring the specific theory that informs the studies in this thesis.

The SJT and other methods of selection

The incremental validity of a measure typically refers to the process of examining one predictor's value over and above another predictor (Schmidt & Hunter, 1998). For example, Lievens & Patterson (2011) note that the SJT and the assessment center (AC) give incremental validity over and above a knowledge test when predicting job behaviour in advanced high stakes settings. Both predictor measures also correlate with actual job performance, as noted above (Lievens & Patterson, 2011).

There are many other selection methods used within HRM and therefore it is important to compare these to an SJT as well. In terms of medical selection, validity has been researched where possible in terms of how a measure can predict job performance (i.e. criterion validity) although as noted this can often be difficult to do. Anderson & Cunningham-Snell (2000) note that methods such as work sample tests (.54), cognitive tests (.51) and structured interviews (.51) hold strong correlations with job performance, whereas methods such as

graphology (.00), astrology (.00) and references (.26) have much lower correlations with job performance.

The assessment of a method itself is difficult due to how the nature of jobs themselves differ and therefore how the method adapts and validity may be altered according to the role it is used to assess for. Therefore, using meta-analytic studies as the most suitable source of information, Patterson, Lane & Carr (2009) have classified the following selection methods and their relative accuracies as below in Table 2. They note that all occupation groups have been used in meta-analyses and hence, interpretation should be guided with caution. However, the table is still useful for a general overview of the method's accuracy in comparison with the SJT.

Table 2. Summary of selection methods, criterion validity, reactions and extent of use.

Selection method	Criterion validity evidence	Applicant reactions	Extent of use
Structured interviews	High	Moderate to positive	High
Cognitive ability	High	Negative to moderate	Moderate
SJT	High	Moderate	Moderate
Personality test	Moderate	Negative to moderate	Moderate
Work sample test	High	Positive	Low
Selection centre	High	Positive	Moderate
Handwriting	Low	Negative to moderate	Low
References	Low	Positive	High

The SJT is seen to have a moderate extent of use, a moderate reaction from applicants and a high amount of evidence for criterion validity, across work occupations and internationally. Hence, the practicality and validity of an SJT within personnel selection has been a widely researched topic (e.g. Lievens, Buyse & Sackett, 2005a; Lievens and Sackett, 2007, 2006; Patterson and Ferguson, 2007; Patterson, Ferguson, Lane, Farrell, Martlew & Wells, 2000; Weekley and Jones, 1999, Ryan & Ployhart, 2013).

McDaniel et al., (2001) summarised findings concerning the internal reliability of the SJT in their meta-analyses and reported coefficients between 0.43 and 0.94. They identified reasons for this large range and suggest that factors such as the length or instructions of the SJT may be the cause of such variation. SJTs are multidimensional measures that are used across many

occupations with many variations. Therefore it is potentially to be expected that due to this they may only show relatively low internal reliability as a tool (Lievens, Peeters and Schollaert, 2008).

SJTs have become popular over high fidelity simulations such as ACs. SJTs show promise due to the reduced cost, effort and time needed in the employment of the technique. For example, paper or video based SJTs can be developed by a small team, produced and distributed in mass (i.e. to many different applicants together) and a consistent scoring procedure is available after a panel of experts has given their responses. There are some concerns over the incremental validity that high fidelity simulations may show over low fidelity ones; the validity of a test should not be compromised for cost benefits or the aim of the low fidelity simulation is futile. However, returning to the Lievens & Patterson (2011) paper, this recent research suggests that the validity of an SJT is equal to that of an AC (a high fidelity simulation) with both emerging from analyses as significant predictors of job performance with correlations of above .30 and then both also showing incremental validity over a knowledge test when looking at overall job performance. Lievens & Patterson (2011) note the specific high stakes setting of their research and that further investigation in other settings is a direction for future research.

It should be noted that in less realistic (lower fidelity) ACs this statistical basis for validity may not be the same and hence these findings may not be generalisable across all ACs in selection. ACs for medical selection tend to be high in fidelity and use very close approximations of job tasks whereas not all ACs hold such close replicas of job roles. For the purpose of this literature review the focus is particularly upon high stakes selection and therefore the high fidelity selection centres that are used within medical selection specifically. In the AC referred to, Lievens & Patterson (2011) describe the AC as lasting for one day and over six weeks approximately 150 ACs were held. The day itself involved high fidelity tasks such as simulated consultations with actors performing as patients, group discussions resolving work related problems and written planning exercises regarding work prioritisation and justification for actions.

In other research concerning the validity of the SJT over other methods of selection, Lievens, Peeter and Schollaert (2007) report numerous studies that have investigated the incremental validity of the SJT over the established measures in selection of cognitive ability, job knowledge, job experience and personality (Chan and Schmitt, 2002; Clevenger et al., 2001;

Lievens et al., 2005a; McDaniel et al., 2001). McDaniel et al., (2007) conducted meta-analyses on this issue and summarised that SJTs had incremental validity over cognitive ability ranging from 3-5%, over personality measures 6-7%, and over both cognitive ability and personality measures ranging from 1-2%. Hence, the SJT appears to have some modest percentages of incremental validity above these measures across research.

This chapter will now introduce medical selection and assessment before discussing the SJT within this field to frame the empirical research within this thesis.

Medical Selection and Assessment

With medical professionals playing such a large and important role to each individual they encounter, their families, within their communities and the larger team of medical care across the country it is clear that ensuring the selection and assessment procedures in place are rigorously accurate, reliable and fair is important. These processes must allow for individual differences to show through and for actual future job performance to be predicted accurately. For example, patients reported wanting to consult with doctors who are good communicators, with sound, up-to-date knowledge and skills, who are interested and sympathetic, involve them in decisions, give them sufficient time and attention, and provide advice on health promotion and self-care (Little et al., 2001). The challenge is therefore upon measures of selection and assessment to illuminate the presence or lack of these characteristics in applicants.

In addition to providing correct assessment, diagnoses and treatment, the General Medical Council (GMC) requires that doctors display and adhere to certain rules and standards. For example, making the patient their first concern, being honest and trustworthy, respecting and protecting confidentiality, listening and respecting patients, recognising one's limits of competence, avoiding abusing one's position as a doctor and ensuring their personal beliefs do not affect job performance/behaviour (Hays, 2009).

The outcome of selecting poor candidates for medical roles clearly would have substantial costs; both human and financial. It is therefore, of great importance that medical selection and assessment procedures are identifying and assessing individual differences between applicants for specific courses and/or job roles. The identification of individual differences, skills, abilities and the consequential profiling of people for potential medical jobs is of upmost importance to the individual, the public and society as a whole.

In fact, organisational psychology is the ideal field to develop, establish and improve the process of selection from within. Scott (1920) described the establishment of the psychology of individual differences as the single best achievement of the American Psychological Association when discussing the science of employee selection following World War 1. The assessment of individual differences is further discussed in the section below.

Individual Differences

Murphy (2012) states that individual differences in cognitive and physical abilities, personality, interest and core self-concepts are all important and relevant for the understanding of job and organisational behaviour. The fact that people differ in so many different ways, both in stable or less stable ways, can affect job performance and therefore should inform job selection and assessment. Relatively stable individual differences may include personality traits, beliefs and core concepts about one's self, whereas individual differences that may fluctuate and change more frequently may include moods or preferences/interests at the time (e.g. music or art interests). Whilst the latter category may still affect performance upon selection and job test/performance, it is the former category that is more widely researched and held to play a major role within the psychological testing of individual differences with regard to employee selection.

Murphy (2012) splits individual differences into four main domains that are demonstrated through individual behaviours and choices related to work organisations; ability (mostly cognitive), personality, interests and self-evaluations. How much these four domains are influenced, caused or altered may be explained in terms of 'nature vs. nurture' theories, i.e. through a combination of both biological bases and environmental or social interactions.

The ability domain engulfs cognitive ability and skill, for example, intelligence and how this differs from one person to the next. Murphy (2012) states that cognitive ability is present in almost any aspect of activity that involves information processing, and that measures are consistently related to success, performance and effectiveness in school, at work and in life (Jensen, 1980, 1998; Kuncel, Hezlett & Ones, 2004; Sackett, Schmitt, Ellington & Kabin, 2001; Schmidt & Hunter, 1998). There are numerous theories and measures of cognitive ability and these have been discussed above in the intelligence section of chapter 2.

Personality refers more to stable subjective traits that one possesses, and the behavioural consistency shown across situations/performance because of these. If a person is seen as highly agreeable it is likely that they will show this trait across numerous situations and

scenarios. However, Murphy notes the importance of personality as a not completely consistent variable, i.e. there will be variations in this in some cases, for example, an agreeable person will likely still show aggressive behaviour in a knife fight. Personality measures and theories have been further discussed in the personality section of chapter 2.

Interests represent consistent preferences that govern behaviour, for example, a football supporter will actively look for fixtures and seek out information regarding football when compared with a non-football supporter. Murphy (2012) notes that interests are especially useful when examining why people choose, accept, pursue or leave a job role. Strong (1943) defines interests as 'responses of liking' to particular people, things or events. It is suggested that high interest levels elicit positive affect, low interest levels elicit little affect and no interest elicits apathy or aversion (Murphy, 2012). Interests are consistently related to job and career satisfaction, motivation, persistence and involvement in work and work organizations (Campbell & Johansson, 1966). Hunter & Hunter (1984) showed that interests are not generally strong predictors of work performance, although Mount & Barrick (2012) more recently have stressed the importance of investigating the relatively unexplored area of interests and personality traits to reveal better predictions.

Finally, self-concepts are beliefs regarding oneself that constrain, demonstrate or exaggerate personal abilities, personality traits and/or interests. For example, confidence and self-efficacy regarding one's skill will affect performance, the choice over challenging or non-challenging tasks and therefore, in turn, likely job progression and performance. Self-concepts are also referred to as Core Self-Evaluations (CSEs) and they reflect beliefs in one's capabilities, competencies and general sense of life turning out well, which in comparison, makes them broader than measures of pure self-esteem (Murphy, 2012). Judge (2009) notes that CSEs predict a range of outcomes including job performance, job satisfaction, life satisfaction, subjective well-being and motivation. Furthermore, Murphy (2012) adds that CSEs give incremental validity over the dimensions of personality in predicting job performance (Judge, Erez, Bono & Thoresen, 2003).

The literature review will now discuss the medical selection and assessment field to frame the research related to medicine that follows therein.

Structure of Medical Training and Selection

As this thesis focuses upon the medical field, and in particular a selection test used within this profession (the SJT), it is necessary to introduce the field itself in more detail, with an

overview of how course and career progression takes place, and in particular the typical selection procedures within this progression. Assessment within medicine is considered following this overview of the course and career route.

A career in medicine begins with an individual applying for a medical school course lasting typically 5 years. Upon acceptance onto this course, they then join and complete the degree programme and following this qualification, move on to basic years training (Foundation programmes for 2 years) followed by speciality training in a specific chosen area of the field. For example, in surgery or general practice. Progression through speciality training can then lead to higher positions and salaries within senior roles, for example the ultimate clinical position of Consultant.

University medical school selection typically follows a Curriculum Vitae (CV) type application through the UCAS system. Following this, interviews are given and conditional offers made upon academic assessment achievement requirements (e.g. three A grades at A – level). For medicine, intelligence levels are expected to be high and consequently, application forms and personal statements through UCAS forms are expected to be equally impressive. The majority of medical schools use the UKCAT (UK Clinical Aptitude Test) as a ‘measure of appropriate mental abilities, attitudes and behaviour required for new doctors’ (UKCAT Consortium, 2015). This is a test taken that assesses four main domains of ability; verbal reasoning, quantitative reasoning, decision making and abstract reasoning. Scores are given per domain and as an overall total. Whereas many Universities still use this test, some have lost faith in it and discarded it (for example, there has been reporting of a gender bias in the measure; James, Yates & Nicholson, 2010). Many universities also use non-cognitive factors for assessment, for example, questioning regarding empathy levels within interview to determine if the individual is correct for the programme and a career within medicine. If applicants successfully pass the entry measures, and gain the necessary A-levels or qualifications, they will then be awarded a place in medical school to enable them to embark upon the degree programme.

A medical degree course is normally a five year undergraduate course that is taught and non-clinical based for years 1 and 2, before proceeding to more practical and clinically focused for years 3-5. There is often an ‘elective period’ in year 4 where students can choose to travel abroad or elsewhere in the country to enable them to gain ‘hands-on’ hospital experience.

Throughout the course there are numerous points of both practical and non-practical assessments with hospital experience and classroom teaching.

Selection into medical school, comparable to gaining a training position for any other profession, should involve selecting those who will hopefully in the future become competent doctors, before an applicant has even begun training. Developing selection systems for this is a difficult challenge (Patterson, Lane and Carr, 2009) and methods should be standardised, reliable, fair, cost-effective, feasible and valid, among other potential criteria.

Within medical school, various assessment and teaching techniques are used. Medical schools will vary in how courses are taught relative to their resources, preferences and results according to methods. Professionals and academics from various disciplines will be involved in the teaching, for example, qualified clinicians, neuroscientists, anatomists and psychologists. Many institutions will also rely upon mentoring and pastoral support units for further guidance. Furthermore, older students may be used as teachers in peer-assisted learning. Peer-assisted learning is rooted in educational theory and is becoming increasingly popular (Cantillon & Glynn, 2009).

For assessment, options of measures include Multiple Choice Questionnaires (MCQ's), oral examinations, poster presentations, and clinical assessments, such as Observed Structured Clinical Exams (OSCEs). In the final year of medical school applicants will go through selection processes again when they apply for their Foundation years (basic training in hospital rotations for two years post qualification as Foundation 1 and Foundation 2 level Doctors). Following this, they will be assessed again when they then choose a 'speciality' (for example, General Practitioner). These selection procedures, for foundation and specialising, typically contain specific knowledge tests, interviews, Assessment Centres and SJTs.

Career speciality may be influenced by many factors including personal preference, personality, temperament, potential job progression, hours of working, family/friends and life events. McManus & Goldacre (2009) note the likeness of a medical professional's career to that of a microscopic particle, being continually buffeted by outside events as 'stuff happens'. Although there is still relatively little known about why doctors end up in particular specialties (McManus & Goldacre, 2009) John Holland's hexagon model of career preferences (1973; 1997) is a widely accepted diagram of how people interested in certain careers may not be interested in others and is often used for examining speciality choice

within career path. It contains dimensions of interest such as investigative, artistic, social, realistic, conventional and enterprising. Such ideas are taken in conjunction with the knowledge that within the real world, what one would like to do is not always what one is able to nor should do.

Compromise is necessary when considering job choice (Gottfredson, 1981; 2002) and also the awareness that job choice is quite often based upon what one does not want to do, as opposed to what one does want to do; most people are more aware of the former than the latter (McManus & Goldacre, 2009). McManus & Goldacre (2009) also note that factors such as gender affect choice, for example, more men generally apply to be surgeons than women. Early choices also appear to change when looking at later specialisations, with many failing to choose general practice, psychiatry, radiology and pathology at first opportunity.

After 3-7 years of successful progression in their chosen speciality the doctor can go through selection/assessment processes again in order to become a Consultant. Hence, from the start of medical school throughout their careers, medical professionals are continually monitored and assessed.

Most doctors end up working in organisations such as the National Health Service (NHS) or within the independent sector in the UK, or managed care systems within the USA. In the latter cases, doctors are usually employed by an institutional provider. Levels of responsibility, autonomy and motivation can vary according to work organization and/or location. Typically, across all medical careers, there is a large emphasis on teamwork and professionalism. The shift towards more patient-centred care has led to a profession that now is evolving around both clinical skills/knowledge and professionalism/competency, as opposed to an older system where doctors controlled most encounters, care and decisions through a clinical knowledge upper hand and passive patient acceptance (Waas & van der Vleuten, 2009).

Once practicing, doctors typically report intentions to remain within medicine for the foreseeable future when asked (approximately 75%). Those who do not say this tend to report that they would practice medicine abroad, as opposed to leaving medicine entirely (Lambert, Goldacre & Parkhouse, 1997). Consequently it appears that medicine is a lasting career choice whether it is in the UK or elsewhere.

Following the overview of the typical medical course and career path, the assessment process within medicine will now be discussed in more detail. The issues relating to assessment both during and after training are central to this thesis as it attempts to further understand the methods used to establish individual differences in medics.

Assessment in Medical Education and Training

Medicine is a career that requires consistent learning and therefore, education can be regarded as a life-long continuum (Waas & van der Vleuten, 2009). It follows that assessment has to match this and also move with the education and ever changing nature of the field. Somewhat traditional methods, for example, examinations as a set end point to curriculum, are being steered away from (van der Vleuten, 2000). This movement can be referred to as in a ‘competency based’ curricula direction. Competency is defined as ‘the ability to handle a complex professional task by integrating the relevant cognitive, psychomotor and affective skills’ (Wass & van der Vleuten, 2009).

Changes have introduced more formative, reflective and supportive assessments that progress and ensure constant development and assessment of both the medical professionals and students. There are other modern issues that are becoming increasingly important to consider as well, for example, the reliability and the validity of a measure, feedback opportunities and quality assurance of a measure. There is an emphasis for medical schools and teaching professionals to protect themselves, and the institution they are employed by, as well as keep the best interests and progression of the students as their focus. In such a high pressure, lengthy and financially demanding course, failure and problems are not easily swept under the carpet, and may even result in legal action being taken following a forced exit from the course, or a final failed attempt at training. Defensibility of teaching, assessment and support for students is key.

Miller (1990) developed a model of competence assessment that is still widely used today to establish the aim of assessment. According to Wass & van der Vleuten (2009) the purpose of an assessment should be clear and transparent, ensuring it mirrors the teaching and aim of educational content. Although the Miller (1990) model has come under some criticism to be inverted or altered (Rethans, Norcini, Baron-Maldonado et al., 2002) the ideas of still using the principles of assessing what one ‘knows’ (basic facts), if one ‘knows how’ (applied knowledge), if one ‘shows how’ (clinical competency) and finally, if one ‘does’ (test performance in workplace) are still criteria used to design assessment procedures. For

example, 'knows' may be assessed by a Multiple Choice Questionnaire (MCQ) on medical knowledge, whereas 'shows how' may be assessed by an Objective Structured Clinical Examination (OSCE; a practical exam of clinical competency). There are real challenges in assessing 'does' as workplace assessments are ethically and realistically difficult to design, measure and administer. There is also increasing pressure for the model to include assessment of domain independent knowledge and aspects of behaviour that are essential for professional behaviour (Waas & van der Vleuten, 2009).

Until relatively recently, assessment in medical education and training was pre-competency; for example, it consisted mainly of high-stakes, summative, intimidating and one point in time examinations, quite often in examination halls and little in the field. Focus is now more upon the workplace and competency as well as basic fact knowledge assessed by written best answer questions. To establish methods and batteries of assessment and selection that satisfy these new criteria, and the needs of the student/trainee, examiners, organisations and stakeholders, is a potentially difficult task. Methods that do fit modern views of assessment and guidelines are known as 'contemporary best practice'. The philosophical shift towards these ideas suggests that selection and assessment will become competency/professionalism based, as well as knowledge based, more formative, open and transparent to all involved and affected by it and consistently rigorous in terms of reliability and validity. Extensions of a basic utility model for modern assessment (van der Vleuten, 1996) suggest reliability, validity, educational impact, acceptability and cost are the main factors to be considered in assessment method choice/design now.

Medicine and the SJT

Medical selection using SJTs is a high volume, high cost, high stakes, and high pressure operation taking place in most countries, with over 8000 candidates annually in the UK for just 3, 250 General Practitioner (GP) jobs alone (Patterson, 2009). The costs involved are illustrated by an example of training a General Practitioner in 2006; the annual training costs for a GP were approximately £87,000, which across the three year training programme totals £261,000 per trainee (Patterson, Lane & Carr, 2009).

There has been extensive, high quality research internationally regarding development of an optimally effective and valid methodology for undertaking the medical selection process in general. In particular, as a relatively new method, the SJT is of interest. Within medicine the SJT has been used for selection for GP speciality posts (from 2007), for ranking medical

students for their training years following graduation (from December 2012) and now even for entry into some medical schools (from course entry for Autumn 2013).

The SJT used in this thesis for two of the three main studies was developed and used for the purposes of ranking medical students in their final year for foundation training year purposes. Within the academic literature, the SJTs used for the medical gateways noted above were all developed in line with methods from other main studies (Weekley et al., 2006). For example, using job analysis and subject matter experts for both the establishment of target domains and marking/scoring decisions. Instructions tend to be knowledge based, as these are less prone to faking (Lievens, Sackett & Buyse, 2009). The performance domains and individual items tend to differ according to the purpose of the SJT. For example, in Lievens & Patterson (2011) the GP speciality selection SJT used was mapped onto five performance dimensions (communication, empathy, professional integrity, coping with pressure and problem solving) and contained 50 items that were part ranking format and part multiple choice format, with all knowledge tendency instructions. The SJT used for foundation selection is mapped onto five different domains (commitment to professionalism, coping with pressure, working effectively as part of a team, effective communication and patient focus) but similarly is part ranking format and part multiple choice format, with all knowledge tendency instructions, but with a greater number of items (64 items in total).

The research base has tended to disperse under three main themes; exploring factors influencing the criterion related validity of the method (e.g. different high-stakes contexts or response instructions), exploring the nature of constructs assessed by the method (i.e. what are the SJTs intending to measure) and lastly, exploring subgroup differences upon SJT performance (for example, factors such as race) (Ryan & Ployhart, 2013).

Beyond these three main directions, research into the general and medical selection SJTs has continued to develop to incorporate new themes and avenues as literature has become more extensive. For example, the method of administration for the test itself has moved from written to video based situations (e.g. Lievens, Buyse & Sackett, 2005a). Recent ideas also include focus upon the employment of the technique and how applicable it can be for different groups or settings (for example, within higher education and/or employment settings, as seen in Lievens & Sackett, 2012). Other new ideas see extensions of variables to further investigate potential links affecting SJT performance, for example, Slaughter, Christian, Podsakoff, Sinar & Lievens (2014) have recently introduced 'anger hostility' as an

interacting variable upon SJT performance, instead of the commonly used standard personality inventory scales.

This review will now discuss the criterion validity of the SJT in relation to job performance, and the construct validity of the SJT from current research. These two main themes are central throughout the remainder of this thesis and the studies presented.

Criterion Validity of the SJT: Job Performance

The criterion validity of a measure concerns how well the variable predicts the outcome of another variable. For example, how much does a personality trait affect performance on a behavioral criterion? Within context, the criterion validity of an SJT is examined in relation to how well SJT performance predicts actual job performance/ behaviour. Assessment of this predictive form of validity is therefore difficult, and involves collection of data at time 1 (SJT result) but also a follow up and collection of outcome data during actual job performance years (longitudinal assessment). In medicine, this is particularly difficult and there are marked difficulties, ethically and practically, in gaining an objective measure of on the job behaviour. Ethically this is difficult due to the nature of most medical job roles as they involve patient interaction and assessment. Often matters can be very personal or private, and therefore the confidentiality of the patient is of utmost importance. Beyond this issue, numerous patients and/or doctors may not want to be observed or assessed due to these factors. Furthermore, another avenue of longitudinal assessment of medical professionals may use supervisor ratings. However, many roles are not continually, or even often, observed or monitored by supervisors. For example, on a busy hospital ward a training Foundation doctor may be left alone to see patients and provide care without constant supervision. Hence, ratings may be unsatisfactory or based upon little valid evidence.

Some studies have in fact managed to obtain on-the-job behaviour ratings of doctors and have been able to use this for criterion validity related studies regarding the SJT. As already noted in this review, Patterson and her colleagues (Lievens & Patterson, 2011; Patterson et al., 2012) embarked upon an extensive research programme to investigate this issue. They investigated the validity of knowledge tests, low-fidelity simulations (SJT) and high fidelity simulations (Assessment Centers; ACs) in predicting job performance in a high stakes setting. Furthermore they examined the incremental validity of the SJT and AC compared to each other and the knowledge test. Incremental validity examines the validity of one measure over and above another measure.

Specifically, the participants were 196 General Practitioners (GPs) working in the UK who were applying for specialty training following medical education (the 5 years degree Programme) and Foundation training years (Foundation year 1 and Foundation year 2).

Firstly, in the development of the study measures, objective and face valid measures of performance in medicine were determined via focus groups, critical incident analyses and an expert panel of experienced assessors. This led to the identification of eleven competency dimensions, of which six (empathy, communication, problem solving, professional integrity, coping with pressure, and clinical expertise) were judged most critical at the point of selection. Over an extensive period, comprehensive item banks were built up and refined for clinical knowledge and problem solving items (knowledge and skill) and for the SJTs, a bank of scenarios was constructed and validated relating to the five dimensions other than clinical judgment. Subject Matter Experts (SMEs) were also used for the development and implementation of a scoring key for the developed SJT.

Participants then undertook a two-stage selection process. Firstly, a composite score was created from the SJT and then for their knowledge/skill. The SJT included items with knowledge response instructions, for example:

‘You are reviewing a drug chart for a patient with rheumatoid arthritis during an overnight shift. You notice that your consultant has inappropriately prescribed methotrexate 7.5mg daily instead of weekly. Rank in order the following actions in response to this situation, 1=most appropriate, 5=least appropriate.

- a. Ask the nurses if the consultant has made any other drug errors recently.
- b. Correct the prescription to 7.5 mg weekly.
- c. Leave the prescription unchanged until the consultant round the following morning.
- d. Phone the consultant at home to ask about changing the prescription.
- e. Inform the patient of the error.’ (Lievens & Patterson, 2011, p. 939).

Only those candidates passing the cut-off point were invited to the second part of the process; the AC. The AC involved three work-related simulation exercises devised from job analysis information; (a) a consultation with patient simulation, (b) a discussion over a work related issue with colleagues, and (c) a written planning exercise concerning prioritisation of work tasks. Experienced General Practitioners (GPs) acted as assessors for the AC exercises, and where possible these altered for participants. The knowledge test used was a machine-graded

assessment of clinical knowledge (declarative knowledge; knowledge of facts, rules and principles) consisting of 98 questions that were also devised by the SMEs.

The successful candidates entered into an actual supervised GP training phase in which they acted as a GP (with an experienced GP as supervisor) for up to 3 years in duration. The supervisors then completed a 24 item assessment of their candidate's performance one year into the job (4 items on a six point scale) for each of the 6 dimensions. This performance criterion measure was then correlated with the selection measures, and a structural equation model developed. All three predictor measures (knowledge tests, SJT and AC) were significantly related to job performance (correlations of 0.36, 0.37 and 0.30 respectively, with higher values of 0.54, 0.56 and 0.50 respectively if a correction for range restriction is applied).

A hierarchical regression analysis indicated that the knowledge and skill accounted (following correction for range restriction) for 29.4% of the variance, the AC then accounted for a further 5.7% and the SJT had 5.9% additional explanation over the knowledge test. The AC was also found to have 3.0% incremental validity over the SJT when predicting job performance meaning that an AC is still a useful selection measure alongside an SJT. The authors investigated further the cause of the incremental validity of AC over SJT and concluded that it arose from the non-cognitive measures (communication, empathy, integrity and coping with pressure). Most importantly, the authors reported that the SJT fully mediated the effects of the AC on job performance, and had a direct effect on job performance independently as well. Procedural knowledge (knowledge of how to do a task) was found to build upon declarative knowledge (knowledge of facts, rules and principles) and it also fully mediated the effects of declarative knowledge upon job performance. Consequently, unsurprisingly, procedural knowledge was also found to be a precursor of AC behaviour as ACs aim to produce job performance examples. Furthermore, procedural knowledge should have causal links to job performance because people, who know what they should do in order to be effective, are logically more likely to perform effectively in a job role (Motowidlo, Hooper & Jackson, 2006). Correlations between SJT performance and job performance do generally support this idea of criterion related validity (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001).

This approach to medical selection nonetheless leaves some doubt in the theoretical completeness and the extent of the criterion validity. From a practical perspective, it appears

that the entire process accounts for a moderate amount of performance variance. As noted earlier, some personnel selection measures correlate with job performance more highly at approximately .50 (e.g. work sample tests; Anderson & Cunningham-Snell, 2000). The costly AC does provide some, albeit modest, unique additional variance over the other two methods. Finally, the SJT is having direct and indirect effects upon job performance and the method therefore needs to be understood completely. This knowledge can then inform the development of theory behind the SJT, and potentially the refinement of the measure, with the aim of increasing the explanation of criterion validity of this low fidelity simulation.

As it has been noted, it is of utmost importance to determine how well a selection method can predict job performance, and it is key therefore to question the criterion related validity of the SJT. In line with the research noted above by Patterson and colleagues, in their meta-analyses McDaniel et al., (2001) used 95 studies of SJTs across various work settings found a corrected correlation with job performance of .34 (uncorrected .26). They suggest that the range of correlations across the studies suggests the presence of moderating variables within the research; for example, whether the SJT was based upon job analysis for development (such studies tend to produce higher correlations than those that are not based on job analysis). In 2003, McDaniel et al., re-analysed and updated the 2001 data. They found that knowledge response instructions had higher validities ($r=.33$) than behavioural tendency instructions ($r=.27$) when predicting job performance.

Following the discussion of the criterion validity of the SJT, this literature review moves to discuss the construct validity of the SJT and introduce the current theoretical modelling of this which will inform study 1. Ployhart (2006) notes that it may have been considered acceptable to discard investigations into what the SJTs as predictors measure from a practical sense, i.e. they are reported to predict job performance and that is satisfactory. However, in reality Ployhart goes on to explain why understanding the underlying construct validity is important for three reasons. Firstly, from a scientist stance the method needs to be understood if it is to be predicted and responses controlled in applicant settings, secondly, practically it is necessary to inform selection systems and implementation, and thirdly, it is important to explain to the general public what is being measured and how it relates to performance, which is one right of a test taker (Joint Committee on Testing Practices, 1998)

Construct Validity of the SJT

Another important aspect of validity, alongside criterion validity, is the construct validity of a measure. The construct validity concerns whether a test accurately measures what it is supposed to measure. For example, does the personality measure actually measure levels of the trait it specifies? In the context of the SJT the proposed construct of measurement is somewhat unclear and generally is labelled as ‘procedural knowledge’ (e.g. Lievens & Patterson, 2011). Therefore, the construct validity of an SJT concerns whether the SJT actually assesses procedural knowledge, or if it potentially assesses something else. Construct validity within the SJT literature appears to be viewed from two angles within research which reflect the versatility and differing nature of specific SJTs from one another. Firstly, those specific constructs that the SJT may have been designed to assess (e.g. communication skills) and secondly, the more general constructs of interest (e.g. personality). For example, if an SJT is designed to assess a domain of interest (e.g. teamwork) this may be assumed to be the ‘construct’ of interest within the construct validity of the SJT in question. Whilst this is true, the construct validity literature surrounding the SJT concerns general constructs assessed across SJTs in broad terms. For example, personality traits or cognitive ability. The SJTs are held to have general constructs of interest, as well as potentially specific ones given in design and development. These specific ones can be held to affect links with the more general ones as discussed later in this section. For example, an SJT designed to assess teamwork skills specifically may relate more strongly to certain personality characteristics over others, and potentially show weaker links with cognitive ability when compared to the personality traits.

Researchers have investigated various constructs of interest that have been linked with the SJT over the years (Lievens, Peeters and Schollaert, 2008) and even early papers (e.g. Wagner and Sternberg, 1985) were acknowledging that the purpose of the SJT is to measure beyond cognitive ability (intelligence/academic attainment). They proposed that the SJT measures ‘tacit knowledge’ or ‘practical intelligence’ (see earlier intelligence review for further information on this concept). Other research however does not support this, and suggests that SJTs are in fact related to cognitive ability (McDaniel & Whetzel, 2005). Research into the construct validity now mainly concerns personality traits and cognitive ability; these are shown to be measured by the SJT to varying degrees, and these correlates make up the main current body of research into construct validity of the SJT (McDaniel et al., 2006).

It is widely established that cognitive ability/ knowledge tests assess declarative knowledge (facts, rules, principles and procedures; Kanfer, Ackerman & Cudeck, 1989) and are often used alongside SJTs which assess procedural knowledge (knowledge of how to execute some task; Anderson, 1982). A basis of sound declarative knowledge forms a foundation for procedural knowledge to develop through experience (Beier & Ackerman, 2005). General knowledge seems to inform the later learning of more specific knowledge (e.g. Hambrick & Engle, 2002). Such cognitive ability has been shown to be a key determinant for further knowledge and acquisition (Kanfer, Ackerman & Cudeck, 1989). This procedural knowledge experience may be domain specific (i.e. referring to the subject area that the task is related to) or it may not be (i.e. learnt in any scenario).

To put these ideas into context, from a medical selection perspective, declarative knowledge concerns the facts and principles behind medical procedures and ability. For example, which drug to prescribe, whether blood pressure is worryingly high or what certain symptoms may mean for an individual's health. The procedural knowledge may concern which behaviours are effective in a certain situation and which ones are not i.e. shouting at a colleague when trying to get their attention may not be as effective as waiting until they are free and then asking politely for their attention or help. It is this aspect of knowledge that SJTs are suggested to assess. This procedural knowledge may have been a result of job specific experience (e.g. a past encounter with another doctor whilst working) or from general life experience (e.g. a past encounter with a family member) or potentially even from others experience (vicarious learning or modelling; Bandura, 1986) or purely from being told (i.e. not from experience but from another's imparted wisdom).

It may be that the links between the SJT and cognitive ability test scores are due to the links with such declarative knowledge providing a basis for procedural knowledge. However, it may be that this is not the case. For example, video based SJTs have lower correlations with cognitive ability than written SJTs (Weekley and Jones, 1997) suggesting that the link is lost the further it moves away from a written test format, and therefore the link may not be in the knowledge the measure assesses, but in the format as well. There are other papers that still suggest the SJT as an alternative measure of job knowledge, job experience or interpersonal variables (McDaniel and Nguyen, 2001). It appears that it is somewhat unclear what the SJT is measuring and in fact this may be dependent upon the job occupation and/or format of the SJT. However, there is great scope for further research into this and the cognition assessed by the SJT.

Meta-analyses have investigated the construct validity of the SJT further (for example, McDaniel, Hartman and Grubb, 2003; McDaniel, Morgeson, Finnegan, Campion and Braverman, 2001; McDaniel and Nguyen, 2001). SJTs were shown to correlate to varying degrees across studies with the commonly investigated variables of cognitive ability and the big five personality factors (McDaniel et al., 2003).

The correlations between the SJT and cognitive ability from meta-analytic research (McDaniel et al., 2003) average at $r=.39$ from 62 papers examined. These values differ for behavioural tendency questions ($r=.23$) and knowledge tendency questions ($r=.43$) and these values somewhat suggest that the construct validity of the SJT may be rooted in basic cognitive ability/ intelligence, as opposed to 'procedural knowledge' specifically. However, as noted, the theoretical basis of procedural knowledge within declarative knowledge may also explain the relationships. Chan & Schmitt (2002; 2005) suggest that variability reported in SJT correlations with cognitive ability measures is that the SJT can be used to assess cognitive ability, but that it does not necessarily do so, and may be used to assess a variety of other constructs that may or may not correlate with cognitive ability themselves. This was based upon the McDaniel et al., (2001) meta-analysis reporting a mean correlation of .36 (corrected .46) of SJT performance and cognitive ability with only 12% of the variance being attributed to artefacts and a credibility interval (10th-90th percentile) of .17-.75. This indicated that in some cases the relationship was high, but then in some cases it was very low or practically non-existent showing large variability in this relationship (Schmitt & Chan, 2006). The correlations of the SJT with the Big Five factors of personality on average in McDaniel & Nguyen (2001) were .25, .26, .31, .06 and .09 for agreeableness, conscientiousness, emotional stability, extroversion and openness respectively. Correlations have also been subsequently reported between SJT performance and the personality traits (e.g. Chan & Schmitt, 2002). However, they are not always present; in Lievens & Coetsier (2002) no correlations between the Big Five and the SJT were reported to exceed .15. Similarly, Clevenger et al., (2001) reported correlations of only .00, .16 and .21 between the SJT and conscientiousness in three different samples. These results support the suggestion that SJTs are designed to correlate with different constructs and can be developed to do so (Chan & Schmitt, 2005).

Schmitt & Chan (2006) reported that a great deal of variance in SJT measures, in their meta-analytic data, was unrelated to performance, cognitive ability, personality and interests. They

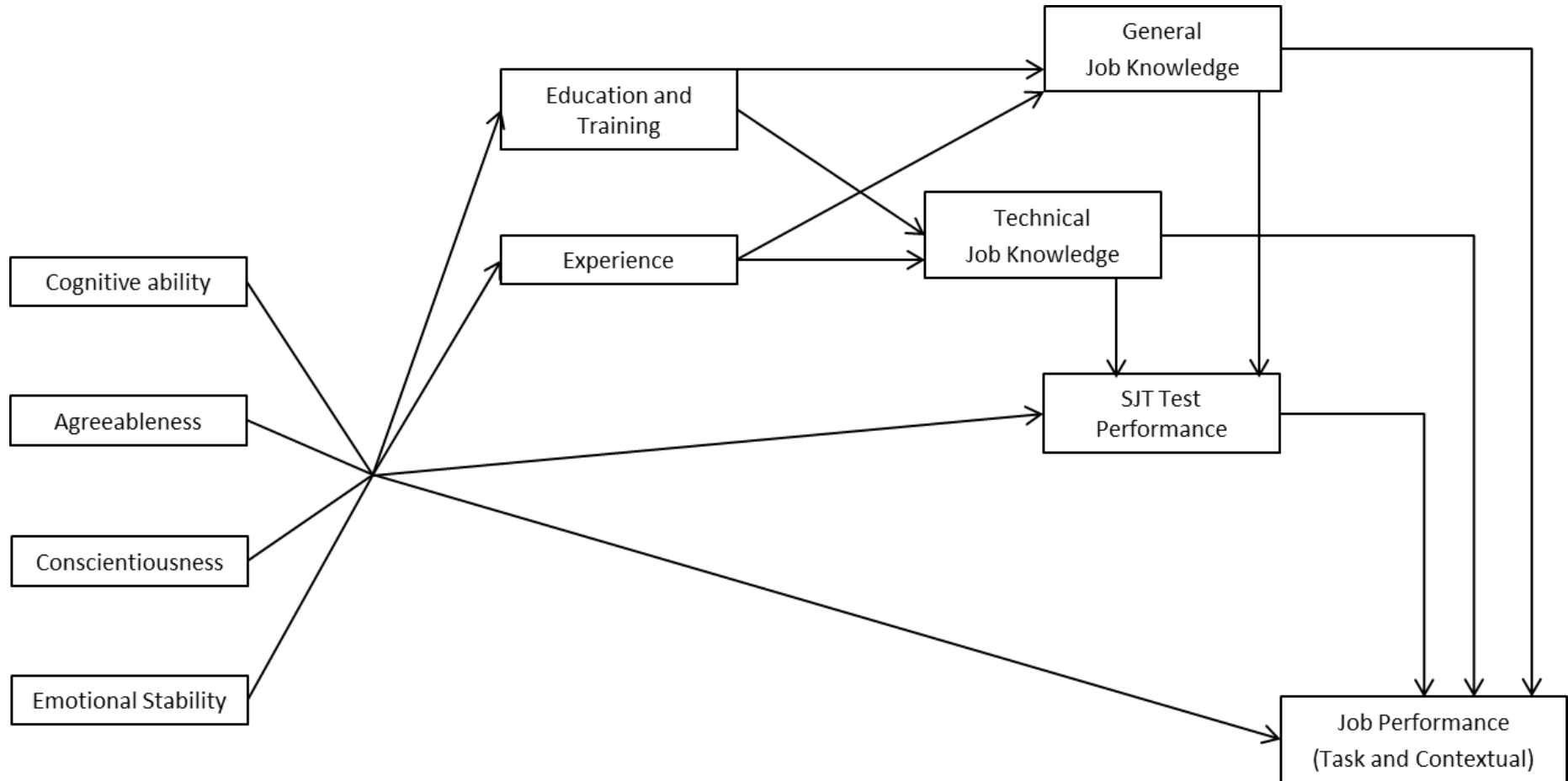
suggest that there may be an underlying situational judgment construct that is independent of most other individual difference variables used typically in personnel selection.

In summary, SJTs are typically correlated with these noted measures (McDaniel et al., 2006) although there are factors affecting these relationships. For example, SJTs with behavioural tendency instructions are found to correlate more highly with personality constructs, than knowledge tendency instructions (agreeableness .53 vs .20, conscientiousness .51 vs .33 and emotional stability .51 vs .11). As noted above, knowledge tendency instruction SJTs were found to correlate more highly with cognitive ability than behavioural tendency instruction SJTs (McDaniel et al., 2003). This is unsurprising as knowledge tendency instructions may be viewed as a maximal measure of performance that focus more upon the correct and most knowledgeable action, as opposed to behavioural tendency instructions that may be seen as a typical performance measure allowing for individual interpretation and the application of personality/individuality.

McDaniel et al., (2006) note cognitive ability, agreeableness, conscientiousness and emotional stability as the most important constructs involved in the SJT based upon past empirical research. This is further supported by meta-analyses such as Schmitt & Chan (2006).

They propose a model of these four constructs and performance on the SJT and this relationship can be mediated by education, training and experience that in turn then affect an individual's general job knowledge and technical job knowledge. The model can be seen in Figure 5.

Figure 5. Model of construct (and criterion) validity of the Situational Judgment Test (McDaniel et al., 2006).



This model is based upon the literature showing links between the four constructs and the SJT that hold consistent correlations (albeit of varying degrees) between these constructs and the SJT. The four personal traits affect how an individual's education and experience then influence job knowledge and general knowledge. For example, McDaniel et al., (2006) note that a smart and dependable individual would gain more through educational and training experiences and opportunities than a lazy and less intelligent person. Furthermore, the four personal traits affect ones' attainment of job knowledge.

McDaniel et al., (2006) have split job knowledge into 'general and technical' to allow for the subtle differences in these to show through. For example, general job knowledge consists of general principles consistent across job roles, such as how one is socialised and accepted into a work community. Principles within this area of knowledge may include punctuality at work, polite and courteous behaviour towards other staff and following instructions from superiors in a sensible and effective manner. The technical job knowledge is more specific to a role and hence is gained mainly through education and training. Within a medical context, this would include knowledge such as awareness of the correct drug to prescribe for a certain illness that would have been learnt during Medical school training courses. Both technical and general job knowledge are seen to directly affect SJT test performance.

The model is summarised as one that shows SJT test performance as a direct function of cognitive ability, conscientiousness, agreeableness and emotional stability, and through these variables acts on job knowledge, as mediated by education, training and experiences (McDaniel et al., 2006). The three personality traits from the five factor model of personality included in the McDaniel model (agreeableness, conscientiousness and emotional stability) directly map on to the broader personality factor α (also known as 'stability' or 'communion'). This factor has been defined as socialization processes, conformity and the extent to which one is consistent in motivation, mood and social interactions (Mount & Barrick, 2012).

Furthermore, it considers the larger framework of criterion validity and the SJT performance relating to actual job performance. Actual job performance is predicted by the four key constructs (cognitive ability, conscientiousness, agreeableness and emotional stability), SJT performance and job knowledge (general and technical). Different SJTs are held to measure the four personal traits to differing amounts hence the incremental validity of a specific SJT will depend upon its' correlates. To illustrate this, an SJT with large correlates with cognitive

ability may have little incremental validity over cognitive ability in predicting job performance, but however, may have substantial incremental validity over personality traits when predicting job performance.

Job performance is also split into task and contextual groups. Knowledge based SJTs are suggested as better predictors of task based job performance, whilst behavioural tendency SJTs are suggested to better predict contextual job performance.

The model is proposed as a heuristic to understand the relationships and is based upon available data (McDaniel et al., 2006). They note that more theory and better models are necessary for the future. Whilst this model is a step in the right direction to fully theoretically mapping the SJT, it appears that personality factors and cognitive ability as constructs to investigate 'may be too restrictive' (McDaniel et al., 2006, p.196). There is potential for researchers to investigate new constructs that the SJT may actually target.

Nonetheless, the model is coherent and consistent with current research and empirical evidence within SJT literature. Furthermore, it appears to be the most developed and current modelling of the SJT based upon empirical evidence, regardless of the stated need for more development and construct investigation. Based upon this requirement and suggestion of development from the model authors, this literature review will now consider other relevant constructs of interest to the SJT from previous literature. Following this the literature review will summarise and discuss the main gaps in the literature.

Other constructs of interest related to the SJT

Whilst the above four personal traits (cognitive ability, conscientiousness, agreeableness and emotional stability) are noted as the most researched and correlated constructs with SJT performance, other constructs have previously been suggested as ones of interest that the SJT may be measuring.

This section will discuss constructs and provide a foundation for this thesis to then target new constructs for investigation in relation to expanding the construct validity measures associated with the SJT from a cognitive psychology perspective.

Returning to the intelligence literature, Stemler & Sternberg (2006) suggest that the overarching link between all SJTs and the aspect of the measure that researchers are missing is a construct of 'practical intelligence'. As noted earlier in the literature, Sternberg (1985) developed a triarchic theory of intelligence involving componential intelligence (i.e. mental

mechanisms used to plan to execute a task), experiential intelligence (i.e. ability to deal with new situations and solve problems that have already been dealt with before) and contextual intelligence (i.e. intelligence reflecting behaviours developed from natural selection). He described metacomponents, performance components and knowledge acquisition components. Metacomponents decide the nature of a problem in hand for example, a skilled reader will decide how long to spend reading a passage dependent on how much information they want to acquire from it (Wagner, and Sternberg, 1985). Performance components are the processes used to perform the task in hand and knowledge acquisition components are used to gain new knowledge from sifting out the unnecessary information. Stemler and Sternberg (2006) also describe creative, analytical and practical thought. Creative thought is invoked when tasks are novel, analytical thought is invoked when familiar problems are fairly abstracted away from life and practical intelligence when components are applied to everyday life experiences in order to adapt to, shape and select environments. Thus, the same components applied in different situations invoke different types of thought.

Stemler & Sternberg (2006) suggest there is a cognitive element (consisting itself of a tacit element concerning guides over behaviour that one cannot explain, and an explicit element acquired through formal training) and a behavioural element to practical thought. For example, one can know how to execute the right solution to the task in hand yet not actually be able to implement this plan. They note that behavioural assessments are difficult, yet they have developed tacit knowledge assessments.

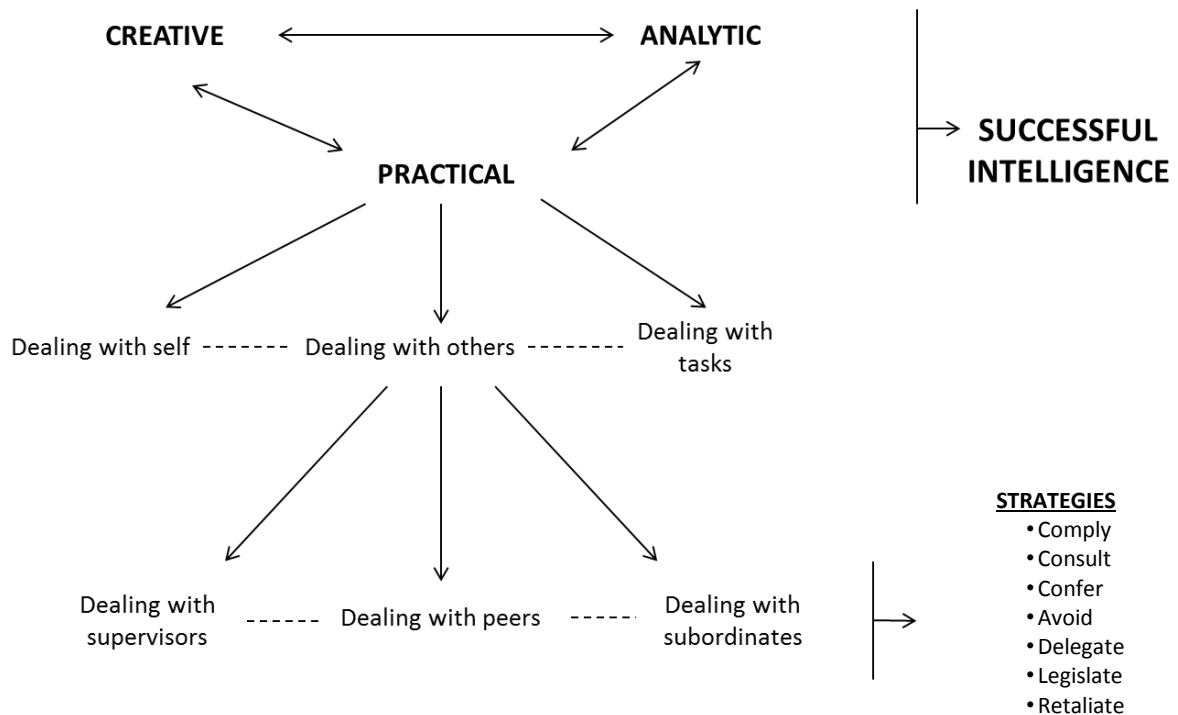
To relate tacit knowledge to the SJT, they describe problem solving situations as consisting of three main parts; the situation, the response strategy and the culture.

The situation concerns the nature of the problem, for example, descriptors of a situation may include uncertainty, insubordination (acting to undermine authority) status exertion (occurs when authority is challenged) and apathy (facing a task one has no motivation or desire to do) (Stemler & Sternberg (2006). Hence, situations are classified into certain motivations or characteristics of that situation. Stemler & Sternberg suggest that there are a finite number of such scenarios. However they also acknowledge the difficulty in noting these as pointed out by Gessner & Klimoski (2006).

The response strategy concerns the way the problem is solved. Stemler & Sternberg (2006) focus upon the strategic decisions people make about how they respond to others in difficult situations (i.e. socially challenging situations in this case). They asked teachers (n=20) about

how they have handled scenarios in the past where they had no formal teaching of how to do so, i.e. tacit knowledge. They targeted teachers ranked as ‘excellent’ by their principals, and also asked these teachers to think of other ways to deal with the scenarios discussed. Teachers reported situations when they dealt with supervisors, peers or subordinates, and from this certain trends were examined and strategies of dealing with these situations were noted (Stemler & Sternberg, 2006). These seven strategies were comply, consult, confer, avoid, delegate, legislate and retaliate. It is noted that each strategy has negative and weak points to it and therefore no one strategy is uniformly best for one situation. These can be seen below in Figure 6.

Figure 6. Triarchic of successful intelligence showing breakdown of ‘dealing with others’ within practical knowledge aspect. (Stemler & Sternberg, 2006).



The seven strategies are seen to work at the level of the ‘dealing with others’ sub categories (either with peers, supervisors or subordinates). The categories of ‘dealing with self’ and ‘dealing with tasks’ are also shown as aspects of practical knowledge. Practical knowledge is shown as one of three main components of successful intelligence. Within theory, practical knowledge is split into a cognitive and a behavioural element, and the cognitive element is further defined into explicit vs. tacit knowledge. These divisions are not shown on the model.

With regards to an SJT used within medical selection, all sub categories of the practical intelligence shown here would be assessed here (i.e. dealing with self, others and tasks) and furthermore, dealing with the all three of the target person groups would most likely be assessed. The strategies would therefore be implemented whilst answering questions in terms of cognition as opposed to the behavioural aspect of tacit knowledge. However, as noted, behavioural consistency ideas and behavioural intention ideas suggest this is a good precursor for later behaviour (Motowidlo, Hooper & Jackson, 2006).

The third aspect of tacit knowledge is culture. This element concerns the context in which the situation unfolds (for example, personal, educational or business). The success of strategy choice and implementation are dependent upon the nature of the situation, i.e. the culture. Sternberg and colleagues (Sternberg, 1997; Sternberg et al., 2000; Sternberg & Hovarth, 1999) conceptualise tacit knowledge as being based upon three features relating to the conditions it is acquired under, the structural representation and the conditions of its use. Firstly, tacit knowledge is acquired with little or no formal instruction and experience is the main form of learning when learning is not the primary objective. When knowledge-acquisition procedures are not activated in everyday life (as they are often facilitated by formal educational environments) the likelihood increases that some people will fail to acquire knowledge (Stemler & Sternberg, 2006). Hence there are individual differences in tacit knowledge based upon the learning environment and capability of learning without facilitation of knowledge acquisition processes such as selective encoding, selective combination and selective comparisons.

Secondly, tacit knowledge is viewed as procedural; knowledge of how to perform a task in a situation. Tacit knowledge can be considered a subset of procedural knowledge that is drawn from personal experience. Much like procedural knowledge, it is often hard to explain how or why one knows a certain behaviour or skill (Anderson, 1983). It is suggested by Sternberg that this is due to such knowledge being based upon condition-action pairings, such as 'If – I am in situation - and – an event happens – and – another event happens- then I will do this – because of this reason.' A type of logical algorithm is used and this allows for behaviour to be reasoned, explained and implemented through tacit knowledge, although this may not always be conscious to the individual.

Thirdly, the final feature of tacit knowledge is that it is relevant to an individual's goals. People are more likely to learn tacit knowledge both from their own practical or another's experiences and advice if the information is goal relevant (Stemler & Sternberg, 2006).

Validity evidence suggests that practical knowledge as measured by tacit knowledge shows low, non-significant or negative correlations with academic intelligence (i.e. typically known as cognitive ability measures. Wagner, 1987; Wagner and Sternberg, 1990; Sternberg et al., 2001). It also appears to be distinct from personality (Wagner and Sternberg, 1990) and as this distinct concept goes on to show correlations with important outcome criteria such as salary, years of management experience and whether a manager worked for a company on the Fortune 500 list (Wagner 1987; Wagner and Sternberg, 1985).

Further inspection of the literature reveals that there is substantial criticism for the concept of practical intelligence as an underlying knowledge affecting performance upon the SJT (e.g. Gottfredson, 2003). Gottfredson (2003) notes that practical intelligence '...claims rest primarily on the illusion of evidence, which is enhanced by the selective reporting of results' (Gottfredson, 2003, p.343) and she goes on to further criticize the theory and lack of clear description and/ or narrative. McDaniel and Whetzel (2005) further argued against the idea of tacit knowledge in the SJT by criticizing the measures used by Sternberg, the lack of evidence for a general factor throughout SJTs (i.e. through factor analysis) and the introduction of moderators (such as instruction form alterations) removing the identifiable practical intelligence result.

In conclusion, the idea appears plausible in theoretical terms and the extensive explanations described above, however, its conceptual involvement in the SJT is debatable and highly criticised from empirical research results and therefore is now somewhat discredited as a construct of interest, regardless of the extensive past literature surrounding practical intelligence.

The idea of an abstract intelligence being assessed by the SJT has been referred to within the literature since the investigation of tacit knowledge. However, whilst this knowledge store may not be supported by empirical research, the idea that the SJT measures beyond procedural knowledge is still a research direction of interest. Further inspection of the literature modelling intelligence needs to be examined in order to identify an area/ construct for further investigation.

Gaps in the Literature: Research Directions

Following the above literature review, the gaps in the research will now be summarised below, firstly within the area of construct and criterion validity of the SJT, secondly, regarding the theory and modelling of the SJT, and finally within medical selection itself.

Construct and criterion validity

There have been attempts to explain the psychological theory surrounding the SJT, from both social and cognitive psychology stances. There are models of the SJT by McDaniel et al., (2006) and later from Motowidlo & Beier (2010). McDaniel et al., (2006) have reported a concise and integrative model that combines results from numerous studies into the SJT and the construct validity of this measure as an approximation of SJTs in general. Motowidlo & Beier (2010) have focused more upon the explanation of personality (a construct) through the SJT and how this affects performance on the measure through the new concept of an Implicit Trait Policy (ITP). Therefore, both these models have contributed to the field.

These models make up a small research base tackling the theory of the SJT that requires exploration, evaluation and expanding. There is acceptance that there is still a lot to learn about the SJT (Lievens, Petters & Schollaert, 2007) and the first point of call is to further understand what the SJT is measuring (construct validity). McDaniel et al., (2006) have produced a coherent model which includes the main constructs of interest when examining SJT construct validity (i.e. cognitive ability, conscientiousness, agreeableness and emotional stability) that is based upon empirical evidence and research. Consistent correlations between these constructs and SJT performance results have allowed for this model to be produced. There are indirect and direct effects of these constructs upon SJT and job performance through factors including education, training and experience. However, the authors themselves note that the model is restricted by using so few constructs and they note that there is much development needed within this area (McDaniel et al., 2006). They state that SJTs can and do measure many constructs and that whilst these can alter according to the specific SJT that still ‘much more work needs to be done to target the constructs measured by the SJT’ and they ‘encourage more theory and research concerning other constructs that are or can be assessed by the SJT’ (McDaniel et al., 2006, p.198). This is not a view they are alone in and other researchers have similar suggestions, for example, Motowidlo, Hooper & Jackson (2006) note the possibility of developing SJTs to measure constructs different from the procedural knowledge construct that is typically presumed to be measured by the SJT, and Weekley & Ployhart (2006) note the importance of the need to know more about the

construct validity of the SJT. Furthermore, Schmitt and Chan (2006) note the benefits of assessing a broad array of psychological measures against the SJT (Oswald, Schmitt, Kim, Ramsay & Gillespie (2004) which included measures of continuous learning (efforts to learn outside the curriculum) and interest in academic learning (positive correlates were found between both measures and the SJT).

This procedural knowledge concept has been broken down and investigated somewhat into the 'practical intelligence' and 'tacit knowledge' literature and this has been suggested as a general factor throughout all SJTs. However ideas have been widely discredited and there is need to look to other psychological concepts that may explain SJT performance. Furthermore, Lievens & Patterson (2011) note the first objective of using simulations in a high stakes selection is typically concerned with broadening the constructs measured. Hence, this topic is at the forefront of current research and in need of exploration.

The criterion validity of the SJT predicting job performance is reported as a moderate relationship (Lievens & Patterson, 2011; McDaniel et al., 2006) yet there may be more to improve regarding the measure and other psychological concepts that could be involved within performance upon the SJT. Ryan & Ployhart (2013) note that two major issues within SJT research are the construct and the criterion validity of the SJT. There is scope for this relationship between job performance and SJT performance to be further investigated and existing evidence of incremental validity over high fidelity assessments when examining actual job performance (e.g. Lievens & Patterson, 2011) to be reproduced and replicated in different settings and fields. The nature of the SJT as a methodology, in that it can differ dramatically between each measure designed due to design and purpose factors, means that research needs to tackle numerous different fields and SJTs to establish a more general theory of SJTs as a whole.

Theory and Modelling the SJT

There is an opportunity to attempt to produce a coherent model and integrated explanation of the SJT including the social, cognitive and intelligence literature, introducing new constructs of interest to expand knowledge of the construct and criterion related validity of the SJT.

From the literature review, it appears that the empirical research continues at a vast rate and there is an excess of this in comparison with the amount of theoretical and conceptual models and explanations of the processes at play within the SJT. Weekley & Ployhart (2006) note

that the absence of robust theory is potentially the largest limitation to the SJT. There is a need to bring conceptual modelling and theory in line with the empirical studies.

As noted, McDaniel et al., (2006) emphasise the need for the development of their model with new constructs and further empirical work on this model in particular. They have stated the restrictions within their model and further construct exploration across different fields, different SJTs and using different constructs may lead to development of the theory. This would ideally lead to an increase in the power of the correlations seen between the constructs of interest and SJT performance. New variables and aspects of cognition may be included in future studies to add to this model and further explain SJT performance. With the evident versatility of SJTs (for example, according to differing instructions, responses and development paths) there are numerous studies that could investigate new constructs of potential interest for differing SJTs and fields.

With regards to the Motowidlo & Beier model (2010) they report empirical evidence for their model, and there is potential now for this to be further investigated, added to or modified, using new cognitive variables and the psychological theory covered in this literature view. This new concept of an ITP is clearly of major interest to the SJT literature and needs further investigation and development. There are extensive aspects of other psychological theories and ideas that may influence the origins, development, expression and manipulation of the ITP as noted in the literature above (p.58-63).

As an aside, there also appears to be a lack of research upon cross cultural studies and following that there appears little research upon adverse impact, beyond the assessment of gender and race as factors affecting SJT performance (Lievens, Peeters and Schollaert, 2008) and these extraneous factors could be incorporated into future models. To summarise, as a relatively new measure within personnel selection there are still many aspects of the SJT that need further investigation and development, both empirically and theoretically. The main themes appear to be investigating new constructs and further explaining the construct validity of the SJT in broad terms, and the development of theory to explain both this and criterion validity of the SJT to keep up with the amount of empirical research reported.

Medical Selection

With the criterion and construct validity of the SJT fluctuating according to format, length, instructions and target occupation of the SJT, there is a need to investigate the key variables (e.g. cognitive ability, personality) involved together within a medical education/career

assessment setting. As noted, the medical field is a high pressure, high stakes and high cost area that has many stakeholders and people with vested interests in the occupations, sustainability and success (for example, patients, the public, the government and the medical professionals themselves). It is of utmost importance that the correct individuals are chosen for these careers, and that the measures used to do so are identifying individual differences in applicants/ workers through reliable, valid and well understood methods.

As noted above, there is a need to investigate further constructs of interest to the SJT. The medical field is in fact a very unique career path, where individuals are required to perform to a consistently outstanding standard in both academic and personal skill areas, as well as maintain a large basis of knowledge whilst adapting to different situations and constantly learning new information and evolving to the progress with the career. Therefore, certain constructs related to such performance may produce correlates and interesting findings for an SJT designed for medical selection, that may not be found or relevant to other career based SJT measures. There is scope to introduce new variables into SJT studies within the medical field, alongside the existing key variables (for example, the four personal traits in the McDaniel et Al., 2006 model) to see if these impact on the SJT performance and could move towards higher correlations of the SJT with job performance, or even higher correlations of the constructs to the SJT through mediating or moderating variables (i.e. assessment of criterion and construct validity of the SJT). Recent research has begun to identify new predictors of performance in medical school for example, the consideration of emotional intelligence as an antecedent (Libbrecht, Lievens, Carette & Côté, 2014).

Summary

Broadly speaking, the investigation of current constructs, the introduction of new constructs and the unification or development of theoretical models and explanations of the SJT are the main directions for research to move in.

These points should all be focused on within different areas of selection (i.e. in this case research focuses upon the current, relevant and highly important medical field). It is therefore necessary to investigate the SJT further within this field and to attempt to explain the SJT in more detail. The selection and assessment method of the SJT is used widely throughout medicine to choose, rank and assess doctors. Hence it is of utmost importance that we fully understand and investigate this measure in the correct context. A brief summary of the thesis aims can be seen below followed by the specific research questions.

Summary of Aims

Following examination of the intelligence, personnel selection, SJT and related psychological theoretical literature it is now logical to define the main issues that are discussed in this thesis and the research questions for each study contained in this thesis.

There are three main studies within this thesis. They tackle two of the major issues within SJT research according to Ryan & Ployhart (2013), the construct and the criterion related validity of the measure. Furthermore, there is the introduction of a new construct of potential interest to the SJT. These studies as a whole aim to further develop the modelling and theory behind the SJT and the construct validity of this tool.

Broadly, the studies firstly investigate the current basic modelling of the construct validity of the SJT whilst introducing new constructs for investigation. Secondly, a new construct of interest to the SJT (formal operational thought) is introduced and validated, prior to examining this new construct within a medical sample alongside an SJT and an actual job performance measure.

These studies are summarised in terms of theoretical rationale for each study below. Full introductions, methods and reports of the studies follow in the remainder of the thesis, but for initial clarity the summary of studies is included here as a foundation for the later individual study accounts. The main research questions for this thesis that underlie the aims are reiterated below.

Research Questions

Research Question 1:

To what extent can variables from existing SJT theory explain the construct validity of the SJT within a medical selection context?

Research Question 2:

Can the inclusion of additional non-cognitive variables assessing individual differences further explain the construct validity of the SJT over and above variables from existing SJT theory?

Research Question 3:

Can the inclusion of an abstract intelligence variable assessing higher level thought further explain the construct validity of the SJT in a medical selection context over and above variables from existing SJT theory?

Research Question 4:

Can the inclusion of an abstract intelligence variable assessing higher level thought further explain medical job performance independent of the SJT in a medical selection context?

Research Question 5:

How can existing theoretical literature surrounding the SJT's construct and criterion validity be expanded and linked to develop an integrative conceptual model of the SJT?

CHAPTER 4: CONSTRUCT VALIDITY OF THE SJT

Study 1: Situational Judgment Test Performance and Demographic, Personality and Non- cognitively Orientated Variables in Medical Students

Acknowledgements and Role in Study

This project was conducted by the Work Psychology Group (Ashbourne, Derbyshire) in collaboration with the University of Cambridge on behalf of the Improving Selection to the Foundation Programme (ISFP) Project Group. The ISFP Project Group, Chaired by Professor Paul O'Neill, includes representatives from fifteen organisations: the Academy of Medical Royal Colleges (AoMRC), the British Medical Association (BMA) Medical Students Committee, the Conference of Postgraduate Medical Deans (CoPMED), the four UK Departments of Health, the General Medical Council (GMC), the Medical Schools Council, the Northern Ireland Medical and Dental Training Agency (NIMDTA), NHS Education for Scotland (NES), NHS Employers (NHSE), the Scottish Board for Academic Medicine, the UK Foundation Programme Office (UKFPO) and clinical tutors. The ISFP Project is funded by the Department of Health.

The Co-directors of the project who developed a Situational Judgement Test for ISFP were Professor Fiona Patterson and Dr David Good. The core project team consisted of Professor Fiona Patterson, Vicki Archer, Dr Maire Kerrin, Victoria Carr, Louise Faulkes and Helen Stoker in conjunction with the Medical Schools Council project team including Professor Paul O'Neill, Professor Tony Weetman, Dr Katie Petty-Saphon, Dr Denis Shaughnessy, Siobhan Fitzpatrick and Amy Stringe.

This study took a lot of effort, planning and implementation over a long period of time from the research team. Following the commencement of my PhD I worked alongside the Work Psychology Group in the study and my primary role within this study was to run the data analysis on the data set following the data collection by the Work Psychology Group. I produced a report and discussion of the data using this analysis and the results can be seen in this chapter.

Introduction to Study

This study investigated performance upon a Situational Judgment Test (SJT) and whether an individual's personality and the non-cognitive measures of Need for Cognition and Occupational Self-Efficacy contribute to predicting SJT performance (construct validity). This study tackled Research Question 1 (how well can variables from existing theory explain the construct validity of the SJT) and Research Question 2 (how well can additional non-cognitive variables assessing other individual differences explain the construct validity of the SJT over existing theory?)

McDaniel et al., (2006) designed a model of the SJT that considered cognitive ability and three main personality factors as the main constructs assessed by the SJT. This study was designed to investigate the construct validity of the SJT from the McDaniel et al., (2006) model and to introduce two new constructs of potential interest to the SJT (Need for Cognition and Occupational Self-Efficacy) based upon the Murphy (2012) framework of individual differences in personnel selection. The investigation of these new constructs alongside an SJT was examined for the first time in this study as potential variables that may affect medical job performance and success. This study followed direction from previous literature, by the examination of new constructs that may affect SJT performance, and by using constructs of a non-cognitive nature.

This thesis was designed to investigate theory surrounding and involving the SJT, whilst considering the constructs involved and attempting to improve the explanation of the measure. It was therefore logical to assess the broader model by McDaniel et al., (2006) as opposed to focusing upon the specific personality expression route and the newly developed ITP concept from Motowidlo & Beier (2010). Due to external constraints (ethical boundaries regarding past examination results during University and access to current examination papers assessing academic ability) the variable of cognitive ability was unable to be examined. This was due to reasons beyond the investigators control, for example, data collection timing and the issue of access to the large medical student sample or their previous academic results. However, as the sample consisted of medical students it can broadly be assumed that they are all of a high cognitive ability due to the rigorous academic demands for acceptance into medical school. These issues are further discussed in the following section below.

Theoretical Basis for Study

Numerous researchers note the importance of the need to gain greater understanding of the construct validity of the SJT (for example, Weekley and Ployhart, 2006; McDaniel et al., 2006; Ryan & Ployhart, 2013; Schmitt and Chan, 2006). Although there has been an increase in research and literature surrounding the SJT in recent years, there has still been relative little success in targeting what the SJTs actually measure; empiricism has left theory and explanation somewhat behind. It is a common theme throughout the literature review that research has been conducted and some basis of theory established, but that much is left unexplored with comments that more research into the theoretical underpinnings of the measure is necessary. In particular, explaining the SJTs in general terms and their construct validity requires further attention.

SJTs have been known to correlate with certain personality traits, cognitive ability, experience and knowledge (McDaniel et al., 2006). These constructs are considered the four main constructs affecting performance upon the SJT (McDaniel et al., 2006) based upon a number of empirical studies showing varying, yet consistent, correlations between SJT performance and these four constructs. To investigate the construct validity of the SJT in study 1, the McDaniel et al., (2006) model was chosen as a recent and integrative model of construct validity of the SJT. This was chosen an appropriate base on which to add further additional new constructs. (This model can be seen on page 84).

Furthermore, medical selection has become better understood, investigated and improved from insights from the generic personnel selection field over recent years, which allows it to now use new methods and approaches, such as the SJT (Patterson, et al., 2012). A noted example of this is the investigation into both non-cognitive as well as cognitive competencies and variables within medical selection, and both high fidelity and low fidelity simulations are being used to assess these (e.g. Assessment Centers and SJTs respectively). It is therefore important that we understand the SJTs being used within medical selection, and attempt to fully explain the construct validity of these measures. It is suggested here, from the McDaniel model, that the SJT already assesses some non-cognitive variables (i.e. personality). The idea that the SJT measures non-cognitive constructs is also now supported by more recent literature from Murphy's (2012) individual differences framework.

As noted in the literature review, the Murphy (2012) framework is broadly split into abilities, personality, interests and self-concepts. These four domains encapsulate areas of individual

differences within selection and all are expressed in behaviours and choices that are directly relevant to work organizations (Murphy, 2012). With the exclusion of the abilities domain, all three can be mapped onto the variables within this study. The personality domain refers to the assessment of behavioural consistencies from personality traits. The interests domain refers to responses of liking to particular situations or circumstances, and self-concepts represent beliefs about ourselves, including perceptions of self- efficacy.

The four main constructs in McDaniel's model measured by the SJT are noted to be the three main personality factors (agreeableness, conscientiousness and emotional stability) and cognitive ability. These are assessed by the SJT, both directly and indirectly through experience and education and knowledge of the job (general and/or specific) according to the model. It has also been established that links between these constructs and the SJT performance are subject to effects from variations across SJTs, such as the response instructions, the applicable occupation and the fidelity of the measure (McDaniel et al., 2006). It is therefore important to examine the SJT and these noted predictive variables affecting performance within a high stakes setting, and within medical selection in particular using an SJT specifically developed for this purpose.

Therefore, a study was designed to investigate the main constructs of the SJT as established from the McDaniel et al., (2006) model and supported by the Murphy (2012) framework. This was designed to test the noted elements of the above model within a medical student sample using an SJT designed for medical selection. Furthermore, McDaniel et al., (2006) together with many other researchers, encourage researchers to investigate theory behind the SJT further and in particular, other constructs that are, or may be assessed by the SJT.

The personality factors (the Big Five) are present in both the McDaniel et al., (2006) model and the Murphy (2012) framework. The addition of two new constructs that have not been assessed in relation to the SJT are introduced. The additional measures are relevant to one's occupation and the capability an individual feels about behaviour at work (Occupational Self-Efficacy) which can be seen as a measure of a self-concept (Murphy, 2012) and also whether one enjoys thinking about challenging concepts, i.e. whether one enjoys cognition and challenge (Need for Cognition) which can be seen as a measure of interests; i.e. how does one respond or feel about the act of thinking/ engaging in thought (Murphy, 2012). These concepts are introduced with the aim of developing the McDaniel et al., (2006) model further and expanding the constructs that are assessed and may be related to the SJT in a medical

setting, i.e. the construct validity of this SJT. These two additional constructs have not been investigated alongside the SJT before and due to the nature of these variables, it is suggested that they will be positively associated with SJT performance.

Medical students were asked to assume the role of a training Foundation Year 1 doctor and therefore assume that the SJT was being used at this gateway for selection purposes at this level. This stage was chosen as it is one of the gateways within medical selection when the SJT is used as an assessment method (i.e. for job specialisation at this point and ranking of students for this purpose). The measures used in this study are as follows: Goldberg's (1992) Big Five Personality Measure, Cacioppo and Petty's (1982) Need for Cognition Scale and Schyns and Von Collani's (2002) Occupational Self Efficacy Scale. Demographic information about participants was also collected alongside their SJT participation. The reasons for the variable choice and how these constructs are theoretically rooted in previous literature can be seen below.

Personality

As it has been noted in the literature review, it is widely established that there are links between the SJT and personality factors, and in particular the three main factors of agreeableness, conscientiousness and emotional stability (McDaniel et al., 2006) and the links between these concepts are therefore widely researched (noted in Patterson, Ashworth, Zibarras, Coan, Kerrin & O'Neill's, 2012 evaluations of the SJT). Factor α (stability/communion) has been defined as socialization processes, conformity and the extent to which one is consistent in motivation, mood and social interactions (Mount & Barrick, 2012) and this factor was noted to encompass the three factors of agreeableness, conscientiousness and emotional stability).

It is important therefore to include these personality traits to establish if there are links shown between personality and the SJT in this format using a medical student sample. Personality is also a main domain in the Murphy (2012) framework concerning the assessment of individual differences in relation to personnel selection. Furthermore, established links can be used as a statistical basis for investigating the Need for Cognition and Occupational Self-Efficacy variables (i.e. within regression modelling as a first block variable).

New variables (Need for Cognition and Occupational Self-Efficacy)

As seen in the literature review, there is extensive literature investigating the SJT in relation to cognitive ability, and the non-cognitive orientated variables of personality and experience.

There is less investigation into other non-cognitive ability related constructs that the SJT may in fact be measuring. Within medical selection, there is great potential to assess non-cognitive ability through SJTs and from the literature it appears that examination of other constructs is necessary.

Research is beginning to move in this direction, for example, Patterson et al., (2012) also note that empirical research suggests that SJTs may be related to a variety of constructs and that further research is required to establish greater conceptual clarity regarding SJTs. Following this, the inspection of non-cognitively orientated variables led to the targeting of two in particular that could possibly further explain the construct validity of the SJT. These are the Need for Cognition (Cacioppo & Petty, 1982) and Occupational Self-Efficacy (Rigotti, Schyns & Mohr, 2008) scales. These two concepts relate to how much an individual enjoys cognitive challenge and thought, and how capable one feels about using the thought to translate it into work related action/success, respectively. These are suggested as two non-cognitively oriented constructs that may be related within the medical field in particular due to the highly challenging nature of the job role; i.e. the intense thought required and the need for confidence (self-efficacy) within an individual's actions and job performance due to the large personal and professional consequences for poor or incorrect performance (for example, risk of life following a practice error or legal proceedings).

Need for Cognition broadly defines the 'tendency for an individual to engage in and enjoy thinking' (Cacioppo & Petty, 1982, p. 116). Evidence suggests that people scoring highly (i.e. those who enjoy challenging thought) are more likely to evaluate, organise and elaborate upon information they have been exposed to than those who score lower on Need for Cognition measures (Cohen, 1957). Higher levels of Need for Cognition are linked with intrinsic motivation for cognitive processing and metacognition (Cacioppo & Petty, 1982) as well as confidence in decision making (Levin, Huneke & Jasper, 2000) and enhanced problem solving (Unnikrishnan Nair & Ramnarayanb, 2000; cited in Madrid & Patterson, 2014). The Need for Cognition variable maps onto the 'interests' domain from the Murphy (2012) framework. The interest in one's thought and engagement of thought is expected to relate directly to the SJT performance due to the affect felt from the level of interest self-reported in cognition. Hunter & Hunter (1984) showed that interests are not generally strong predictors of work performance, although Mount & Barrick (2012) more recently have stressed the importance of investigating the relatively unexplored area of interests and

personality traits to reveal better predictions, and both will be investigated in this study together.

Occupational Self-Efficacy refers to an individual's beliefs about their ability to successfully complete a task. More specifically, Occupational Self-Efficacy concerns 'the competences that a person feels concerning their ability to successfully fulfil the tasks involved in his or her job' (Rigotti, Schyns & Mohr, 2008, p.238). Evidence suggests that individuals high on self-efficacy tend to perform to a higher standard (Grandey, 2000). Low self-efficacy can also hamper both the frequency and the quality of behaviour-environment interactions, whilst high self-efficacy can facilitate both (Martin, Carlson & Buskist, 2007). The Occupational Self-Efficacy variable directly maps onto the 'self-concept' domain from the Murphy (2012) framework. This may also be known as a 'core self-evaluation' (CSE) and refer to broad judgments that people make about their basic worth and ability to influence events (Murphy, 2012). Occupational Self-Efficacy is a CSE as it reflects beliefs about one's competence and capabilities, and a general sense that life will turn out well. Due to the nature of this research, the measure of self-efficacy chosen reflects behaviour in a work setting, i.e. occupationally related. Judge (2009) states that CSEs predict job performance, along with other outcomes such as motivation or job satisfaction. CSEs also add incremental validity over the dimensions of personality in predicting job performance (Judge, Erez, Bono & Thoresen, 2003) and hence this will be investigated.

At the time of this research there were no other papers investigating these concepts in relation to the SJT within the medical field. In the medical field it is held that through the rigorous screening and selection procedures, based mainly upon cognitive intelligence and academic achievement to enter university that those in this position are of a consistently high level of cognitive ability. For example, applicants are typically required to have achieved at least three A grade A-levels (usually including a mixture of science subjects and mathematics) and in the absence of these academic qualifications are asked to have completed other academic qualifications such as Scottish Highers, an undergraduate degree or a foundation entry medical course. In such a situation, where academic success is held as a standard throughout the sample, it is therefore important to investigate the non-cognitive constructs that the SJT may be measuring and showing individual differences through as opposed to focusing upon cognitive ability which is evidently strong throughout a medical student sample.

It is important to recognise that cognitive ability was not measured in this study. It would have been preferable to have been able to measure this as it is a main factor in both the McDaniel et al., (2006) model, and the Murphy (2012) framework. However, due to the timing of the data collection, the researchers were unable to administer a test of cognitive ability. Furthermore, due to the ethical and institutional constraints and privacy concerns relating to information regarding the student's cognitive ability (i.e. academic history or university career success) it was not possible for the researchers to establish access to this information from the respective universities. However, as entry and progression through medical school already relies upon extensive and consistently excellent academic achievement and intelligence it can be reasonably assumed that cognitive ability is generally high with somewhat limited variation across individuals. The inclusion of the other factors in the model, such as cognitive ability, may be used alongside these new variables as an option in future research.

The variables examined in this study were therefore personality traits, the Need for Cognition, Occupational Self-Efficacy and an SJT measure with demographic information in a medical student sample.

Hypotheses

Hypothesis 1: Personality and SJT performance

As has been noted in the literature review, it has been widely established that there are links between the SJT and personality factors, and in particular the main three factors of agreeableness, conscientiousness and emotional stability as noted in the McDaniel et al., (2006) model and that directly map onto factor α of personality (stability/ communion) which has been defined as socialization processes, conformity and the extent to which one is consistent in motivation, mood and social interactions (Mount & Barrick, 2012). These three facets of the factor α which relate to conformity and consistency in mood and motivations are suggested as more likely traits to be shown in effective medical care practices as well as the SJT, as opposed to the extraversion and openness to experience domains which map onto a factor β , representing plasticity, non-conformity and searching for new experiences.

Hypothesis 1 states that there will be significant positive association found between the three main personality factors (agreeableness, conscientiousness and emotional stability) and SJT performance.

Hypothesis 2: Need for Cognition and SJT performance

Evidence suggests that people scoring highly (i.e. those who enjoy challenging thought) on the Need for Cognition scale are more likely to evaluate, organise and elaborate upon information they have been exposed to when compared to those who score lower on Need for Cognition measures (Cohen, 1957). This suggests that those who score highly on this measure will have a greater and more detailed amount of knowledge that is more organised, logically represented and explained cognitively than those who score lowly on the scale. Therefore, one would expect that these ‘high thinkers’ would have more declarative and more procedural knowledge than the ‘low thinkers’, hence they would be expected to score more highly upon the SJT.

The interest in one’s thought and engagement of thought is expected to relate directly to the SJT performance due to the affect felt from the level of interest self-reported in cognition. Interests are strongly related to job persistence and work involvement (Campbell & Johansson, 1966) which suggests that those who are high on the Need for Cognition scale will report higher SJT scores due to increased involvement, engagement and persistence with the challenge and task of the SJT.

The study was designed to attempt to explain further variance in SJT scores beyond the McDaniel et al., (2006) model by including the new constructs of Need for Cognition and Occupational Self-Efficacy. It is therefore suggested that these non-academic constructs will have incremental validity over the personality variables when explaining variance in SJT performance due to the non-academic intelligence nature of the SJT.

Hypothesis 2 states that there will be a significant positive relationship between SJT performance and Need for Cognition scores, and that Need for Cognition scores will have incremental validity over the personality variables in explaining variance in SJT performance.

Hypothesis 3: Occupational Self-Efficacy and SJT performance

Evidence suggests that individuals high on self-efficacy tend to perform more highly on measures of ability (Grandey, 2000). When applied to an organisational setting, self-efficacy can mean that people feel more capable to act, and maintain actions in the face of adversity in their place of work. Self-efficacy is based on perceptions of our environment and positive reinforcement of such behaviours and of our actions carried out in the situation. Low self-efficacy can also hamper both the frequency and the quality of behaviour-environment

interactions whilst high self-efficacy can facilitate both (Martin, Carlson & Buskist, 2007). Therefore, it is expected that those with high occupational self-efficacy will feel more capable to act, feel more confident in doing so and feel more in control of themselves in the situation at work. The perceptions they have made of behaviour and positive reinforcement of correct behaviours will further increase self-efficacy.

Occupational Self-Efficacy is a Core Self-Evaluation (CSE) from the Murphy (2012) framework. CSEs are found to predict job performance and motivation, among other outcomes (Judge, 2009) which suggests that they would also predict SJT outcomes as a behavioural indicator for later job behaviour.

The study was designed to attempt to explain further variance in SJT scores beyond the McDaniel et al., (2006) model by including the new constructs of Need for Cognition and Occupational Self-Efficacy. It is therefore suggested that these non-academic constructs will have incremental validity over the personality variables when explaining variance in SJT performance due to the non-academic intelligence nature of the SJT. CSEs have been found to add incremental validity over and above the big five domains of personality when predicting job performance (Judge, Erez, Bono & Thoresen, 2003).

Hypothesis 3 states that there will be a significant positive relationship between SJT performance and Occupational Self-Efficacy, and that Occupational Self-Efficacy scores will have incremental validity over the personality variables in explaining variance in SJT performance.

Method

Study Design

The study was a cross sectional within-participants design with all participants taking part in all measures. Analyses were conducted within and across the measures. Descriptive statistics are reported for all measures before hypothesis investigation.

Participants

Participants were medical students from five medical schools across the United Kingdom (Cambridge, Sheffield, Manchester, Cardiff and Keele). Initially data from 453 participants was collected, but after omitting cases due to lack of unique student identification numbers (referred to as RA-ID numbers) and/or lack of scoring on either the questionnaire measures

and/or the SJT, $n = 399$ (male = 139, female = 253, 7 did not specify) with a mean age of 24.68 ($sd = 2.98$) and a range of ages from 22- 43 years old.

Tests were set in the context of the Foundation Programme (two years training post medical qualification) with applicants assuming the role of a Foundation Year 1 doctor (first year job after graduating from medical school). Participants were identified between according to their RA-ID numbers that were given before the SJT and for the questionnaire.

General Procedure

Participants completed a questionnaire and participated in a Situational Judgment Test (SJT). Demographic information was also collected from each participant. Detailed descriptions of the individual measures can be seen below.

Task Procedures

Questionnaire

The questionnaire comprised three sections, each employing a different measure and intending to assess a different aspect of the participant's character or abilities. Questionnaires were completed in a paper and pencil format. The three sections can be seen explained below (personality, Need for Cognition and Occupational Self-Efficacy variables).

Big Five Personality

The Big Five Personality traits were measured using Goldberg's 1992 scale. Goldberg breaks personality down into five major factors and these are referred to as I. Extraversion; II. Agreeableness; III. Conscientiousness; IV. Emotional stability and V. Openness. These domains are further broken down into fifty factors that can then inform the degree to which a certain trait is present in one's personality. For each factor there were ten further factors representing it and these were marked accordingly upon a nine point unipolar scale. The breakdown of these points is seen in Appendix II.

The scale ranges from 1 to 9 with the two contrasting terms at either end; 1 and 9 represent 'very much like me' and 'not like me at all' respectively with 5 referring to 'neither.' The maximum score for each factor is therefore one of 9 (top of scale) multiplied by 10 (number of questions per personality factor) resulting in 90. Following this, the minimum possible score for any factor is 1 (bottom of scale) multiplied by 10 (number of questions per personality factor) resulting in 10.

There is extensive endorsement and use of the 5 factor personality idea within personality research and it appears to be a valid and reliable measure with the five noted variables being present consistently within research (e.g. Goldberg, 1992; Norman, 1967; Bond, 1979). The Cronbach's alpha reliability coefficients for the measures of extraversion, agreeableness, conscientiousness, emotional stability and openness were .87, .88, .79, .87 and .80 respectively.

Example items include *'I am the life of the party'* (extraversion), *'I have a soft heart'* (agreeableness), *'I am always prepared'* (conscientiousness), *'I change my mood a lot'* (emotional stability) and *'I have a rich vocabulary'* (openness).

Need for Cognition

Need for Cognition was measured using the 1982 Cacioppo & Petty Need for Cognition scale. This measure assesses the 'tendency for an individual to engage in and enjoy thinking' (Cacioppo & Petty, 1982, p. 116). The scale employed 18 items and is widely used. Participants were asked to note how much or little they agreed with each of the 18 statements. Participants rated themselves on a scale of 1-5 points (1=being extremely uncharacteristic of me, 2=quite uncharacteristic of me, 3=neither uncharacteristic nor characteristic of me, 4=quite characteristic of me and 5=being extremely characteristic of me) and half of the statements were negatively structured. Hence, due to this there was a maximum possible score of 90 and a minimum possible score of 18.

Those scoring highly may loosely be referred to as 'thinkers' but for further clarity they should be thought of as individuals who enjoy engaging in thought and cognitive challenges, enjoys solving problems and applying solutions that they have thought of for problems in their lives. Sadowski & Cogburn (1997) draw links between high Need for Cognition scorers and conscientious personality traits, and furthermore openness to experiences. Other links have been drawn between academic achievement (Sadowski & Gulgoz, 1996) and the ability to sort, process and apply information effectively (Cacioppo & Petty, 1982). Higher levels of Need for Cognition are linked with intrinsic motivation for cognitive processing and metacognition (Cacioppo & Petty, 1982) as well as confidence in decision making (Levin, Huneke & Jasper, 2000) and enhanced problem solving (Unnikrishnan Nair & Ramnarayanb, 2000; cited in Madrid & Patterson, 2014).

There is extensive evidence supporting the instrument's validity and reliability (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein & Jarvis, 1996; Sadowski, 1993; Sadowski & Gulgoz,

1992b). Cacioppo and Petty (1982) also noted that neither social desirability nor test anxiety were problematic biases for the scale.

The Cronbach's alpha reliability coefficient for the Need for Cognition scale was .81.

Example items include '*I would prefer complex to simple problems*', '*I like to have the responsibility of handling a situation that requires a lot of thinking*' and '*thinking is not my idea of fun*'. (The last example item is a reverse coded item).

A complete list of the 18 statements used can be found in Appendix III.

Occupational Self-Efficacy

Occupational Self-Efficacy was measured using the 2002 Schyns & Von Collani Occupational Self-Efficacy scale. Bandura (1977) introduced the concept of self-efficacy and described it as 'people's judgements of their capabilities to organize and execute courses of action required to attain designated types of performance' (Bandura, 1986, p.391). It concerns an individual's beliefs about their ability to successfully complete a task. More specifically, Occupational Self-Efficacy concerns 'the competence that a person feels concerning their ability to successfully fulfil the tasks involved in his or her job' (Rigotti, Schyns & Mohr, 2008, p.238).

The scale was developed with the intention of capturing a level of self-efficacy towards job performance and achievement by Schyns & Von Collani (2002). An 8 item version of the scale was used for this research to avoid participant disaffection and tiring. This short version scale validity was established for use across five European countries, including Great Britain, in 2008, by Rigotti, Schyns & Mohr.

A six item likert scale is used ranging from 1 (not at all true) to 6 (completely true) and participants mark how strongly they think each statement applies to them. There is therefore a maximum possible score of 48 (6, the top of scale multiplied by 8, the number of questions) and a minimum possible score of 6 (1, the bottom of the scale multiplied by 8, the number of questions). High scores suggest a high level of self-efficacy and low scores suggest low levels of self-efficacy.

The Cronbach's alpha reliability coefficient for the Occupational Self Efficacy scale was .86.

Example items of this measure include '*thanks to my resourcefulness, I know how to handle unforeseen situations in my work*', '*if I am in trouble in my work, I can usually think of*

something to do’ and *‘I can remain calm when facing difficulties in my work because I can rely on my abilities’*.

A full list of the statements used can be found in Appendix IV.

Situational Judgment Test (SJT)

Before completion of the SJT participants disclosed demographic data. This information included their RA-ID number for identification purposes within data analysis, gender, ethnicity, religion/belief and nationality. The location of the centre where the test was undertaken was also recorded and participants were asked to provide feedback upon the test administration and material. For example, participants were asked whether the questions were clear and easy to understand and the time taken to complete the SJT.

The SJT used in this case was composed of 64 questions mapped onto the five main domains. These domains were commitment to professionalism, coping with pressure, working effectively as part of a team, effective communication and patient focus. The domains of the SJT were mapped onto a ‘Professional Attributes Framework’ with related behavioural descriptors of these behaviours. These domains were initially established based upon ratings by Foundation Year 1 doctors upon their importance whilst balancing this with their importance within job selection.

The reliability of this SJT as a whole was reasonable ($\alpha = 0.74$, 64 items).

Development of SJT

A job analysis was conducted which then allowed for a panel of researchers (N=7, from the Work Psychology Group) to decide upon 5 target attribute domains to assess in the SJT. This happened prior to study commencement (development in 2009). The domains were decided upon based upon the job analysis and their relative importance within job selection, and also ratings from actual Foundation Year 1 doctors upon the importance of the domains.

A brief description of each domain can be seen below, each with an example of how an individual might be assessed on the relative domain by an SJT:

- *Commitment to professionalism:* Displaying honesty, integrity, awareness of confidentiality and correct ethical procedure, e.g. challenging inappropriate behaviour.

- *Coping with pressure*: Capability to work under pressure and remain resilient. Demonstrates adaptability, e.g. dealing with confrontation.
- *Working effectively as part of a team*: Capability and willingness to work within a partnership and show competence and skill in doing so, e.g. recognising and valuing the skills of a peer worker.
- *Effective communication*: Actively and clearly engages patients and colleagues in equal/open dialogue. Concise and clear in verbal and written communication, e.g. negotiating for a scan from radiology.
- *Patient focus*: Ensuring patient is the focus of care. Understanding shown to patients' needs, e.g. considering that a patient may have different needs from those around them

Items used in the SJT were content valid (established from Foundation Year 2 focus groups and item review) unambiguous, and equally mapped over the target attribute domains.

Response instructions

The SJT was developed in two sections that had two different types of formats.

Section one consisted of 42 items and asked participants to rank answer options in order of the likelihood they are to take that action. Section two consisted of 22 items and asked participants to answer by choosing the one course of action they are most likely to take. All instructions were 'knowledge tendency' instructions and asked what one *should* do, as opposed to 'behavioural tendency' instructions which ask what one *would* do. This was designed with the intention to minimise susceptibility to coaching or faking. Knowledge tendency instructions are more complex and cognitively loaded as well as less susceptible to self-deception and impression management, and therefore to coaching (Lievens, Buyse, Sackett & Connelly, 2012; McDaniel, Hartman, Whetzel & Grubb, 2007; Nguyen, Biderman & McDaniel, 2005).

Results

Questionnaire Measures and SJT

Demographic information, SJT scores and the questionnaire data were all entered and coded as necessary.

A table of the means, standard deviations, correlations and reliabilities was constructed using the scores for each section of the questionnaire measures (each of the five personality variables, Need for Cognition scores, and Occupational Self-Efficacy scores), the demographic variable of age and the SJT scores. The results are shown below in Table 3.

Table 3. Means, standard deviations, correlations and reliabilities of variables.

Variables	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9
1. Extraversion	65.91	10.02	0	90	(.87)								
2. Agreeableness	76.47	8.50	0	90	.353**	(.88)							
3. Conscientiousness	72.37	9.78	0	90	.291**	.526**	(.79)						
4. Emotional stability	64.27	10.97	0	90	.419**	.526**	.381**	(.87)					
5. Openness	70.65	8.50	0	90	.415**	.547**	.459**	.385**	(.80)				
6. Need for Cognition	64.31	8.15	18	90	.237**	.083	.193**	.083	.313**	(.81)			
7. Occupational self-efficacy	34.23	5.08	0	48	.324**	.233**	.221**	.352**	.330**	.281**	(.86)		
8. Age	24.68	2.98	-	-	.080	.084	.077	.049	-.004	.123*	.151**	-	
9. SJT	872.42	34.47	0	932	.027	.001	.078	.065	.089	.042	.041	.012	(.74)

(N=399, * p<.05 **p<.01

Reliabilities are displayed in parentheses on the diagonal. Minimum and maximum possible scores for measures noted to give context to mean values).

All of the personality variables, Need for Cognition and Occupational Self-Efficacy variables were significantly positively related to each other, with the exception of the Need for Cognition and the agreeableness and emotional stability traits. None of these variables were significantly associated with the SJT scores.

Questionnaire measures and SJT domains

The SJT was further categorised into five main domains.

A linear correlation was conducted between the questionnaire measures and each of the SJT domains. The results of the correlation analyses can be seen below in Table 4. There is a distinction in the table between the domains of the SJT (top half of table; numbered 1-5) and the questionnaire measures (lower half of table including the personality traits, Need for Cognition and Occupational Self-Efficacy variables).

Table 4. Correlation matrix of questionnaire measures, SJT domains and total SJT score.

Variables	1	2	3	4	5
1.Commitment to professionalism	-				
2.Coping pressure	.344**				
3.Working as a team	.411**	.320**			
4.Effective communication	.200**	.267**	.273**		
5.Patient focus	.290**	.243**	.386**	.202**	
Total SJT score	.698**	.672**	.755**	.571**	.590**
Extraversion	-.026	-.004	-.085	.024	.029
Agreeableness	-.024	.058	.002	.000	-.043
Conscientiousness	.027	.022	.115*	.042	.042
Emotional stability	-.033	-.057	-.055	-.035	-.031
Openness	-.063	-.096	-.055	-.016	-.062
Need for cognition	.096	-.004	-.034	.075	.018
Occupational self-efficacy	-.010	-.047	-.057	-.036	.032

(N=399, *p<.05 **p<.01)

Correlations are seen between all the domains of the SJT (which is unsurprising as these are not uni-dimensional and constructs are likely to measure more than one domain). There are also, as to be expected, strong significant correlations (all domains $p < .01$) between the domains of the SJT and total SJT score, with the 'working effectively as part of a team' domain score showing the highest correlation.

Significant correlations are also seen between conscientiousness scores and the working as part of a team domain scores. There are no other significant correlations between any questionnaire measures and any SJT domains. This was to be expected as the SJT was not designed to give results for each domain, but to give a total score on the measure.

Personality, Need for Cognition, Occupational Self-Efficacy and SJT performance

A hierarchical regression was conducted to investigate the hypotheses, with the SJT scores as the dependent variable and three steps. Firstly the demographic control variables of age, gender, nationality and disability were entered, secondly the personality factors scores and thirdly the Need for Cognition and Occupational Self-Efficacy scores were entered. Measures were entered in this order to allow for controlling for the demographic factors (which were all coded into two categories), then due to past research establishing links between personality and SJT performance as noted in the literature review above. The three main constructs from the McDaniel et al., (2006) model were entered (conscientiousness, emotional stability and agreeableness). The Murphy (2012) variable of 'personality' further includes the remaining two personality factors and hence these were also entered (extraversion and openness). Some studies report links for all five personality variables and the SJT: McDaniel and Nguyen (2001) reported average correlations between the big five and the SJT performance within their meta-analyses. The correlations were .25, .26, .31, .06 and .09 for agreeableness, conscientiousness, emotional stability, extraversion and openness respectively. Furthermore, Chan and Schmitt (2002) reported correlations of .23, .24, .29 and -.20 for conscientiousness, extraversion, agreeableness and emotional stability respectively.

The two newly tested variables were entered in the third step to see if they added incremental validity above the established personality measures. All the necessary checks on the variables regarding format and distribution were run before analysis.

Table 5 reports the results of this analysis. Step 1 of the analyses entered the control variables which accounted for 2.5% of the variance in SJT scores. The inclusion of the personality factors in step 2 accounted for an additional 3% of the variance (Total $R^2 = .054$, $p = .012$).

Conscientiousness ($B = .537$, $p = .015$) and intellect ($B = -.635$, $p = .017$) scores provided the only significant effects. Step 3 included the Need for Cognition scores and Occupational Self Efficacy scores. These scores accounted for another 0.5% ($\text{Total } R^2 = .059$). However, the change in R^2 was non-significant and neither of the new variables had unique effects in the regression model.

Table 5. Hierarchical regression of SJT performance with three steps.

	Step 1 B	Step 2 B	Step 3 B
Disability	9.83	8.95	9.67
Gender	10.88*	8.72*	8.89*
Nationality	-3.52	-4.44	-5.26
Age	.284	.114	-.016
Extraversion		.069	.026
Agreeableness		.161	.205
Conscientiousness		.537*	.516*
Emotional stability		-.289	-.299
Openness		-.635*	-.732*
Need for Cognition			.315
Occupational Self-Efficacy			.150
Change R^2	.025*	.030*	.005
R^2	.025*	.054*	.059
$F(df)$	2.39* (4, 380)	2.41*(9, 371)	2.29*(11, 374)

($N=386$, $*p<.05$)

This partially supports hypothesis 1 (personality and SJT performance) in that personality factors do significantly explain 3% of the variance in SJT performance along with the control measures, although only the openness and conscientiousness traits had main effects upon the SJT scores. The personality scores themselves only accounted for an additional explanation of 3% in SJT score variance. The openness b value was negative suggesting that high openness scores are related to low SJT performance.

The regression analysis does not support hypothesis 2 (Need for Cognition and SJT performance) or hypothesis 3 (Occupational Self-Efficacy and SJT performance) as these

variables do not significantly explain variance in SJT performance in step 3. Consequently, the analysis does not support the additional elements of hypothesis 2 and hypothesis 3 that suggested the Need for Cognition and Occupational Self-Efficacy variables would have incremental validity over the personality variables in explaining SJT variance.

Discussion.

Hypotheses Summary

The results indicated that the hypotheses, with the exception of hypothesis 1, were not supported by the findings. The SJT failed to show significant associations with the three personality factors from the McDaniel et al., (2006) model, the Need for Cognition scores or the Occupational Self-Efficacy scores. These two latter new constructs also failed to give incremental validity over the personality factors when looking at SJT performance.

Variance in the SJT scores was significantly explained (albeit this was only a small amount of variance at 3%) by all 5 personality factors together along with the control measures in the regression model seen in Table 5, although only openness and conscientiousness had significant main effects upon this step. Furthermore, the openness effect was a negative one suggesting that higher openness scores on the Big Five measure suggest lower SJT scores and the five factors together only explained an additional 3% variance in SJT scores over the control measures. Hence, hypothesis 1 is partially supported.

SJT Performance and Personality

The five personality factors resulted in a significant regression model with SJT performance as the outcome variable. This lack of significance for two of the three main personality factors from the McDaniel et al., (2006) model may be due to the specificity of the SJT used (i.e. an SJT written for medical students and assessing the five domains noted). Correlates of the SJT are noted to differ according to the SJT itself and factors such as response instructions (McDaniel et al., 2006). The instructions for this SJT were all knowledge tendency however, and these instructions are expected to correlate less with personality traits than behavioural tendency instruction SJTs. Schmitt & Chan (2006) emphasise this point and report how SJTs can be developed to either correlate with personality measures or with cognitive ability. Within this idea they also state how an SJT can be developed to assess different constructs (Chan and Schmitt, 2005). Therefore, the lack of relationship between personality traits and the SJT performance does not necessarily infer that this SJT is a poor tool but more possibly that this is a result of it's' construction and the instructions used.

Much as any interview can differ within selection one SJT must differ from another according to circumstance. It is likely that had a cognitive ability variable been included in this study there would have been a significant positive relationship with the SJT performance.

However, conscientiousness as a factor alone did have a unique positive effect in the regression model suggesting that it has contributed to the explanation of variance, although the model itself is not robust proof of the McDaniel et al., (2006) model as a whole from this particular SJT. McDaniel et al., (2006) suggests that the three factors of personality mapping into the broader personality factor α (stability) would correlate significantly with SJT performance as constructs of interest. The separation of the five factor model of personality into two broader factors (α , as described here referring to consistency of mood, motivations and conformity, and β referring to plasticity, nonconformity and searching for new experiences) is not supported by the results of this study. There are some strong correlations between the conscientiousness and agreeableness ($r=.526$) and emotional stability and agreeableness factor ($r=.526$) but a smaller correlation between the conscientiousness factor and both emotional stability ($r=.381$). Extraversion and openness to experience show a moderate relationship ($r=.415$) although there are still stronger relationships shown across facets of the two broader factors, e.g. (openness to experience and agreeableness, $r=.547$) suggesting that the five factors individually are important to examine as well as potentially considering them as split into two main factors. There is a need for a measure of cognitive ability in this study and a more robust testing model of variables, i.e. examination of general and specific job knowledge, experience and education/qualifications in detail. Cognitive ability measures have shown weak relationships with personality traits suggesting that they are independent to personality traits in explaining job performance (Barrick & Mount, 2012) and hence it would be expected would be independent within the SJT performance in turn,

The five factors of personality as a whole were found to significantly add incremental validity over the demographic controls in the regression analyses; albeit only to only explain an additional 3% of the variance in scores. Conscientiousness and openness both had unique effects in this model (positively and negatively respectively). This is consistent with previous research which tends to find conscientiousness as most strongly associated with SJT performance scores (e.g. McDaniel and Nguyen, 2001; Kuncel, Hezlett and Ones, 2001). It is potentially surprising that openness has a negative relationship with SJT performance as one would expect that one with high levels of intellectual curiosity and creativity would perform well upon the SJT as it potentially involves thinking creatively about work place scenarios

and future behaviours, or options for behaviours. It follows that this skill may be important in actual job performance, yet not assessed by this particular SJT.

Mount et al., (2005) described different dimensions of how personality traits are reflected in interests. One of the dimensions identified personality-interest clusters with 'striving for personal growth versus striving for accomplishment'. The personal growth end was defined by openness to experience, extraversion and artistic interests with a preference for working with thinking, creative materials and abstraction. The accomplishment end was defined by conscientiousness and conventional interests.

It follows that if the meta analytic data supporting factor α (containing conscientiousness along with agreeableness and emotional stability) is the construct of interest for the SJT, that factors relating to the other broad factor of personality (β -consisting of extraversion and openness) would potentially either be unrelated or negatively related to SJT results. Hence, conscientiousness is positively related to explaining SJT performance and this is rooted in conventional interests and striving for accomplishment. Whereas there is a negative prediction from the openness to experience trait, as this is rooted in creativity and striving for personal growth, which may not be applicable to success on this SJT designed for best medical practice in foundation training years. Introduction of New Constructs: Need for Cognition and Occupational Self-Efficacy

The introduction of new constructs into the McDaniel model was hypothesised to add incremental validity above and beyond the five factors of personality and the control measures. This was not the case and the three step regression model was non-significant. This suggests that these variables cannot predict performance upon this medical selection SJT. It should also be noted that only a small amount of the variance in SJT scores (5.4%) could be explained from the demographic control measures and the five personality factors together. As noted, this is potentially an artefact of the SJT design and construction itself.

Interestingly, as the McDaniel et al., (2006) model suggested that only the three factors of personality mapping onto the broader factor α (stability) i.e. conscientiousness, agreeableness and emotional stability were constructs measured by the SJT, it is noteworthy that both the Need for Cognition and Occupational Self-Efficacy variables correlated significantly and with stronger positive relationships for the two personality facets making up the other broad personality factor of β (plasticity) in comparison.

Occupational Self-Efficacy and Need for Cognition are two scales that measure subjective opinions of oneself that one is aware of, and hence can express a rating of for the scale. The SJT may be more affected by knowledge and ability that one is unaware of, perhaps on a subconscious level. For example, procedural knowledge is the knowledge of how to perform a task successfully and such abilities may be more important when completing an SJT than whether an individual believes themselves capable of being successful in the workplace. Hammond (1955) discussed a 'quasi rationality' as a type of common sense basis for good judgment and in fact these abstract knowledge stores may be the focus of the SJT, as opposed to the more available and self-accessible cognitions. Furthermore, these ideas are similar to the tacit knowledge and practical intelligence concepts thought to be involved in good judgment derived from work by Sternberg and colleagues. The intelligence, and important constructs, to investigate for further explaining variance in SJT performance may therefore be involved in the aspects of intelligence and cognition that do not require conscious accessing, manipulation or application. Hence, answering SJT questions is not aided by the motivation to 'think' harder nor the idea of self-efficacy regarding job behaviours. This further suggests that thought in the SJT is from an abstract intelligence that cannot be assessed by self-rating, but potentially by objective assessment. Measures investigating new constructs of interests that may result in greater explanation of variance may therefore need to be quantitative and objectively scored using a maximal performance test, as opposed to a typical performance one that somewhat depends upon an individual's own self-awareness and knowledge of conscious, readily accessible and basic knowledge stores/opinions as these measures used here.

Limitations

A limitation of this study is the lack of inclusion of one of the main variables in the McDaniel et al., (2006) model of the construct validity of the SJT; a cognitive ability measure. This was unfortunately out of the research team's control and timing options. Before the commencement of this research project, the study planning and organisation had already taken place. Resources and ethical constraints meant that testing had to occur at a specific time and there was no option for additional measures beyond those reported in this thesis. Furthermore, had time, resources and the situation permitted it would have been preferable to include further variables to examine the other main theoretical model of the SJT explained by Motowidlo & Beier (2010) alongside the McDaniel model. As noted, Motowidlo & Beier have made a step forward in the literature to include new concepts (e.g. the ITP) and this is

the other main model of the SJT. This needs investigation itself and other research should focus on replicating studies involving the ITP since this was not an option within this research project. Before tackling newer concepts, it seemed logical to assess a more basic model within a medical sample when the opportunity, resources and sample were available which could then act as a basis for the development of ideas and further theory that may be interlinked with existing models in the remainder of this thesis. In fact, the cognitive ability variable as a construct of interest with regards to the SJT is the most researched relationship and links between cognitive ability and job performance and success are extensively found in the literature (as noted by Murphy, 2012). It would have been preferable to have a measure of cognitive ability included in the research to allow for investigation of the relationships between the other additional variables and cognitive ability.

The study is also domain specific as the SJT was designed for a medical population and completed by medical students. The results are therefore not fully generalisable to other domains or occupations. Furthermore the participants were actual medical students who were only 'assuming' the role of Foundation year 1 Doctors, and hence, performance may not have been of maximum effort due to the fact that it was not actually a selection procedure but only a simulation of one. However, for the purpose of this thesis, where the SJT in medical selection is the central focus, this is also a strength of this study.

Summary and Conclusions

In summary, the three main personality factors of conscientiousness, agreeableness and emotional stability failed to significantly explain variance in the SJT scores, and the newly introduced constructs failed to add incremental validity above and beyond the five factors of personality. Only a small amount of variance in SJT scores though could significantly be explained by the control measures and the five personality factors together (5.4%) suggesting that there is a lot more to be learnt about this medical selection SJT and what constructs are of interest in relation to the measure. The lack of relationships seen in this study may also be a consequence of this particular SJTs structure and construction; for example, a stronger underlying focus on cognitive ability constructs as opposed to personality ones. The study, and investigation of these new variables, should therefore not be discarded, but investigated using different SJTs in different fields. Evidently the McDaniel et al., (2006) model is not supported by this study in particular.

Nevertheless, this study itself was the first of its kind to investigate these new constructs alongside a personality trait measure and an SJT within a medical sample. The importance of research into medical selection and assessment as a whole has been noted already in the literature review and researchers have been encouraged to explore new constructs and variables in relation to SJT performance from previous authors in the SJT research field. It is therefore necessary to investigate this area in particular, within medicine, and to use the introduction of new constructs that may affect SJT performance, where this may not be the case for other occupations due to the individuality and uniqueness of medicine as a profession. Whilst these new variables of Occupational Self-Efficacy and Need for Cognition have shown no significance for this SJT, they may be highly important for other SJTs within medicine, or potentially other careers. Furthermore, whilst these variables did not explain performance upon the SJT, they may explain job performance and this is a direction for future research to consider.

The obvious need for further investigation and exploration of other potential factors and constructs affecting the SJT is clear. This study gives an awareness of how much is lacking in terms of explaining results regarding this SJT performance specifically from a psychology point of view and the inclusion of new constructs has not improved the explanation of variance. The links between the main constructs of interest and SJT scores has been shown to vary according to the specific SJT. For example, the instruction format and what the SJT is designed to measure can alter these relationships. The results here are in line with these ideas. Relationships in existing literature appear to be accepted and widely acknowledged (i.e. cognitive ability and personality are held to be the main constructs influencing SJT performance, as highlighted in McDaniel et al., 2006), rather than explained in terms of theoretical science. Furthermore, the percentage of variance explained by such models is relatively low and investigations should attempt to explain more from new constructs. This research suggest that the existing constructs (i.e. personality constructs from the McDaniel model) do not appear to consistently significantly explain variance in this medical selection SJT, and nor do the newly introduced constructs give incremental validity over the personality factors. New models of the non-cognitive processes at work need to be developed in order to allow for falsifiable ideas and a clearer understanding of what the SJT is measuring. Further inspection of the literature is required and other constructs and measures should be targeted. It has been noted that potentially the constructs to target now for investigation should be objective measures, as opposed to self-rating scales. This may then

allow for development of a clearer conceptual model of constructs of interest to the SJT, and in particular, within medicine.

From this study, it can be concluded that this medical selection SJT as a method appears to not to be explained to a satisfactory level from the current model of construct validity put forward by McDaniel et al., (2006). Other current theories (e.g. Motowidlo and Beier, 2010) and new constructs should be investigated and ideally a coherent conceptual explanation of the SJT, explaining larger amounts of variance in both SJT scores and consequently job performance, using existing and new theoretical ideas to do this, should be constructed.

The difficulty now lies in a threefold problem. Firstly determining the variables to be targeted as potential constructs to establish clearer ideas of the construct validity of the SJT. Secondly, how can such constructs be examined and measured in a practical sense, and thirdly, how can this help us to understand the SJT as a method in terms of developing conceptual clarity in a theoretical model. The first two of these issues is discussed in study 2.

CHAPTER 5: PIAGETIAN THOUGHT

Study 2: Formal Operational Thought Proficiency and Cognitive Profiles in Control and Dyslexic Samples of Young Adults

Introduction to Study

The first study examined personality and two existing non-cognitive variables assessing individual differences (Need for Cognition and Occupational Self-Efficacy) alongside a medical selection Situational Judgment Test (SJT) performance within a medical student sample. The study indicated that other constructs should be targeted in an attempt to further explain variance in SJT scores. It was suggested that measures involving objective scoring, as opposed to self-rating measures could allow for sub-conscious knowledge or skill to be assessed and this may be the key to explaining SJT performance further.

This second study focuses upon intelligence and targets a certain abstract aspect of intelligence called formal operational thought (FOT) from contemporary intelligence theory (CHC theory), as a new construct of interest in relation to the SJT (i.e. examination of the construct validity of the SJT in relation to FOT). A new measure of FOT is developed and piloted across two different samples using a range of cognitive tasks with the aim of validating this measure. Broadly speaking, this study develops and establishes a measure of higher level thought that can then be used in study 3 to tackle the remaining research questions (3-5).

From the small amount of variance explained in the first study by personality, a variable widely held to explain variance on the SJT from previous literature (e.g. McDaniel et al., 2006) it appears that there may be other constructs to introduce in order to gain incremental validity over existing variables, and to gain a clearer theoretical understanding of the SJT. This may be true for SJTs in general, or the lack of relationships from study 1 may be applicable only to the medical selection SJT used. Alternatively, the SJT design itself may be the cause of these results as opposed to the field of occupation (i.e. knowledge tendency instructions showing only weak links with personality variables and yet stronger ones with cognitive ability variables).

Investigation of the literature relating to intelligence and the good judgment aspect of the SJT suggested that there was an underlying practical intelligence ability of ‘tacit knowledge’

involved in the SJT. Indeed it was suggested by Sternberg and colleagues (1985), as noted in the literature review, that this was in fact what the SJT was measuring and was behind the judgment aspect of the measure. This idea has though come into disrepute and it is argued, amongst other points, that much analysis shows no evidence for such an underlying ability (see Gottfredson 2003).

New models of the processes at work, in terms of cognitions and variables, need to be developed in order to allow for falsifiable ideas and clearer understanding of what the SJT is measuring (i.e. construct validity); existing literature encourages the development of theory and constructs (McDaniel et al., 2006) but the constructs already identified do not appear to affect SJT performance in the SJT used in study 1, nor explain a substantial amount of variance in scores. Further inspection of literature is required and new constructs/measures should be targeted.

There may not be an overarching influential judgment factor, such as tacit knowledge, but insights into intelligence and cognitive literature may lead to similar constructs regarding ability that affect performance; an aspect of abstract intelligence. Examination of the contemporary intelligence theories (see chapter 2) led to the pinpointing of the broad abstract area of intelligence of fluid reasoning. Within this, the narrow ability of Piagetian reasoning (originally from Piaget's 1952 Cognitive Development Theory) has been pinpointed with the aim of bringing forward an area of intelligence that may be able to explain variance within the SJT. This quantifiable construct was identified as one that has not been investigated in relation to the SJT before, and one that can be objectively assessed. This construct is referred to as 'Formal Operational Thought' (FOT) and is a higher level thought ability defined within Piagetian thought (1952). The introduction to study 2 discusses the literature concerning this concept. Schmitt & Chan (2006) reported that a great deal of variance in SJT measures, in their meta-analytic data, was unrelated to performance, cognitive ability, personality and interests. They suggest that there may be an underlying situational judgment construct that is independent of most other individual difference variables used typically in personnel selection. It is proposed that FOT may make up, in part, this underlying construct at play.

Ryan & Ployhart (2013) note that beyond the key constructs associated with the SJT (personality and cognitive ability) the construct validity of the SJT is somewhat of a mystery. Researchers are beginning to examine new constructs such as 'personal initiative' (Bledow & Frese, 2009) and 'team role knowledge' (Mumford et al. 2008). This study is therefore in line

with current research that aims to target and investigate new constructs that are potentially related to SJT performance.

A novel test was developed to assess these higher level thought abilities from Piagetian thought within intelligence models. An existing measure was added to, resulting in a new test of FOT. The measure was piloted using two differing samples of young adults (i.e. a sample of typical cognitive ability vs. a sample with dyslexia) to see if the measure was able to establish variance, stand as a valid tool and it was expected that the cognitive impairments or risk factors associated with dyslexia would not be affected by FOT proficiency. The following section offers full explanation and justification for this sample choice as it is noted the samples used were not as would have been initially chosen by the researcher.

Nonetheless, the scientific merit of this study lies in the fact that there were three main research benefits from this study. Firstly, a measure of FOT was added to that allowed for a novel and not entirely science topic based assessment of the higher level reasoning ability to be developed and piloted for the first time.

Secondly, from a theoretical point of view, the conceptual location of the FOT/ Piagetian reasoning ability was of interest and the key point that allowed for validation of the measure. The dyslexia sample adds value because it enables a good test of the validity of the measure; if the measure is a valid one, it would be expected that the dyslexia sample would score similarly to the typical sample for higher level thought assessment; FOT proficiency is unaffected by deficits in lower level cognitive tasks caused by dyslexia. Only differences in performance between the groups for lower level cognitive tasks would be expected (i.e. the typical group score more highly on these tasks compared to the dyslexia group). Validating the measure used existing and well established dyslexia literature to ensure that the FOT test was assessing the higher level thought ability, as opposed to lower level cognitive skills.

Finally, and although potentially somewhat beyond the scope of this thesis, from a developmental psychology point of view, the study allowed for comparisons to be drawn between a typical sample and a sample with dyslexia concerning their FOT and higher level reasoning abilities for the first time.

Theoretical Basis for Study

The 'situational' element of the SJT is well researched and explained (e.g. the discussed social psychological literature from Barker, McGrath and Hammond). The 'test' element is

also widely assessed with extensive research using the SJT across occupations, into the development/structure of the SJT, regarding validity as a measure, response instructions and marking of the SJT (as noted in Weekley & Ployhart, 2006). However, the 'judgment' element of the test is less well explained. Beyond a general consensus across literature that the SJT measures a 'general judgment' (Weekley & Ployhart, 2006) that is above and beyond that of typical cognitive intelligence there has been little evolution of this aspect of the construct validity and related theory behind the SJT. This is noted and encouraged as a research topic (McDaniel et al., 2006).

The most comprehensive attempt at explaining intelligence with a conceptual model comes from Carroll (1993) and the consequential CHC (Cattell-Horn-Carroll) Theory (see pages 28 & 29). This can be viewed as the end point of cognitive development (or at least the latter stages as opposed to the developmental periods). Within Carroll's model he described the 'g' ability as a third order construct with five main secondary constructs and two abstract secondary constructs. These abstract constructs are known as fluid and crystallised ability (labelled as g_f and g_c respectively). These refer to the ability to deal with novel problems and come up with creative solutions (fluid intelligence) and the ability to apply learnt skills and known solutions to problems (crystallised intelligence). CHC theory went on to incorporate Carroll's model in an integrated theory with Horn and Cattell's intelligence research; there were small differences between the theories and the main one of these being the lack of 'g' in Horn and Cattell's ideas vs. the inclusion of 'g' for Carroll. A resultant set of 10 broad abilities (e.g. reading/writing ability) and narrow abilities within these broad topics were defined.

To investigate the links with non-academic intelligence it was sensible to look to the abstract areas of intelligence within this modern model. Furthermore, upon inspection of this model and the SJT literature it was apparent that these abstract areas of intelligence had not been investigated in relation to the SJT measure before. Therefore, the constructs were examined and, in particular, the potentially difficult to examine skills involved in fluid intelligence (sequential reasoning, inductive reasoning, quantitative reasoning and Piagetian reasoning). The other second order constructs are also supposedly involved in the SJT performance, for example, verbal comprehension, knowledge and achievement (typically known as cognitive ability), learning and memory and perceptual speed are all aspects of intelligence that are likely to be targeted when undertaking an SJT. However, the concept of dealing with novel problems and coming up with creative solutions (i.e. fluid intelligence) suggests a concept

very relevant to success upon an SJT, and has not previously been investigated in relation to the SJT.

Exploration of these concepts, and the Piagetian reasoning aspect of fluid intelligence, appears to have not been considered in relation to SJT performance before now and this was investigated with the aim of further explaining SJT results through 'Formal Operational Thought'. The literature relating to formal operational thought (FOT) is discussed within the introduction for study 2, as this is when the literature becomes relevant to this thesis.

A measure of Piagetian reasoning was extended based upon previous work from Adey, Shayer & Yates (2001). In order to establish the validity of this measure and conceptualise the cognitive placing of FOT, the Piagetian reasoning measure was piloted using a sample of typically cognitively developed students and a sample with the developmental disorder of dyslexia, i.e. rudimentary problems with reading and basic cognitive skills. It is suggested that the Piagetian reasoning (FOT) is a higher order ability and hence the proficiency of this would not be affected by dyslexia or basic cognitive impairments as this level of cognition is below the level of thought that FOT is used for. The comparison of performance between the dyslexia students and the typical students will allow for validation of the measure. The theory surrounding dyslexia is very well established and developed, whereas this is not the case for the SJT. As a cognitive developmental disorder with established theories explaining the cognitive skills and impairments of thought, dyslexia may be able to offer a measure to compare FOT performance against and to validate the measure.

The measures used in this study were: (a) a novel test of FOT, and (b) a selection of tests from the DST-S (Dyslexia Screening Test for Secondary School, Fawcett & Nicolson, 2004). These measures are further discussed below.

Piagetian reasoning

Piagetian reasoning and theory (Piaget, 1952) was investigated within developmental and cognitive psychology literature. Based upon this literature investigation, and the CHC theory of intelligence, the concept of 'formal operational thought' was specifically targeted as a construct involved in the narrow ability of 'Piagetian reasoning' which is conceptually located within the broad ability of 'fluid intelligence'. As noted above, this has not been investigated in relation to the SJT before now. Schmitt & Chan (2006) reported that a great deal of variance in SJT measures, in their meta-analytic data, was unrelated to performance, cognitive ability, personality and interests. They suggest that there may be an underlying

situational judgment construct that is independent of most other individual difference variables used typically in personnel selection.

A measure of FOT was developed based upon FOT questions used in existing Cognitive Acceleration Projects using Piagetian testing from Adey, Shayer & Yates (2001). These projects were designed as programmes that attempted to develop students' thinking skills within schools through task work relating to FOT skills.

A section of non-scientific questions was developed and included within the scale in order to broaden the measure from a science-domain specific category. All questions that existed in the measure from Adey, Shayer & Yates (2001) were science based and dominated by reasoning within this domain. For example, examining principles of volume or mass. Therefore, it was important to develop this measure to include non-science based questions so as not to disadvantage those who may not hold strong scientific knowledge or skills.

This developed scale was piloted using a typical cognition and a dyslexia group in order to establish the conceptual cognitive placing of FOT (i.e. FOT should not affect basic cognitions such as reading, or more complex ones, or both) and to further establish the validity of the tool from this and eliminate any errors with the measure/ ensure it was representing variance across the sample (i.e. no ceiling or floor effects in results).

The questionnaire was designed to assess the different levels of thought building up to and involved within FOT in a hierarchical nature. It involved problems based upon the principles of volume and heaviness, balance and equilibrium and the additional brainteasers not related to scientific principles. Further detail upon the measure development and detail is included in chapter 5 (study 2).

Cognitive profiles

A battery of dyslexia assessment tests were administered to both samples with hypotheses predicting better performance for the typical group across sub-tests and poorer performance for the dyslexia group. Tests selected from the DST-S were the Rapid Naming, the One Minute Reading, the Postural Stability, the Phonemic Segmentation, the Backwards Digit Span, the Nonsense Passage, the Nonverbal Reasoning and the One Minute Writing tests (all included in the Dyslexia Screening Test Battery: Secondary; DST-S by Fawcett & Nicolson, 2004). These tests were selected in order to give a well-rounded picture of a number of participant's abilities through well-established assessments, for example, memory, reading,

writing, working memory, etc. The specific rationale for each test and further detail on each assessment is given in the method for study two.

Similar performance across both groups was expected upon the FOT test as this is an assessment of higher level reasoning and thought, as opposed to lower, rudimentary and more typically used every day cognitive skills such as reading or writing.

This study aims to establish validity in a measure of a construct of interest to the SJT (FOT) and to allow movement in study 3 towards a coherent and integrative model (i.e. social, cognitive and intelligence research related) of the SJT using this within the explanation of the construct validity of the SJT. The piloting of this FOT measure was the basis for the inclusion of a new construct in SJT modelling and the collaboration of cognitive literature regarding intelligence, FOT and the SJT in medical selection and assessment.

Study Design

A range of tests were selected from the Dyslexia Screening Test Battery: Secondary (DST-S) (Fawcett & Nicolson, 2004) assessing the necessary cognitive domains to give a conceptual picture of a range of basic cognitive abilities. FOT problems were developed and then grouped into a worksheet based upon Adey, Shayer & Yates (3rd ed) Thinking Science DVD (2001). The measure itself, and the effect of FOT proficiency as a concept, could therefore be examined using the cognitive tests of different abilities in the DST-S. The DST-S was chosen as a suitable battery of assessments following the inclusion of a dyslexic sample in the study.

Sample Justification

The strength of using the dyslexia sample was the possibility to discriminate between cognitive profiles using established dyslexia theory was an effective way of validating the FOT scale and conceptualising the intelligence aspect.

The focus of this thesis was upon the SJT and in particular, the SJT within a medical selection and assessment context. With this in mind, the FOT measure would have ideally been piloted using either a medical student or medical professional sample. It could have potentially been administered alongside an SJT and/or other measures related to the modelling of the SJT (i.e. cognitive ability tests or personality measures). However, due to issues beyond researcher control it was unfortunately not a possibility at the time. Due to time constraints in particular, and the difficulty of contact and ethical constraints in obtaining a medical student/professional sample, it was necessary to move forward with the research

programme and to pilot the test where, when and with whomever was available to the research team at the time. Hence, this study was designed to validate the FOT measure to allow for confidence in it for the third study using a medical student sample.

Consequently, an opportunity sample was used for study 2 of students from the University of Sheffield. The design of the study moved towards the investigation of the tool using the typical vs. dyslexia sample to gain a richer insight into FOT as a concept and, it was suggested that as a construct it would not affect performance on basic cognitive tasks (i.e. lower level cognitive tasks such as reading, as opposed to higher level cognitive tasks that require more mental effort and processing). The two samples were expected to differ upon cognitive abilities displayed among the dyslexia battery tests that target rudimentary and basic level skills (e.g. reading, writing, memory) yet the measurement of FOT targets a higher level ability that therefore would not be expected to show such differences in performance between the samples.

If both sample groups perform similarly upon the FOT test yet differ in their primary cognitive test results (i.e. the dyslexia battery of tests) then confidence can be held in the validated novel measure; it is measuring what it was expected to measure. Furthermore, this study allows for an opportunity to further conceptually define formal operational thought from the results. The aim of this conceptual development is to inform developing the model of SJT performance and the role of FOT within this in the following study. The reasoning and abstract relation abilities involved in FOT may be above and beyond that of the lower level cognition skills assessed in the DST-S and this aspect of fluid intelligence is hypothesised to be an ability that may be beneficial in SJT performance.

FOT may be a construct assessed by the SJT measure; higher order thought abilities may in fact aid the ability to consider the situation, options, reason regarding these and assess the pros and cons for each choice, before deciding upon a sensible and logical answer. This is further discussed in the literature discussion of formal operational thought below.

Due to the sample choice it is necessary to briefly define the developmental disorder of dyslexia and how this learning disorder may impair ability and performance upon certain tests.

Dyslexia

Developmental dyslexia is traditionally defined in terms of a discrepancy between reading performance and general intellectual function. Dyslexia is a developmental disorder where “a child, despite conventional classroom experience, fails to develop the reading, writing and spelling skills, commensurate with their intellectual ability” (World Federation of Neurology, 1968, p.26). The distinction between children with dyslexia and garden variety poor readers can be made by the criterion of those with dyslexia failing to attain the skills expected for their IQ level, whereas garden variety poor readers show low IQ concordant with their poor reading. Stein and Walsh (1997) found 15% of boys and 5% girls with normal intelligence failed to learn to read and write at the same level. Other research has criticised the need for the necessary discrepancy between IQ and literacy skills and argues that “poor readers of all IQ levels show equivalent difficulties with reading, spelling, phonological processing, short term memory and syntax” (Siegel, 1989, p518, cited in Fawcett, Nicolson & Maclagan, 2001). However, Nicolson & Fawcett (2006) argues the importance of still distinguishing between poor reading ability, general learning difficulties and dyslexia as the same array of symptoms may be present as a result of different underlying causes, and in turn, will require different treatments in order to help.

It is a particularly challenging paradox for psychologists to explain; how an able and intellectually capable individual can demonstrate such difficulties with one of life’s most routine skills (that of reading). Reading involves many processes being combined in a seamless manner and at the very least involves automatic word recognition, eye movement control and speech monitoring (Nicolson & Fawcett, 2008). Even though the acquisition of this ability takes thousands of hours of practise and considerable effort most people tend to master the ability. Natural variation of the human species has however led to populations that do not acquire this ability to the same level.

The disorder has a high incidence of approximately 5% across Western societies (Badian, 1984) and thus is a high profile area for research. As a consequence of the large practical implications of research there are numerous theories attempting to explain the underlying causes of dyslexia. Over the history of research, ideas have moved from specific deficits in ability to more general explanations due to the recognition of such widespread symptoms concerning sensory, motor and reading impairments. The ‘core deficits’ of dyslexia are held to be the literacy related symptoms (Nicolson & Fawcett, 2007).

Theories consider causes of dyslexia symptoms at behavioural, cognitive and biological levels of explanation, but broadly, since the 1970's, they can be considered as focusing either upon cortical and sub-cortical brain abnormalities, or upon aspects of phonological processing deficits.

Almost all dyslexia research has focused on the reading deficits. In many studies now, researchers probe abilities rather than disabilities (c.f. positive psychology ideas). Recent research has begun to look from this perspective and focuses upon the strengths of individuals as opposed to their weaknesses. For example, those with dyslexia may show extremely high levels of creativity or adaptability. In fact they may have some high intellectual skill that is of special value in the workplace (Eide & Eide, 2011). In particular and of importance to this study they may show high levels of formal operational thought. Furthermore, by using this population, it could be possible to establish formal operational thought as an ability separate to certain other rudimentary abilities; i.e. how impaired reading or other deficits in core cognitions, relate to formal operational thought proficiency.

Study 2 Literature

The SJT and Intelligence

As mentioned above, it is widely accepted that SJTs tap into the assessment of procedural knowledge. However, there has been little investigation to further our knowledge of this.

It is well known that SJTs correlate strongly with general cognitive ability, with the meta-analysis of 16,984 people (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001) revealing a correlation of 0.46. Hence, there are established links with intelligence and cognitive ability, but the reasons why this is so are less clear. It appears that examination of intelligence literature may lead to avenues that have not yet been investigated, for example, abstract intelligence and fluid reasoning (and FOT which sits within this modelled branch of intelligence).

The essential feature of the SJT is that a situation is presented, and the student/applicant has to make a judgment, from a set of options, where there is no clear correct answer. The test therefore requires subtle cognitive skills – reasoning in conditions of uncertainty, applying multiple general principles to a complex situation, predicting the effects of actions and ranking the outcomes, and risk analysis. For each alternative the candidate may mentally list the pros and cons, the likely outcomes of the proposed action, possible risks and mitigating

factors, then attempt to rank the possible alternatives in order of overall benefit, taking into account the fact that there are various overriding principles such as ethics, empathy, efficacy, risk, teamwork, impression management and so on, which are highly unlikely to all point towards the same possible answer. The sophisticated candidate may also take into account the fact that the 'best' choices have been selected by a group of experts who have a particular set of perspectives. All in all, this is a very complex reasoning and judgment task that may require higher level thought abilities. This judgment element is the area that appears to require further clarification, modelling and investigation, for example, what constructs are affecting the individual's 'judgment' on the SJT?

The most comprehensive account of intelligence within a conceptual breakdown model comes from Carroll (1993) and the later developed CHC theory, and is described in the literature review (see chapter 2). This spread of intelligence and its' branches can be viewed as the end point of cognitive development (or at least the latter stages as opposed to developmental periods). Within Carroll's model, he described the 'g' ability as a third order construct, with five main secondary constructs and two abstract secondary constructs. These abstract constructs are known as fluid and crystallised ability (labelled as g_f and g_c respectively). These refer to the ability to deal with novel problems and come up with creative solutions (fluid intelligence) and the ability to apply learnt skills and known solutions to problems (crystallised intelligence).

These abstract abilities are targeted in this thesis as areas of intelligence that may affect performance upon the SJT. These areas of intelligence were focused upon and in particular, the potentially difficult to examine skills involved in fluid intelligence (sequential reasoning, inductive reasoning, quantitative reasoning and Piagetian reasoning). The other second order constructs are likely to be involved in the SJT performance, for example, verbal comprehension, knowledge and achievement (typically known as cognitive ability), learning and memory and perceptual speed are all aspects of intelligence that are likely to be targeted when undertaking an SJT. However, the concept of dealing with novel problems and coming up with creative solutions (i.e. fluid intelligence) suggests concepts very relevant to success upon an SJT.

It follows that from this model of intelligence, and assumptions of the processes required to perform successfully upon the SJT that Piaget's formal operations stage (Piaget & Inhelder, 1958), in which a person learns to reason in abstract rather than concrete terms, applying

general principles to specific situations could be involved in this judgment aspect. The Wechsler Adult Intelligence Scale (Wechsler, 1999) distinguishes four major categories: Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index and Processing Speed Index. None of these assess higher level thought abilities such as formal operations and hence, there appears to be a gap in the literature to assess this aspect of intelligence. There is also a new dimension in that factorial models of intelligence, such as Carroll's, see intelligence as a stable and final result, but when examining Piagetian thought, there is scope to view intelligence from an ontogeny viewpoint and as a developmental concept; there is potentially always scope to develop and improve the ability. Thought can be developed and improved across time, as opposed to intelligence being viewed as a stable concept.

In fact, Piagetian ideas opened doors to ideas of genetic epistemology based on the premise that knowledge is a consequence of continuous development and building of the mind from interacting with the surroundings and environment. Piaget held that development was a process of moving from one developmental stage to another and that these cognitive developments were based upon world interactions (Piaget, 1970).

Piagetian Thought

Jean Piaget's (1896-1980) Cognitive Development Theory (1952) provided one of the first and, still today, the most influential and plausible description of a child's cognitive development. He described four periods of development, each of which are identified in reality by distinctive behaviours and realisations by the child over the sensorimotor, preoperational, concrete operational and formal operational stages.

Firstly, the sensorimotor period occurs from ages 0-2 years old. Cognition then is described as developing from reflex actions into complex behaviours and reactions based upon symbolic reasoning and thought. The main features that Piaget describes as marking this stage are the acquisition of 'object permanence' (i.e. the idea that an object still exists even when it is hidden or out of sight) and 'deferred imitation' (i.e. the ability to represent experiences or actions that have been observed in mental representations that the child will then recreate at a later time).

The preoperational period follows and ranges from the ages of 2-7 years old. Both language abilities and symbolic representation/thinking move on very rapidly in this period, for example, imaginative play and using objects as tools to represent other objects. The child is

still very much different to an adult in terms of cognition though; in this stage they fail to grasp the principles of conservation (i.e. the realisation that a volume of liquid remains consistent when poured into a differently shaped container) and the principles of egocentrism (i.e. that others can see the world in a different way to how you personally do).

The concrete operational period follows and extends from aged 7 to 11 years old and the end of this period marks the transition into adolescence. The child should master the principles of conservation as described above and also mature in thinking to allow for logical analysis of problems and the origin of development for mature emotions, such as empathy. They are able to reason when considering concrete objects, for example, blocks or sticks, however they have difficulties when asked to consider hypothetical scenarios or objects.

From the beginning of the formal operational thought period (aged 11 upwards) most children are able to think and reason in abstract terms; they can think and reason about hypothetical objects or emotions or perspectives and look at how their or others' behaviours may differ according to different circumstances and situations. Piaget describes this type of thought as 'culture free' i.e. it is not influenced by cultural variables, especially formal schooling (Piaget, 1972). This period is also thought to be important when developing an understanding that other may view you in different ways, some may agree with your views yet some may not, and you are able to imagine what other people think of you. This is helped along by the onset of physical maturity through puberty and self-awareness principles. Teenagers demonstrate these abilities by holding a view that they are constantly the centre of attention and that people are constantly talking about them (Elkind & Bowen, 1979). These views will peak during puberty before declining over further maturation. Adey & Shayer (1997) note a list of higher order thinking skills that excellently characterise the generality of formal operations. These included points stating higher order thought is non-algorithmic (path of action is not fully specified), is complex (path is not clearly 'visible'), often yields multiple solutions, involves the application of multiple criteria, involves uncertainty, involves self-regulation of the thinking process, involves imposing meaning (i.e. finding structure in apparent disorder) and is effortful (mental work is required; Resnick, 1987). Adey and Shayer further note the specific reasoning patterns involved in formal operations as controlling or excluding variables, classification of variables, ratios and proportionality principles, compensation and equilibrium principles, correlation, probability and the ability to construct and use formal models, and logical reasoning.

It is important to note that although there are four main stages of cognitive development in this Piagetian theory it has been shown that not all adults will reach the final stage of formal operations (Keating, 1979; Cok, 1990). For example, some tribal cultures have not shown the capability to reason with formal operational ability at all (Gellatly, 1987). This may potentially be as a result of lack of opportunities to engage in hypothetico-deductive reasoning. Hypothetico-deductive reasoning involves logical and scientific thought concerning causes and effects and the manipulation of certain variables to alter the consequences of a reaction or event. Deductive reasoning is noted as a key element of FOT (Cowan, 1978) which allows for a person to draw valid conclusions from information or arguments, whether or not the argument appears to actually be valid.

Piaget also held that the acquisition of propositional thought was a key aspect in gaining formal operational thought. This thinking involves understanding that the wording of an explanation has to be logical and possible, be true and make sense. Furthermore, some adults may only show this level of thought within their area of expertise, for example, only when thinking about problems or scenarios within their field of work, such as a physics professor may only be able to reason abstractly about physics problems, yet not when considering problems within another subject area. Crain (2005) has also suggested that formal education may be a necessary basis for the development of FOT, and that this ability is established in areas that people find most interesting. It is unclear whether the education in this specified area comes before or after the interest, hence, which variable is of interest with regards to formal operational thought development is unclear.

Therefore, it appears that FOT may only develop when it is used and practised on a regular basis; the basis of it lies more within a declarative memory field than a procedural one and its use (akin to rehearsal within memory) may be key to develop, use and retain over time the skill of formal operational thought.

The flexibility and logic that develops with FOT allows for mental structures to interrelate identity, negation, reciprocity and correlation (INRC; Cowan, 1978). This INRC group represents abilities to use abstract reasoning according to Muuss (1996). The four main abilities involved are 'Identity' which allows for the knowledge how something can be changed without major alterations, for example, the addition of 0 to another number does not change the original digit. Secondly, the 'Negation' ability allows for the knowledge of the undoing of an operation. Thirdly, the 'Reciprocity' ability gives the knowledge again that

undoing can occur, but with a neutralisation rather than annulment allowing for multiple hypothesis testing. Finally, the 'Correlative' knowledge allows for relationship understanding that if one variable alters it may and will affect another variable. Muuss (1996) holds that this system develops over FOT to become more skilled and results in a systematic process to achieve any possible combination of answers to problems. This systematic process is only mastered in the second stage of FOT (sub stage III-B; full formal function) where it is not in the first stage (sub stage III-A). For example, in stage 1 an individual may use trial and error to reach a solution for a problem, whereas in stage 2 an individual will develop a plan of action before commencing the task. Muuss (1996) also notes that those in the latter stage can produce evidence of their thoughts when required to i.e. they can prove how they have solved a problem or reached a decision.

It is suggested in this thesis, that the acquisition of FOT could involve the acquisition of a complex cognitive organisational system involved in the structure of thought and the application of principles to relevant situations and problems in either real life or hypothetical situations. Adey and Shayer (1997) reasoned that FOT includes a set of reasoning patterns and that these reasoning patterns had to all work together for one to be fluent alone (Shayer, 1979). This suggested network is discussed below including the possibilities for the processes involved in acquiring and making these components of this cognitive network of capabilities. This section demonstrates the potential complexity and intricate thought necessary for FOT capabilities.

Breaking Formal Operations Down

Formal operations may be a combination of rudimentary cognitive skills reaching a mature and refined way of functioning that, when working in conjunction, then allow for abstract thought and looking beyond the present reality to hypothetical situations and the unknown future. Hence, whilst it is known that not all people reach this stage of sophisticated thought, this may be a result of many reasons that may have hindered the development of one of many cognitive functions making up a network of skills contributing to the final formal operational thought.

There may be scope to consider the maturation of these cognitive components over development in comparison to the development of a computer's central processing unit over time. The first basic computers used instructions on magnetic tapes to run their systems and have improved due to technological advances to now be the highly sophisticated desktop

computers, laptops and tablets that we use today. From magnetic tapes, a central processor was developed with certain rules programmed into it for the computer to run and produce relevant outputs. Furthering this idea, these units then consisted of rules for a computer to follow to allow for more sophisticated outputs and processes. Nowadays, in an ever developing culture of exciting information technology, computers contain central processors with a huge amount of rules and guides that allow for them to function in such an intelligent way. This evolution of the acquisition of an input to output conversion, to a specific rule then to a general rule that allows for a principle to be applied to numerous scenarios and allows for the ability to know the correct scenarios to apply these certain rules to can be likened to the cognitive development that a child and young adult goes through.

This development and maturation includes the development of all the cognitive tools and bodies involved. In terms of Carroll's (1993) model of intelligence, it is possible that these cognitive abilities are processes involved in and influencing all of the second order abilities (i.e. these abilities are part of 'g'). These processes and abilities are internal speech, relativity, observation and interpretation, self-awareness and intangible concept acquisition and schemata. These are further described in more detail below. The refinement and practice of these allows for the sophisticated and abstract thought allowed and implemented in formal operational thinking.

Internal speech

The first component of the complex network may be internal speech proficiency. Vygotsky (1986) believed internal speech to be the social commentary a child uses early in development, that later becomes self-directed and involves thinking in pure meanings. Internal speech can be described as "a prearticulatory, but otherwise fully elaborated, and temporally organized representation of verbal utterances" (Ackermann, Mathiak & Ivry, 2004, p.14). More informally, and as a result of folk wisdom, it may be referred to as 'the little voice inside your head'. It may be the ability to think words by silently speaking to oneself (de Guerrero, 2005).

Internal speech has been extensively linked with the ability of reading, for example, in early research by Huey (1908). Studies have consequently shown dysfluent reading in those with poor verbal short term memory (Kibby, 2009). Research has also established links with internalisation and working memory ability, as mentioned by Baddeley & Hitch themselves in 1976. More recent research has shown internal speech measures are related to task

performance, speech action co-ordinates, high level cognitive functioning ability, and can predict later risk for attention problems (Winsler, Diaz, Atencio, McCarthy & Chabay, 2000). Hence, internal speech appears to be related to performance over many cognitive tasks and domains.

A lack of internal speech has been identified to be associated with many developmental disorders. For example, Barkley (1997) noted a lack of inner speech within Attention Deficit Hyperactivity Disorder (ADHD) participants and suggested this as a reason for their impulsive behaviour and lack of high cognitive function. Corkum, Humphries, Mullane & Theriault (2008) found children with ADHD use and employ inner speech in a less mature and relevant way when doing tasks compared to typically developing children, and Berk & Potts (1991) supports this idea and argues that inner speech is important for self-control, forming rules and when forming plans. Within autism research, Wallace, Silvers, Martin & Kenworthy (2009) found diminished inner speech within autistic children from using an articulatory suppression test as an internalisation manipulation on the Towers of Hanoi paradigm. Whitehouse, Maybery & Durkin (2006) also found limitations upon autistic children's inner speech abilities. Frawley (2008) reports similar limitations and notes structural deficits within inner speech when it is used. To consider another disorder, inner speech was found to be lacking in dysphasia when children's school reading and writing abilities were assessed (Njiokiktjien, 1993). Hence the breadth of application of internal speech impairments may be wide within this spectrum of disorders.

However, some research shows increases in inner speech within developmental disorder populations. For example, Berk & Landau (1993) showed a learning disability sample to show higher levels of inner speech relevant to the task in hand when compared to typically developing children. Whilst ideas such as this are controversial they do seem present within the literature; research can be contradictory and in need of clarification, but internal speech proficiency is thought to be involved with many developmental disorders.

Consequently, in terms of formal operational thought and a set of cognitive components making up the underlying network, internal speech may be a tool itself (as in setting a mental goal and seeing the end point to a set of thoughts/reasoning in abstract and hypothetical terms) yet it may also act as an overlaying 'glue' holding the other components together and allowing the sensible and logical integration of the other abilities to allow for abstract thought.

Relativity, observation and interpretation abilities

As teenagers develop they become more astute to their surroundings and others around them. As Piaget described, they become more self-aware and develop ideas of an imaginary audience leading to feelings that they are being watched constantly and that they are of the upmost importance within the world. Along with this development they also develop a personal fable meaning that they see themselves as being uniquely more or less capable than others within certain domains, for example, being smarter than a friend (Elkind, 1967).

These heightened and faulty thoughts tend to peak around mid-adolescence and then taper off slowly with increasing maturity. This dampening effect on these thoughts is potentially key in allowing for formal operational thought. Attributing reactions or consequences for events or seeing oneself as more important than one actually is in reality may be processes that could cause extreme and incorrect attempts at formal operations. For example, an attempt to empathise with a person in a dispute across the cultures will not be accurate or just unless one can see their relative insignificance within the bigger picture of the world as a whole.

Self-awareness and intangible concept acquisition

In order to be able to reason abstractly about the self and the future, one must have self-awareness. Contrary to popular opinion, people do not spend the majority of their time thinking about themselves and may not always be self-aware. Certain events for example may bring such self-awareness, such as performing a public speech. Much in the same way, an act of formal operational thought may trigger a moment of self-awareness and an assessment of oneself in the present and in the hypothetical situation that is relevant. Along with the self-awareness of an individual at the present time and potential future time there may also be a noted distinction between an individual's private self (an internalised representation of the self that is not seen by or expressed to others) and an individual's public self (the self-displayed to others) (Carver and Scheier, 1981). One must also demonstrate self-awareness relevant to how they personally react to events and circumstances in order to be able to predict their future or hypothetical reactions to situations.

Furthermore, in order to be able to reason abstractly about others as well as the self an individual must have knowledge of intangible concepts such as empathy, guilt, jealousy and joy. These need be represented internally in order for an individual to draw inferences based upon them. To manipulate and reason with such hypothetical emotions and situations, an individual must first of all acquire these ideas as mental concepts. This in turn requires a solid

basis of concrete operational reasoning. The interplay of an individual's self-awareness and representations of their different selves (possible and potential) allows for abstract thought about the self and others.

Schemata

Alongside representations of the self and intangible concepts, it is necessary to have an organised set of mental schemata allowing for information stereotypical to a situation or event to be grouped together. This allows us to have some predictability when looking at the future and yet these schemata can be altered according to new and contradictory information if it is necessary. As well as allowing for some predictability, schemata will also allow the cognitive load to be reduced when reasoning about hypothetical events, allowing for more investment into other cognitive resources tapping into abstract thought.

Consequences and relevance of formal operational thought

The ability to think abstractly and be able to apply principles to different situations in hypothetical terms could possibly be tied closely to educational and career success. A low level of proficiency may hinder the ability to plan and see the necessity of present actions for future success. An adult without formal operations remains unable to compare future scenarios and choose the logically superior path of action to take (Fejfar, 2007) with a very tunnel-vision point of view; i.e. they are not open to alternatives and have a set idea of right and wrong actions from which they cannot be shifted. An adult with abstract thought capabilities is open to assessing a situation and looking at alternatives, whilst experimenting with variables and other manipulating factors of a situation. A more open-minded thinker is a result of a formal operations acquisition process.

With regard to teaching in educational institutions, there is a need to understand the levels of cognitive ability in a sample or individual in order to allow for maximal knowledge gain and implementation. Sensible, relevant and stimulating interactions with the environment, within educational settings, can therefore act as a basis for cognitive construction, development and refinement. These interactions and stimulations can only be relevant, appropriate and useful if they are targeted and designed for the correct cognitive ability. Therefore, to develop clear lesson plans and appropriate tools, it is of clear importance that we can assess and evaluate the levels of thought in pupils.

Looking outside of the educational implications there may be real life consequences for people if they fail to reason with intangible concepts and especially if they fail to look to

future and realise consequences from their actions. This may be important within the context of criminality, for example young criminals describing that they had not thought what could or would happen to them if they were caught for a crime that they had committed. Whilst this lack of foresight into future events or consequences may be difficult to grasp for some people it is clearly a problem and one that may be improved through the teaching of abstract thought.

The teaching of formal operations could involve identifying areas that an individual does show abstract thought capabilities for and then teaching these principles to generalise them to different fields. In fact, even in 1984, Halpern notes that educators were criticised for not teaching pupils how to think.

‘A central purpose of education is to provide the conditions which will foster the development of students’ reasoning abilities and schools should do more than teaching collections of discipline-specific facts and concepts’ (p169, Valanides, 1997).

Teaching FOT may take influence from Cognitive Acceleration research. This aspect of research is based upon the grounding of FOT as a concept within Piagetian reasoning and Piaget’s model. Interventions concerning Cognitive Acceleration (CA) are generally based upon three main principles.

Firstly, the idea that there is a general intellectual function akin to the general intelligence ‘g’ and that this g underpins further complex and detailed sub-intelligence functions. Secondly, ‘g’ is held to develop with age and thirdly this development is influenced by nature and nurture, .i.e. maturation such as brain growth, and the environment, for example, cultural influences (Shayer & Adey, 2002). Intelligence and the abilities involved within intelligence models are therefore held to be malleable and influential. Logically, they can then be developed, expanded, nurtured or eradicated altogether. Following these principles from CA interventions, if formal operational thought is involved within SJT performance, there may be scope to influence and improve this intelligence ability, and in turn improved performance upon the SJT and/or job performance.

Nickerson, Perkins & Smith (1985, p.324) said that

‘If teaching thinking cannot be done, and we try to do it, we may waste some time and effort. If it can be done, and we fail to try, the inestimable cost will be generations of students whose ability to think is less than it could have been.’

CA interventions are further based upon the understanding of the makeup and structure of cognitive functioning within a FOT framework. These cognitive functions are then targeted in a stimulating and intellectually enhancing environment. They take focus from teaching 'subject matter' to teaching 'intellectual capabilities and resources'; theoretically, they aim to develop cognitive functioning and integration to allow for more critical and efficient thought and analysis of information, regardless of the field that the information is specific to. CA relies upon teacher strategy and the emphasis on principles such as metacognition and reflection; i.e. thinking about thinking. Students are asked to describe how they see and work out problems. They are encouraged to reflect upon behaviour and think about their thinking. This can be described as metacognition and the ability is a different one to FOT. In fact, it is noted in Adey & Shayer (1993) that it is widely accepted that students are more likely to develop wide-ranging thinking skills if they do think about their own thought (Nickerson et al., 1985). In fact, Flanagan, Ortiz & Alfonso (2013) note recommendations and interventions for low proficiency fluid reasoning including using demonstrations to externalize reasoning processes, targeted feedback, cooperative learning, reciprocal teaching, teaching metacognitive strategies and comparing new concepts to previously learned ones.

Typically, past CA interventions have taken place targeting children of different ages and with the intent of showing mental ages that are superior to chronological ages based on a Vygotskian idea that learning is only 'good learning' if it 'is that which is ahead of development' (Vygotsky, 1978, p.89). For example, CA interventions produced long term effects and effects outside of the targeted subject with improvements in GCSES grades in three core subjects (Maths, Science and English) (Adey & Shayer, 1993). Further successful interventions include Adey and Shayer (1994) and Shayer (1999a). There is also some evidence that the teaching of FOT principles can be generalised by putting students in situations where they have mastered concrete operations and then find a problem unsolvable without higher levels of thought (Adey and Shayer, 1994). This suggests that students have to construct their own networks of cognitive processes and consequently, their own higher level thinking.

Specifically being concerned with targeting the development and acceleration of abstract thought towards the formal operational stage of development, interventions would not target children, but could potentially target individuals of any age once they have demonstrated the capacity for concrete operational thought constructs. There appears to be potential to develop educational programmes (using visual media such as videos or computer programmes) or

modules within educational establishments (for example, a module within a university degree to accelerate intellectual tools) to do so.

Currently there are some differing opinions over the proportion of the population that actually reach the higher levels of cognitive abilities. It appears that adults of a typical level of intelligence (IQ approximately 100) remain at the concrete operational stage and do not progress further. Hence, potentially a large proportion of the general population fails to reason in abstract terms (Neimark, 1979). Approximately two thirds of high school USA students are reported not to think in FOT (Huitt & Hummel, 2003). Furthermore, even less of the population are expected to go on to think post-formally. Thinking post-formally (proposed by theorists known as neo-Piagetians, Andrade & May, 2004) is more complex, and incorporates making decisions using situational constraints and circumstances, and using and assessing emotion with logic to give relevant principles, including flexible and justified ideas and bases (Kail, Cavanaugh & Ateah, 2006). This low proportion achieving this level of thought is potentially due to a typical person finishing a normal undergraduate degree at approximately 21 years old, and subsequently leaving formal education. To develop full post-formal thought it may be that one has to remain in education or another stimulating environment until later in life.

Formal Operational Thought and Contemporary Intelligence

As noted, FOT within Piagetian reasoning fits into modern theories of intelligence (Carroll, 1993; CHC theory) as an element of the second order abstract construct of fluid intelligence. It can therefore be classified as an abstract thought entity that is involved in the deeper and more specified levels of intelligence. Fluid intelligence can loosely be described as the ability to reason and deal with complex information. The first study to find relationships between fluid intelligence and formal operational thought was Hooper, Hooper & Colbert in 1985.

More recently, further distinctions regarding FOT have been made by Emick & Welsh (2004). They have suggested that the large individual differences observed in FOT proficiency mirror those seen in executive function ability. They define executive functions as cognitive processes that underlie future oriented and/or goal-directed behaviour, for example, working memory, inhibition and self-monitoring. They note that the systematic problem solving skill from FOT is very similar to the skills associated with executive function. Their results suggest that FOT has incremental validity over and above fluid reasoning when explaining results on an executive function task. However, they do note

limitations of the study and the results are only applicable to one task with a small and unbalanced sample. Nevertheless, this is preliminary evidence suggesting the specific ability of FOT may be closely involved or related to executive functions.

Assessing Formal Operational Thought

There are various practical tests for FOT; for example, the pendulum, the balance beam or billiard balls bouncing off plane surfaces (Adey and Shayer, 1994). One very famous test used by Piaget, is one that uses a pendulum and asks the participant to estimate how quickly the pendulum will swing according to differently weighted objects and lengths of string; this uses the principle of the pendulum and allows for principles of deduction and application to be used. Research has been plentiful, although methods have rarely strayed from Piaget's typical tests (e.g. Gray, 1973; Kenny, 1978; Hudak & Anderson, 1990; Danner & Day, 1977).

Other tests may generally describe a situation or scenario offering options that have to be considered logically and in a way that is required to solve the problem at hand, often incorporating intangible concepts, for example, the future. Adey, Shayer & Yates (2001) have consequently developed a set of questions assessing FOT and the thought levels surrounding this cognitive stage for their numerous CA programmes. These questions are further discussed in detail in the method section as a basis for the measure used in study 2.

Field Specific Tests: Medicine

After finding a non-field specific test to assess formal operations (as described above) it may also be possible to develop field-specific tests so that they may be used within career and education development and assessment where it may be necessary and in public interest (safety and economic factors) to introduce higher levels of abstract thought. There is some criticism in the field of FOT assessment that tests are domain specific (i.e. science and/or mathematics) as tests tend to fall into these areas (Adey and Shayer, 1994).

Within medical selection and assessment, it may be of great interest to consider assessing levels of abstract thought and then to consider teaching those with lower levels of formal operational thought different ways of thought through educational bodies and resources. This may not only improve performance upon the SJT, but in turn alter thought processes in general and improve job performance through more reasoned, logical and maturely considered behaviour. FOT may have important consequences for diagnostic techniques and decisions, medication choices and follow up routines for patients. Looking at symptoms and

reasoning through possible causes, then referring back to symptoms and working in such a forward reasoning way is the correct way to diagnose a patient; it leaves the least room for error and allows for personal and tailored assessment rather than drawing upon past experiences with other patients and potentially choosing the statistically most common explanation for the symptoms upon display. FOT may allow for this type of principle within diagnosis to be picked up by practitioners and used in day-to-day practice. It may also allow for foresight into the effects of the situation, potential medication or the illness itself upon the patients. This well rounded care that considers all options, perspectives and consequences is surely the care that every individual strives for from a healthcare professional.

Furthermore, there may be benefits from a personal and professional view for the doctor themselves. Career progression and advancement, along with managing a busy and stressful career with a social and family life, may be made easier by the considering of arrangements, possibilities, intangible emotions of the self and others and future planning; all of which can be grasped through FOT.

Whilst this management of life and career may lead to a happier and more satisfied individual, it may also prove to benefit professional bodies, the government and in the grand scheme of things, the public themselves. Through helping healthcare professionals to manage their lifestyles, careers and to show a more professional manner with more correct and quicker diagnoses there may be benefits shown through money invested and used by the government and NHS. If training of FOT is introduced into an earlier stage of the career as well; for example, within medical school, it may also prevent students dropping out of the course if they cannot perform or cope with the path they have chosen. These are all directions for future researchers to consider as whilst such questions demonstrate the importance of FOT, they are beyond the scope of this thesis.

Following this introduction, the hypotheses are noted for this study below. This study aims to develop the measure of FOT and provide the validity data for this by examining the differences in cognitive performance (or lack of) in the two different sample groups.

Hypotheses

Hypothesis 1: Dyslexia and DST-S performance

The DST-S battery of tests is used in the measurement and assessment of risk and severity of dyslexia. The tests were designed for this purpose and have been used extensively in practice and research (e.g. Fawcett and Nicolson, 1992, 1994, 1995b; Moores, Nicolson & Fawcett, 2003; Taroyan, Nicolson and Fawcett, 2007). Dyslexia itself is primarily a disorder associated with reading, writing and phonological difficulties. Some theories of dyslexia even suggest that motor ability is impaired due to cerebellar abnormalities (Nicolson, Fawcett and Dean, 2001).

Hypothesis 1 therefore states that those with dyslexia will perform worse upon the DST-S tests that target abilities associated with cerebellum related skills, phonological processing, reading and writing ability (all noted in the literature to be impaired as a result of dyslexia).

These tests include Rapid Naming, the One Minute Reading, the Postural Stability, the Phonemic Segmentation, the Backwards Digit Span, the Nonsense Passage and the One Minute Writing tests (all tests excluding nonverbal reasoning which is considered a test of IQ related ability).

Hypothesis 2: Dyslexia and nonverbal reasoning performance

The DST-S test of nonverbal reasoning can be classed as a test closely related to the assessment of IQ and not one of the skills primarily associated with weakness of the disorder dyslexia (Fawcett & Nicolson, 1998).

Hypothesis 2 therefore states that there will be no effect of group membership (dyslexia or typical) on performance upon the non-verbal reasoning test.

Hypothesis 3: Dyslexia and formal operational thought

FOT is a higher level thought ability that is not expected to be affected by the developmental disorder of dyslexia, i.e. the impairments within this disorder are independent to the abilities needed for FOT.

Hypothesis 3 states that there will be no effect of group membership (dyslexia or typical) on performance upon the formal operational thought measure.

Hypothesis 4: Formal operational thought and lower cognition

FOT has been explained as a higher level reasoning ability that allows for abstract thought about intangible, complex and hypothetical scenarios. It is proposed as an abstract form of knowledge (Carroll, 1993) and it is expected that it works when complex reasoning is required by a cognitive challenge. It is therefore hypothesised that FOT will not be related to performance upon these DST-S tests that require only basic level cognitions, or the nonverbal reasoning test which assesses general intelligence (FOT is hypothesised to be conceptually above this level of thought). FOT is not a primary or basic cognition, it is a more complex, abstract and procedural knowledge that can assist and help with cognitive challenge or difficulty and reasoning.

Hypothesis 4 states that formal operational thought proficiency will not directly affect performance upon DST-S tests.

Method

Study Design

A cross sectional within-participants design with analyses that looked at between and within group differences in DST-S dyslexia battery subtest performance and FOT test scores.

Participants

Typical sample: A sample of 27 students from the University of Sheffield. The sample were of mixed gender but predominantly female (female $n=26$, male $n=1$) and were between the ages of 18 and 23 years old (mean age = 19.8 years old, $sd=1.26$). Ethical approval was obtained from the University of Sheffield's Institute of Work Psychology Ethics Committee. Participants for this sample were obtained through the Psychology Departments Research for Credit system where first year undergraduates can opt to participate in an experiment or study in turn for credit rewards so that they too can use this system when doing their final year projects.

Dyslexic sample: A sample of 23 students from the University of Sheffield. The sample were both male ($n=7$) and female ($n=16$). Participants were between the ages of 18 and 37 years old with a mean age of 23.7 years old ($sd=4.8$). Ethical approval was obtained from the University of Sheffield's Institute of Work Psychology Ethics Committee. Participants for this sample were obtained through the Disability and Dyslexia Support Service (DDSS). Contact was made through emails and this went on to allow the sending of a consent form

and letter of explanation about the study out to member of this University service. An incentive of £10.00 was offered to each participant to compensate for their time.

Materials and Procedure

DST-S tests

Tests selected from the DST-S were the Rapid Naming, the One Minute Reading, the Postural Stability, the Phonemic Segmentation, the Backwards Digit Span, the Nonsense Passage, the Nonverbal Reasoning and the One Minute Writing tests. Some of these tests required specific apparatus and materials; these can be seen listed below. The specific ability each test assessed is noted and the rationale was to give a clear picture of basic cognitive abilities. The skills were rudimentary and basic ones, for example, reading and writing, and can be classed as lower cognitions for this study (as opposed to higher level thought and formal operational thought as higher cognitions).

The test retest reliability correlations of the chosen tests have been reported and can be seen below in Table 6 (from Fawcett & Nicolson, 1998).

Table 6. Test-retest reliability for specified DST-S tests.

DST-S Test	Correlation
Rapid naming	.68
One minute reading	.90
Postural stability	.71
Phonemic segmentation	.90
Backwards digit span	.64
Nonsense passage reading	.92
Nonverbal reasoning	.75
One minute writing	.87

All tasks required the use of an instruction manual by the researcher which contained the information noted below in the individual test procedure section.

All tasks also required the use of a DST-S marking sheet that is used by the researcher to mark down the scores for each individual task upon. One score marking sheet is required per participant.

Individual test procedures

- (1) **Rapid Naming:** The researcher required the necessary stimuli sheet with the pictures on for the participant to name. They also required a stopwatch to time the participant in doing so.
- A DST-S test used to assess speed and accuracy of naming objects. Those with dyslexia have shown to be slower at this, even when naming familiar objects. This test was chosen to give a measure of processing accuracy and speed.
- The participant was asked to name the pictures on the stimuli sheet, starting at the top left hand corner and going along each row until they reach the bottom right hand corner. They were asked to do this as quickly as possible and were told that they will be timed whilst doing so. The researcher then demonstrated by naming all the pictures on the card whilst pointing to each picture in the correct order.
- The participant then completed the task and the researcher timed them on the stopwatch. The number of errors and the time taken were recorded on the score sheet.
- (2) **One Minute Reading:** The researcher required the necessary stimuli sheet with the words listed upon for the participant to read. They also required a copy themselves to mark any errors on and a stopwatch to time the one minute period.
- A DST-S test used to assess reading ability and fluency. The participant was required to show accuracy and speed in their reading. This test was chosen to give a measure of the reading ability of the participant. Reading is typically the most problematic cognitive area for an individual with dyslexia and hence it would be expected that the dyslexia sample would perform worse than the typical sample upon this measure.
- The participant was asked to read aloud as many words as possible in one minute. The participant was timed by the researcher using a stopwatch over a period of one minute. The stimuli to read were listed on a page of individual words graded in difficulty. There was a practise test of six words for the participant to attempt and then he/she goes onto the main test if they were successful in the practice. The form has four columns of thirty words (this is opposed to children who use a stimuli form of four columns of twelve words.) There are two versions of the form (A or B) and either could be used. Form choice was varied per participant.
- The researcher explained in which order to read the words and columns and started the stopwatch. Any errors made were recorded by the researcher marking with a water soluble pen on a sheet of acetate over another copy of the form. The number of passes

made was also recorded and how far through the stimuli the participant managed to read. One mark was given for each correct word read. If the whole sheet of words had been read, the time on completion was recorded. If this happened, two points were added to the score for every second less than sixty the participant finished at. (For example, finishing at 57 seconds resulted in an extra 6 points added onto the score.) The test was discontinued if the participant made five consecutive errors or if they had passed on five or more consecutive words, the researcher asked if they could read any of the next five words and gave credit for correct responses.

(3) **Postural Stability:** The researcher required the balance tester (see Appendix V for diagram of this) and a blindfold for the participant. The balance tester was a yellow device within the DST kit with a sliding collar on the internal shaft. To calibrate this to a certain weight before using for testing a pair of scales was also needed.

- A DST-S test used to assess the participant's balance and underlying cerebellar abnormalities. This test was chosen to give a measure of motor skills and underlying cerebellar difficulties that may be linked to dyslexia. Some theories of dyslexia hold that cerebellar abnormalities are the cause of dyslexia and hence, as well as difficulties with literacy, difficulties with motor skills such as balance would also be expected.

In preparation for this test, the balance tester calibration was checked. This was done by pressing down the tester upon kitchen scales (5Kg). The tester was calibrated to apply a 2.5 Kg force.

The researcher asked the participant to stand up straight with their arms by their sides. They were asked to remain as upright and straight as possible and warned that the researcher was going to apply a small amount of pressure to their back. The balance tester was placed into the back gently, the collar slid slowly down the shaft, stopping pushing just before the collar meets the pommel (see Appendix V again for a diagram).

The participant was then blindfolded and the balance tester used to apply the pressure. The amount of sway was assessed as either 0 (no movement), 1 (a slight sway), 2 (rising up on toes), 3 (small step forward/marked sway), 4 (marked step forward), 5 (two steps forward) or 6 (several steps/loss of balance).

The participant was then asked to resume the original position and told that the procedure would be repeated. The researcher proceeded to repeat the procedure as before. The scores were noted again upon the score sheet.

Finally, the participant was asked to outstretch their arms at 90 degree angles in front of them and told that the researcher was going to push them again. The researcher went on to do so and recorded the amount of sway again.

(4) **Phonemic Segmentation:** No specialist equipment was required for this test.

- A DST-S test used to assess the participant's phonological skills and working memory ability.

The researcher explained that they were going to play a short game involving the breaking of words down into sounds. The researcher said a word with two constituent sounds and asked the participant to repeat back the word without one of the sounds. An example was given to the participant with the researcher stating 'words have different sounds, if you listen to me say eyelid you can hear eye and lid, or if I say headache, the two sounds are...' The participant may have needed prompting but a correct response signalled that they could go on to try three practise words and then twelve words in the main test (see Appendix VI for stimuli used).

A mark was scored for each correct response. If they got the first four words wrong, or three consecutive later errors, the test was discontinued.

This test also included spoonerisms; a DST-S test that examined the participant's phonological processing abilities.

The researcher explained that they were going to play a game involving swapping the sounds of words around. It was then explained that the researcher was going to say a couple of words and the participant should swap the first sounds of each word around, giving an example, 'hello Jack becomes jello hack.' The participant was given four practise sets, for example, 'try this one, red hat.' For a correct response ('hed rat') the participant was told 'yes, that's right.' The test was discontinued if the participant scored one or zero out of the four examples.

After the practise set the actual test of three sets using names was attempted by the participant after an example was given (Michael Jackson becomes Jichael Mackson). (Full stimuli shown in Appendix VII).

The participant scored one mark for each correct word.

(5) **Backwards Digit Span:** No specialist equipment was required for this test.

A DST-S test used to assess working memory ability. Normal digit span is included in IQ tests and is modestly associated with dyslexia. Backwards digit span is particularly vulnerable to developmental disorder dyslexia.

The researcher then told the participant a list of numbers at approximately one number per second (for example, 4, 2, 6, 10) and the participant was asked to repeat the numbers back to the researcher in the reverse order (for this example, 10, 6, 2, 4).

The researcher then told the participant that there would be two numbers, and then three, and then four and so on. The participant was given three practise trials and after they correctly answered these, the main test was administered. (See Appendix VIII for full set of practise and main test stimuli used.)

A mark was given for each correct reversal of a sequence. Discontinuation instructions held that if a participant answered two incorrectly at any length (that is, if the participant got the first right, but the second two wrong at the 4 digit length, you would continue to do both at 5 digit length).

(6) **Nonsense Passage:** The researcher required the necessary stimuli sheet with the nonsense passage on it for the participant to read aloud. They also required a stopwatch to time the participant reading.

This test was chosen to assess reading ability and speed of nonsense and non-nonsense words in a text.

The participant was informed that they would be shown a passage of printed text and that they would be asked to read this passage aloud to the researcher. They were also informed that the text will make sense but would contain lots of nonsense words (i.e. words that are completely made up that mean nothing to the researcher nor to the participant). The participant was asked to read the passage naturally, and to attempt the words unless they did not want to and they then could say pass and move on, but told that they would be timed on their effort. After the explanation the participant was given the practise text sentence to read aloud. Upon completion of this, the main test began and the researcher timed the participant using the stopwatch. Time taken and errors made for nonsense words and non-nonsense words are recorded. The final score was therefore one of time taken less the number of errors made.

(7) **Nonverbal Reasoning:** The researcher required the necessary stimuli sheet which had the reasoning problems upon it for the participant to solve. The participant was also given a sheet of numbered paper to write their answers on. The researcher also required a stopwatch to time the sections of the nonverbal reasoning task.

This test was chosen to assess reasoning ability in a non-text format. It is often used a measure of typical intelligence within a dyslexia battery; those with dyslexia tend to not show poorer performance than typical individuals.

The participant was told that there would be three sections to this task. They were given the numbered sheet to write their answers on and a pen. Following this, the researcher went through the instructions for section A (what is next in the sequence questions) and through the practise with the participant ensuring that they understand what to do and how long they have to do it. The task then took place with the researcher timing for 1 min 30 seconds. The researcher then explained section B (analogy questions) and went through the practise question for this section with the participant. Following this, the main test went ahead with the researcher timing for 1 minute 30 seconds again. Lastly, section C (grouping questions) was explained to the participant and the practise for this section run through. The participant was then timed for 1 minute to complete the main questions for section C. The participant's answers were then recorded on the researcher's score sheet. The score was a total of correct answers on these questions.

(8) **One Minute Writing:** The researcher required the necessary stimuli sheet with the text passage on for the participant to copy. The participant was also given a sheet of blank paper to write their attempt at copying the passage on. The researcher required a stopwatch to time the one minute writing period.

This was a test chosen to assess writing accuracy and speed. Those with dyslexia often demonstrate problems with writing due to issues with literacy that follow from reading impairments.

The participant was told that they would be shown a passage of printed text and then would be asked to copy this as quickly and as accurately as possible. They were informed that they would be timed for one minute and that they should write as much as possible in this time, keeping it accurate and legible for the researcher to read. On understanding this, the researcher then showed the practise sheet to the participant and asked them to copy this sheet. After completion of this the main test began. The

participant was given the plain sheet to write upon and the researcher timed them writing for one minute. The number of words written and errors made are recorded on the score sheet. The total score was therefore one of words written in the one minute less the number of errors made.

Formal Operational Thought

The FOT worksheet required no specialist equipment besides the problem worksheet (see Appendix VIII) for the participant to work through.

One worksheet was required per participant. The total score given was that of the number of questions correct on the paper.

Development of formal operational thought methodology

Investigation and research concerning previous testing of FOT has taken place (e.g. Taylor & Francis, 1976; Gray, 1973; Kenny, 1978; Hudack & Anderson, 1990; Danner & Day, 1977) although, within the literature methods varying against Piaget's typical tests are scarce. Furthermore, the assessment of FOT appears to remain relatively untouched, in terms of theory and empirical research, since approximately 1990. The most recent use and practical application of formal operational thought testing and assessment comes from Adey, Shayer & Yates (e.g. Shayer & Adey, 1993; Adey & Shayer, 1994; Shayer, 1999; Adey, Shayer & Yates, 2001). Their focus has been within the Cognitive Acceleration field. They have extensively used assessments of Piagetian reasoning within these interventions and programmes.

The materials used in their Thinking Science DVD (3rd edition, 2001) were used to base some questionnaire problems on for this FOT questionnaire. These tests were used in the CASE project and the science reasoning tasks were part of a series developed by the team 'Concepts in Secondary Maths & Science' at Chelsea College, University of London, in the period 1973/78 in order to investigate the relationship between the optimum Piagetian level at which a pupil can function and the understanding of Science which he or she can achieve.

There are two main themes of science tasks used from Adey, Shayer and Yates' work; Volume & Heaviness Tasks and Balance and Equilibrium Tasks.

The Volume and Heaviness Tasks are based upon most of the chapters in J. Piaget and B. Inhelder, 'The Child's Construction of Quantities', Routledge, London, 1974. The task works

on the principle that initially, for the participant, there is a global concept of 'size' in which mass, weight, volume and density are not differentiated clearly from each other.

The Balance and Equilibrium Tasks are based upon chapter 11 of Inhelder and Piaget's "The Growth of Logical Thinking", Routledge, London, 1958, and investigates the participant's ability to recognise and use inverse proportions in a simple beam balance.

From these two tasks, a measure incorporating questions that are based upon both task topics was developed.

The test is hierarchical with measures assessing initially early, middle and mature concrete operational thought followed by concrete operational generalisation, early formal operations, mature formal operations and finally, formal operational generalisation thought. The breakdown of the questions according to these categories of thought can be seen below in Table 7.

Table 7. Breakdown of formal operational thought questions; origins of and level of thought assessment.

Section & Question Number on developed questionnaire	Adapted from Thinking Science Question (Adey, Shayer and Yates, 2001)	Thought Assessment Level
Section A (a)	1 (Volume & Heaviness Task)	Early concrete
(b)	2 (Volume & Heaviness Task)	Early concrete
(c)	5 (Volume & Heaviness Task)	Middle/Mature concrete
(d)	6 (Volume & Heaviness Task)	Middle/Mature concrete
(e)	10 (Volume & Heaviness Task)	Concrete generalisation/early formal operations
Section B (a)	13a (Volume & Heaviness Task)	Concrete generalisation/early formal operations
(b)	13b (Volume & Heaviness Task)	Early/ mature formal operations
Section D (a)	6 (Equilibrium & Balance Task)	Early formal operations
(b)	14 (Volume & Heaviness Task)	Mature formal operations
(c)	9 (Equilibrium & Balance Task)	Formal operations generalisation

(d)	10	Formal operations
	(Equilibrium & Balance Task)	generalisation

(*Section C is that of abstract thought non-scientific problems as described below).

Additional question development

As noted earlier, due to concern over the FOT testing focusing upon scientific questions only, as opposed to domain general testing, it was necessary to include an assessment of formal operations that steered away from scientific principles themselves. Furthermore, it was necessary to target problem solving skills that focused upon hypothetical and intangible scenarios that had many different possible answers that required cognitive consideration. The researcher developed three additional questions to address these issues. These three developed questions and the explanation of the correct answers can be seen below. Where possible, the explanation of answers and thought behind a given answer was requested as part of the question due to this being used as an element of FOT testing in past research (e.g. Walker, Hendrix & Mertens, 1979). The measure development is further discussed below.

Measure Development

It is important to note that the FOT questions were based directly on the very longstanding development of the Piagetian concept of Formal Operational Thought (Piaget, 1952) and then developed and fully established for use in UK schools by the Adey, Shayer & Yates (2001) programme. The researcher was therefore utilising a pre-existing and validated set of questions rather than developing a new scale as such. In particular, the theoretical constructs of early concrete operations, mature concrete operations, early and mature formal operations do not sit comfortably within the standard psychometric approach to scale development, in which the ideal is that if the scale attempts to assess three constructs, then these constructs should be to some extent orthogonal, or independent. Given that the Piagetian constructs are actually cumulative, with formal operations built on top of concrete operations and so forth, it is by no means clear that the traditional psychometric approach to scale development is applicable.

Nonetheless, it was necessary to establish whether the scale was indeed fit for purpose in working with university students, in terms of providing a measure that discriminates between different individuals, and does indeed yield interpretable results.

For completeness, it is valuable to provide further detail on best practice for scale development if one was indeed developing a new scale. Best practice approaches (e.g. Simms 2008) discuss the methods of scale construction as a dynamic process of construct validation, rooting the guidelines within classic work from Loevinger (1957), Cronbach & Meehl (1955) and Campbell & Fiske (1959). Simms presents a unifying framework with three main steps of scale construction between the initial ideas for a new measure and the final complete scale. Loevinger's (1957) theory-driven approach to scale construction informed the three main stages of substantive validity phase, structural validity phase and the external validity phase.

To summarise the guidelines within the three phases, Simms (2008) recommends that a larger-than-needed set of questions be developed, and then be administered to a selected group of the target population. The resulting data may then be analysed using confirmatory factor analysis to establish the underlying factor structure. Clearly there should be a good fit to the intended structure. This allows a set of items to be determined for each construct, and a Cronbach alpha item analysis will then indicate the set of most reliable items within each factor-set. Low reliability or cross-factor items should then be discarded, leading to a relatively reliable and sensitive set of items for each factor. In general, the sample size required for such an analysis depends upon the number of constructs in question. Nunnally (1978) recommended having 10 times as many participants as variables for exploratory factor analyses.

The three developed questions can be seen below. Following the question description there is a justification and explanation of the correct answer.

Question 1:

'Sing Right' is a competition to find a new successful recording artist who will sell records and earn money for themselves and the record company that they are signed for. There are two judges, Mark and Ian. Together they have to choose a winner from two remaining acts.

Darren is a 25 year old male singer whom Mark considers has a lot of potential to be successful in the public eye but is unlikely to sell many records once the competition is over. He also considers that Darren is a handsome man with a good sense of humour. Ian also likes Darren as an act and thinks that he has a lot of charisma.

The second act is a 32 year old female singer called Daisy. Ian prefers Daisy to Darren and considers that she will sell many records in the future. Mark believes she will sell many

records and be successful both during and after the competition. He does not get along with Daisy though and finds her to be rude and quite arrogant. Ian thinks that Daisy is pretty but that she does not have a sense of humour.

Which do you think is the most likely outcome of the competition?

- (a) The two judges decide that Darren will win the competition
- (b) The two judges decide that Daisy will win the competition
- (c) The two judges will not be able to agree upon a winner

Why have you chosen this answer?

.....

Question 1 answer:

The correct answer is Daisy because she was the only act who both judges considered would be successful selling a lot of records which was the aim of the competition.

This problem required the participant to consider each of the singing acts and how likely they would be to win the competition. The participant was required to consider both of the judge's points of views and the merits of each act and furthermore must eliminate irrelevant information and focus upon the actual purpose of the competition. For example, 'a sense of humour' is not relevant to the competition, yet 'selling records after the competition' is. Reasoning about the judges' opinions and considering the facts of the problem will demonstrate the presence of formal operational thought.

Participants were asked to choose one of three answers to allow for distinguishable marking. Finally, they were asked to explain their decision so that thought processes and understanding of the correct answer could be shown.

Question 2:

An audience watching a ballroom dancing competition final are asked to vote for a winner out of three finalists, Jane, Jamie and Henry. A panel of judges have passed comments on the finalists' performances, although the final decision of who will win is entirely decided by the audiences' vote.

The judges have described Jane as a very talented dancer who technically performs the dance moves almost perfectly nearly all of the time. Her performances have been described as captivating and a joy to watch. She has not been unpopular with the audience, although she has failed to win their hearts fully as she has been consistently good throughout the competition and not shown any improvement or development.

Jamie has been described as the joker of the competition who entertains the audience, although his dances are not so enjoyable to watch as they are technically poor and lack finesse. He is a high spirited and handsome young man who has proven popular in particular with the female members of the judging panel and the audience.

Henry is older than Jamie but younger than Jane. His dances have been described as challenging and technically quite accurate. He has shown improvement over the competitions dances from the start of the competition, although his technical ability is still not up to the same standard as seen in Jane's dances. Henry has somewhat charmed audiences and included elements of comedy in his dances, although judges have described these additions as juvenile and as undermining to the competition itself.

Who do you think is most likely to go on to win the dancing competition?

- a) Jane
- b) Jamie
- c) Henry

Why do you think this person will win the competition?

Who do you think is most likely to come second in the competition?

Why have you chosen this answer?

.....

Question 2 answer:

Henry is most likely to win the competition.

Unlike the first thought problem, this question required the participant to consider two dimensions (popularity and technical ability). Henry scored fairly highly for both of these

dimensions whereas the other two contestants scored much lower on either one of the dimensions.

The participant was again required to discard irrelevant information and focus on the facts that affected the audience's decision. They were asked to consider the hypothetical situation and reason about the positive and negative attributes of each contestant. The problem was worded using negative descriptive terms and many unnecessary personality descriptions in order to complicate the problem clarification process further. Participants were also asked to justify their answer to show thought processes and understanding of why the chosen contestant would be the winner.

Jane was most likely to be second in the competition as whilst Jamie was popular, he lacks technical ability and Jane had been consistently popular.

Question 3:

On a planet in outer space there are two species of creatures that have been identified, these are species x and species y.

If there are no x's that aren't slim and no y's that aren't x's, then which statements are always true?

- (a) There is not one slim creature that isn't an x
- (b) All y's are slim
- (c) Any creature that is slim is also a y
- (d) None of the above

Question 3 answer:

Answer: 'All y's are slim'

This problem was developed in order to use algebraic functions to assess formal operational thought. The mental representation of an 'x' and a 'y' in this problem required formal operational capacities. Furthermore, it required complex cognitive capabilities to reason about these abstract entities and to be able to manipulate hypothetical, intangible and unknown situation and creatures.

General Procedure

The data collection was carried out within the University of Sheffield in the Institute of Work Psychology building inside a private office.

Following introductions the researcher explained that the study had two parts and that the study was likely to take approximately one hour to complete both parts. Participants were also informed that they had the right to withdraw from the study at any time and that all data gathered was completely confidential. The researcher also ensured that informed consent had been obtained through a signed consent form. If they had not already done so, they were asked to read the information letter and sign the consent form (see Appendix X).

The researcher instructed the participant through each part of the study and explained instructions clearly, allowing time for practices for the tests that required them and marking down scores as necessary upon the score sheet. 'Tick crosses' were used, as opposed to plain crosses, to avoid the participant thinking they had performed badly upon a test (although where possible showing the participant the score sheet was avoided). Right and wrong feedback to questions was not allowed, but participants were encouraged wherever possible to continue and to strive to their best performance level. Discontinuation instructions allowed for a fairly smooth transition between tasks and time to motivate the participant if necessary.

The order of the parts of the study was random and varied for participants to reduce order effects. Once the tests were completed the participant was asked if they had any questions or if they would like to receive feedback from the study. Following this, the typical sample participants were given their research credits (via the online internet system) and the dyslexia sample participants were given their £10.00 in cash.

Results

The participant's raw data scores for the DST-S tasks were standardised relative to the established normative group for the age range (mean= 100, sd=15).

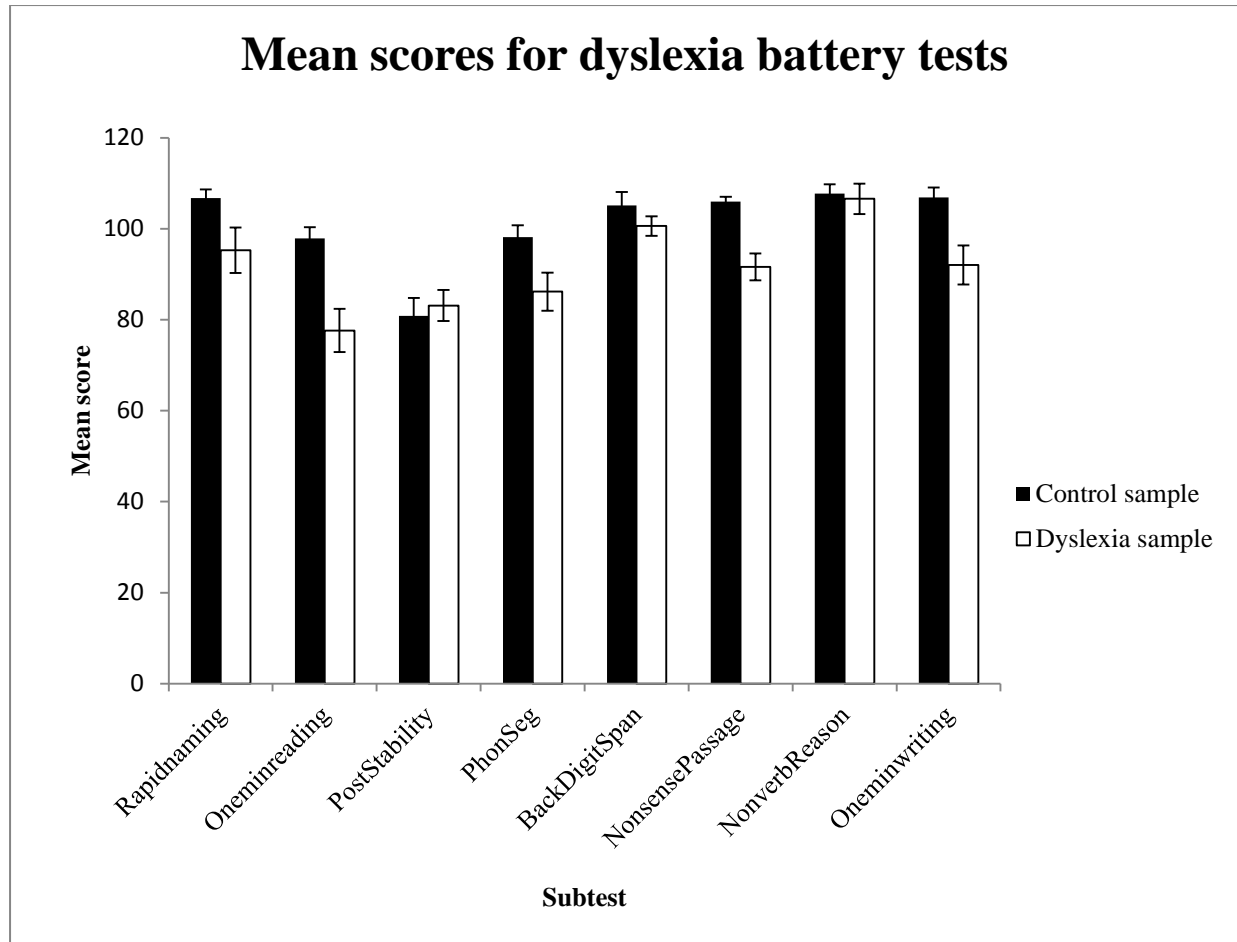
Higher scores indicated a higher level of ability, for example, a score of 110 on the one minute reading test indicated a better reader than a participant with a score of 95.

DST-S General Performance

The overall mean for the scores for the typical sample was 101.16 and the standard deviation 9.08. The overall mean for the scores for the dyslexic sample was 91.63 and the standard

deviation 9.34. Comparative performance between the groups across all the DST-S tests can be seen below in Figure 7.

Figure 7. Dyslexia battery test performance for typical and dyslexia groups.



(*Typical N= 27, Dyslexia N= 23. Error bars showing standard error).

The graph suggests that the typical sample generally performed better on all but one of the subtests, in particular the one minute reading and one minute writing tests. Postural stability scores were relatively low for both groups and this may be due to the task itself and subjective nature of the examiner scoring the balance of a participant; the typical sample in fact scored slightly worse than the dyslexic group. Non-verbal reasoning task scores were similar for both groups. This was to be expected to be the case as non-verbal reasoning is the closest measure in this battery to that of typical intelligence which the samples should not differ on. This test was in fact not one of the diagnostic components of the DST-S, and is described in the manual as a potential strength and assessment of general intelligence.

However, the reading and rapid naming tests were those that a dyslexic sample would tend to struggle with, as shown above.

There were significant differences between samples on the rapid naming test, $t(49) = 2.74$, $p=.009$, the one minute reading test, $t(49) = 3.77$, $p<.001$, phonemic segmentation, $t(49) = 2.80$, $p=.001$, nonsense passage reading, $t(49) = 4.70$, $p<.001$ and one minute writing, $t(49) = 2.22$, $p=0.030$. The typical sample performed significantly better than the dyslexic sample on these tests.

Hypothesis 1 (the dyslexia and DST-S hypothesis) is therefore mostly supported; the only tests that were expected to be related to dyslexia group membership that did not show significantly better performance for the typical sample were postural stability and backwards digit span. The fact that there was no significant difference between the groups on nonverbal reasoning scores suggests that Hypothesis 2 (the dyslexia and nonverbal reasoning hypothesis) is supported.

A linear correlation was performed between all the DST-S battery subtests across the entire sample (both typical and dyslexia groups) and FOT scores. The results of this analysis can be seen in the Table 8 below.

Table 8. Linear correlation between DST-S battery subtests.

Variables	1	2	3	4	5	6	7	8
1.Rapid naming	-							
2.One minute reading	-.034							
3.Postural stability	.195	-.130						
4.Phonemic segmentation	.266	.284*	-.020					
5.Backwards span	.258	.161	.070	.257				
6.Nonsense passage	.422*	.414*	.036	.459	.373*			
7.Nonverbal reasoning	-.066	.431*	-.021	-.002	-.036	-.049		
8.One minute writing	-.080	.649*	-.342*	-.004	.112	.197	.392*	
9.Formal Operations	-.019	-.057	-.008	.054	-.063	-.163	.187	-.291

(N=50, *p<.05)

The reading ability tests of one minute reading and nonsense passage reading unsurprisingly correlated with the most other tests. This was logically expected due to the many cognitive abilities needed that coordinate to produce successful reading. In general, the verbal measures correlate with each other (e.g. nonsense passage, rapid naming and one minute reading) and the number and motor tests were independent of the others. This linear correlation suggests that the DST-S tests were assessing the abilities that they were expected from their descriptions.

Formal Operational Thought

Formal operational thought performance

After omission of incomplete questionnaires N=43 (typical sample=23; 53.5% and dyslexia sample=20; 46.5%).

The mean age across both samples was 21.43 years old (sd=3.79) with 36 females and 7 males (83.7% and 16.3% respectively).

FOT abilities were assessed through the worksheet made up of scientific thought based problems and abstract thought problems. The entire paper had a possible 19 marks for achievement. The mean scores for the paper and for the length of time taken to complete the paper for each group are shown below in Table 9.

Table 9. Mean formal operational thought paper scores and time for completion according to group.

	Typical Sample (N=23)	Dyslexia Sample (N=20)
Mean score	10.52 (sd 2.44)	11.60 (sd 3.28)
Mean time taken (mins)	17.52 (sd 4.74)	21.26 (sd 5.91)

From these results it appears that the dyslexia sample took longer to complete the paper but scored slightly higher on the paper than the typical sample. It appears that the dyslexia group took longer to process the instructions and scenarios and/or work out the answers. Both groups overall mean scores for the paper were quite low considering that the paper had a total of 19 potential marks. There were no significant differences between the groups' performance on the FOT test. This gives support for Hypothesis 3 (dyslexia and formal operational thought hypothesis).

A linear correlation showed that there were no significant correlations between any of the DST-S battery tests and the total marks for the FOT paper. This can be seen above in Table 9. This suggested that FOT proficiency did not affect performance upon any of the DST-S tests. The one minute writing test was the closest to being significantly associated with the formal operational thought scores ($r=-.057$, $p=.06$) and in fact this is a negative correlation. Hypothesis 4 (the formal operational thought and lower cognition hypothesis) is supported.

Detailed scoring of formal operational thought

Following an overall mark for the paper (as shown above) each of the questions were scored as assessing different levels of higher thought according to their classifications from Adey, Shayer and Yates (2001). The questions fell into the following categories:

- (1) Early concrete operational thought (assessed by two questions focused upon volume and heaviness principles).
- (2) Mature concrete operational thought (assessed by two questions focused upon more complex volume and heaviness principles).
- (3) Mature concrete/early formal operational thought (assessed by two questions focusing upon brainteaser science questions focusing upon volume and mass principles).

- (4) Early formal operational thought (assessed by one question focusing upon volume, mass and density principles, and two further questions focusing upon equilibrium and balance principles).
- (5) Formal operational thought (assessed by six abstract thought problem questions including the three new developed questions noted in the method).
- (6) Formal generalisation thought ability (assessed by three questions focusing upon equilibrium, balance, weight and proportion principles).

Individual scores for each question were recorded and from these scores each of the six thought level category scores were calculated. The means of these category scores can be seen in Table 10 below.

Table 10. Mean performance for Piagetian stages of thought (raw scores and % of sample showing evidence of thought level).

	Typical Sample Score (N=23)	Typical %	Dyslexia Sample Score (N=20)	Dyslexia %
Early concrete (/2)	1.77 (sd 0.66)	88	1.39 (sd 0.94)	69.5
Mature concrete (/2)	1.15 (sd 1.00)	57.5	1.44 (sd 0.91)	72.0
Mature concrete/ early formal (/2)	1.55 (sd 0.84)	77.5	1.57 (sd 0.84)	78.5
Early formal (/3)	1.00 (sd 1.31)	33.3	1.58 (sd 1.47)	52.7
Formal operations (/7)	4.04 (sd 3.48)	57.7	4.09 (sd 3.27)	58.4
Formal generalisation (/3)	1.15 (sd 1.19)	38.3	0.96 (sd 0.81)	32

On closer inspection of the breakdown of the FOT scale, the methodology does appear to establish variance in performance as there is no evidence of either floor or ceiling effects (i.e. all the scores are neither all high nor all low) and there appears to be higher percentages for

the lower levels of formal operational thought (i.e. the concrete levels of thought) than the higher levels (formal and formal generalisation) which seems logically correct from the nature of one level building upon a lower level within acquisition of stages of thought.

Furthermore, the dyslexic group appear to perform better than the control sample at the higher stages of thought (early formal operations and formal operations). For both of these stages, the dyslexic students out-performed the typical students on these items. This difference was statistically significant for the early formal operations stage of thought ($t(43) = 2.54, p < .03$). None of the other contrasts were significant between the groups, except that there was a trend toward better performance of the controls on the early concrete tasks ($t(43) = 1.85, p = .07$).

Discussion

Hypotheses Summary

The typical sample performed better than the dyslexia sample on the rapid naming, one minute reading, phonemic segmentation, nonsense passage and one minute writing tests. There were no significant differences between the samples for the non-verbal reasoning, backwards span or postural stability tests. Hypothesis 1 (the dyslexia and DST-S hypothesis) is therefore partly supported; the only tests that were expected to be related to dyslexia group membership that did not show significantly better performance for the typical sample were postural stability and backwards digit span. The fact that there was no significant difference between the groups on nonverbal reasoning scores suggests that Hypothesis 2 (the dyslexia and nonverbal reasoning hypothesis) is supported.

Both samples performed similarly on the FOT test with the mean values for each group reported as approximately 50-60% correct only. The dyslexia sample had a slightly higher mean but they took, on average, four minutes longer to complete the paper and also showed more variance around the mean. This is likely a reflection of difficulties the dyslexia sample may have had with reading speed and processing ability, rather than a reflection of FOT ability. There were no significant differences between total scores of groups. This supports Hypothesis 3 (dyslexia and formal operational thought).

There were also no significant associations between FOT scores and any of the DST-S tests (i.e. those assessing the lower cognition abilities noted in the measures section or non-verbal

reasoning which is held to assess general intelligence). This supports Hypothesis 4 (formal operational thought and lower cognition).

The dyslexia sample performed statistically significantly better than the typical sample on the latter stages of formal operational thought (i.e. the higher stages) whereas the typical sample was stronger for the lower stages of thought.

When taking the sample as a whole, there were no significant relationships between any of the dyslexia battery test scores and the FOT scores. When looking at the samples separately there appeared to be a trend of better performance from the controls on early levels of FOT and worse performance for the dyslexia sample, whilst the dyslexia sample performed better on the later levels of FOT and the typical sample performed worse. The variance of participants across this measure and the notable distinctions between performances across the different levels within FOT gives further support for the measure itself.

All hypotheses were mostly or entirely supported by the data suggesting that the measure behaved as expected. The measure allows for certain inferences to now be drawn about formal operational thought based upon this data. These are discussed further below.

Formal Operational Thought

The lack of significant correlations between the DST-S tests and the FOT scores suggests that FOT is an independent cognitive entity that is not impaired as a result of dyslexia, nor involved within the disorder, nor in the successful completion of any of the tasks assessed here. The fact that those with dyslexia showed higher levels of the latter stages of FOT suggests that the basic cognitions impaired with dyslexia, are not involved in FOT and that even when the primary cognitive abilities (such as reading) are diminished through dyslexia, FOT is not. In fact, the results suggest that FOT may be more proficient within the dyslexia participants than the typical ones for the later stages of thought. Furthermore, the closest significant association between any of the DST-S tests and FOT scores (one minute writing) was a negative one suggesting that better performance upon the DST-S test is associated with lower FOT proficiency. This is in line with the trends for the dyslexia group to outperform the typical in FOT stages; FOT appears unrelated to primary cognitive ability and may in fact inhibit performance upon lower cognitive tasks.

It is therefore suggested that as expected this indicates FOT is an independent cognitive entity that is used for and employs separate and higher level thought than that used or employed by

basic and lower level cognitions. The DST-S tasks themselves target lower basic cognition abilities, and the non-verbal reasoning task is described as an assessment of general intelligence. In fact, the FOT proficiency (i.e. ability to reason abstractly and use higher level thought skills) appears not to be related to performance upon tasks for basic reasoning and cognitive tasks such as reading, writing and basic manipulation of information.

The results of this study reflect the ideas in existing intelligence literature that there is an abstract area of intelligence (i.e. fluid intelligence) and that Piagetian reasoning is an aspect of this (Carroll, 1993), as opposed to other basic intelligence abilities such as decision speed or reading ability. It is seen in this study to be independent from the measure of a lower general intelligence from through the non-verbal reasoning measure. Furthermore, the literature has distinguished between this abstract intelligence, alongside crystallised intelligence, from the other second order constructs in intelligence modelling. These other constructs include visual perception, memory, knowledge and achievement and perceptual speed. From this study it is logical to suggest that these non-abstract elements of intelligence are some of the primary cognitions used for cognitive tasks (e.g. the DST-S tests used here or typical cognitive ability tests).

Furthermore, the consideration of tests that do not typically assess cognitive ability or primary intelligence aspects as discussed here, leads to the broader issue of the SJT, and what is measured. The need for further understanding, theoretical clarity and explanation of the SJT, and in particular, the constructs affecting performance upon this measure, has been noted throughout this thesis.

FOT could be a possible skill that is used or employed to achieve success upon the SJT; the abstract and higher level thought gained through this ability may be a key aspect of how one organises one's thoughts, plans and consideration of options on the SJT. It may allow for abstract reasoning and complex thought that goes far beyond that assessed by typical cognitive subjects such as mathematics, English or even more specialist subjects such as medical knowledge. In particular, this may be a useful ability for practitioners within medicine due to the ever changing, unique and cognitively sophisticated nature of this occupation. The formal operational thought construct has scope to be involved in research surrounding the SJT and may potentially be able to explain additional variance beyond currently investigated constructs such as cognitive ability and personality measures (see study

1) in SJT scores. Consequently, even job performance itself may be able to be further explained through the measurement and assessment of FOT through the SJT.

Following investigation, there may then be potential to incorporate intelligence models, FOT and SJT models into one theoretical base for proposing construct and criterion validity relationships within and around the SJT from a cognitive psychology stance.

The Formal Operational Thought Measure

Due to the individual questions being used extensively by Adey, Shayer and Yates and the data noted above in this study, the scale was deemed valid; the measure assessed what it was expected to assess giving evidence for its' construct validity.

The measure appears to have established a clear spread of variance of scores across the samples, and did not show any floor or ceiling effects (i.e. the scores were neither all very good nor very poor). Furthermore, the measure appears to produce results in line with current literature, in that these students performed relatively poorly on the measure, as suggested by FOT literature (Piaget stated that not all will reach this stage of thought) and more recent papers such as Shayer & Ginsburg (2009).

It should also be noted that where a typical measure of intelligence and ability (non-verbal reasoning) failed to find any significant differences between the two groups, the FOT test did. When taken as an overall score, the Piagetian test failed to show any group differences, but when the stages of thought themselves are examined, then there is a significant dissociation. Hence, the test itself may be a useful method for distinguishing underlying individual differences in a population.

Dyslexia

Whilst dyslexia as a disorder is not primarily relevant to this thesis, it is necessary to discuss the findings here briefly in relation to the results for the completeness of the discussion and study itself.

With regards to the investigation of dyslexia as a disorder, there was an interesting dissociation between the samples identified when examining the different levels of higher thought. Taking the results of the control group first, these are perhaps as expected. Even though the overall level of performance was low, with only 33.3% achieving early formal operations, this is in line with the disconcerting figures found by Shayer and Ginsburg (2009) for 16 year old children. Their strong performance on concrete operations but failure to move

on is also consistent with pattern observed by Shayer and his colleagues. Given that the Psychology students will in general have gained the top scores in their A-level subjects, this confirms the need for concern raised by Shayer. The results for the dyslexic group are particularly interesting; they performed significantly better than the controls on the early formal operations and they also performed worse than the controls on the early concrete operations test. Given that it is likely that the early concrete stage is achieved in late primary school, there appears to be a situation where the dyslexic children continue to develop cognitively in secondary school, whereas the controls have quickly reached the lower plateau but then not moved on.

There are potential practical implications for interventions and in particular teaching methods (for example, the extensive literature upon Cognitive Acceleration by Adey and Shayer, as noted in the literature). Shayer & Ginsburg (2009) even note the potential need to examine FOT as a further exploration of the Flynn effect to give further insight than standard psychometric tests of intelligence. The dissociation between samples may have also been due to the fact that those with dyslexia may struggle more with the rudimentary school teaching of the early stages of higher level reasoning, yet find strategies and alternative ways to deal with problems and can therefore tackle the later questions more easily.

Limitations

It is also necessary, however, to interpret the results with caution. Firstly, it was a relatively small study with small sample sizes within each group and consequently overall. Further larger replications of this research are needed to establish the generality of the results across populations.

Secondly, there were demographic differences between the groups. There were more males in the dyslexia group, and the average age of the dyslexic group was greater. Furthermore, the entire typical group consisted of student psychologists from within the University of Sheffield, whereas the dyslexic group came from a range of disciplines and occupations from outside and within the University.

Thirdly, the development of the formal operational thought scale has not conformed to the expected protocols on typical scale development. As discussed in the method section of study 2 due to the nature of the FOT concept and the cumulative scoring the traditional psychometric approach to scale development is not applicable. Best practice was noted with regards to Simms (2008) and how typical scale development may be conducted.

The validation process of the FOT measure in study 2 included the assessment of the basic construct validity of FOT through examination of discrepancies between results and the theory behind the scale. This was referenced in study 2 as the results were in line with what was expected in relation to the two samples differing cognitive profiles and similar FOT, and the higher levels of the lower stages of thought vs the higher stages of thought.

It would have been valuable to investigate the construct validity of the measure as a whole by correlating the measure with similar measures of formal operational thought that have been used in the literature. For example, the Group Assessment of Logical Thought, GALT; as used in Bitner (1991), and the Noelting (1980a) 'Orange Juice Test' as used in Liu & Shen (2011). Similarly to the Adey, Shayer & Yates measure, the Noelting (1980a) test assesses FOT hierarchically and the different levels of thought (e.g. concrete operations vs. formal operations) could have been correlated accordingly to assess construct validity of the new questions.

It would also have been valuable to take guidance from a recommendation of the American Psychological Association that new measures require a manual or specified instructions for the developed scale to describe the methods used to construct it, how to use it, how to score it and how to interpret it (APA, 1999). Conclusion

Methodology

The FOT test appears to be a robust measure from the results of this study. The measure showed variance across the sample and interesting variations of abilities across and within the samples for the different stages of thought. There were no ceiling or floor effects in the distribution of data. Due to the relatively low mean it appears that the test is challenging and therefore does require cognitive manipulation, reasoning and analysis. These results are also in line with the existing literature surrounding FOT assessment and what one would logically expect for a higher level thought reasoning test. However, the measure should be used with different populations and in studies with larger sample sizes. The generalisability of the results and use of the measure needs further investigation but this first step suggests promising results.

Theoretical conclusions

The results of this study indicate that if this measure has successfully assessed formal operational thought, as it is thought, FOT can be conceptualised based on this data as a separate cognitive entity to the lower cognitions used to tackle basic cognitive tasks, such as

reading and writing. FOT should be conceptualised as an ability for higher level reasoning on tasks that are cognitively more challenging than the basic ones used in the DST-S for this study. This is in line with the current literature which proposes FOT to exist within ‘fluid intelligence’ which itself is an abstract aspect of intelligence (Carroll, 1993). The lack of relationships between FOT and any of the DST-S tests (including the intelligence assessment of non-verbal reasoning) supports the idea of higher level thought entity that is independent to general intelligence abilities.

This study has established a measure for assessing FOT proficiency and suggests that future research uses this measure to further assess FOT performance, and in particular in relation to performance on tests that do not assess basic cognition/typical cognitive ability, for example, the SJT. One could hypothesise that for assessments of non-typical intelligence aspects, FOT may show significant associations or relationships with the results.

This study aimed to establish a measure of FOT and to further conceptualise it from empirical results so that it can be used alongside an SJT measure to investigate relationships in the final study. The third study of this thesis is further explained and introduced following this chapter.

CHAPTER 6: FORMAL OPERATIONAL THOUGHT AND THE SJT

Study 3: Situational Judgment Test performance, Formal Operational Thought, University Attainment and Job Performance in Final Year Medical Students

Introduction

In study 2, the formal operational thought (FOT) tool was established as a valid measure that could establish variance across samples with both typical and atypical cognitive abilities. Furthermore, the FOT concept itself was confirmed as a separate cognitive entity above and beyond the lower entities of thought and other knowledge stores for everyday and primary cognitive tasks. Finally, FOT appeared to be a new dimension of thought that had not been investigated in relation to personnel selection methods, and in particular not in relation to an SJT.

The findings that FOT is likely an abstract base for reasoning in complex cognitive tasks makes this an interesting and important direction to investigate within research concerning a Situational Judgment Test (SJT) within medicine.

The SJT is a well-established measure and research tool that is widely accepted as a selection measure that does not target lower levels of cognition (for example, it does not assess the declarative knowledge used for facts and principles). Furthermore, as a tool there is still a large amount to be understood in terms of theory and specifically, within the areas of construct and criterion validity. FOT may be a construct of interest that could affect performance on the SJT adding incremental validity above and beyond that found in existing research through current variables of interest, e.g. cognitive ability and personality factors. Furthermore, since this construct (FOT) has not been investigated before in relation to the SJT, there is scope to also explore the possibility of it's relationship with job performance and how FOT can fit into these ideas surrounding the SJT.

This study tackles Research Question 3 (can the inclusion of an abstract intelligence variable assessing higher level thought further explain the construct validity of the SJT in a medical selection context over and above variables from existing SJT theory) and Research Question 4 (can the inclusion of an abstract intelligence variable assessing higher level thought further explain medical job performance independent of the SJT in a medical selection context?)

Theoretical Basis for Study

Following the establishment of the validity of the FOT measure and affirming the concept as an independent and higher order thought ability, for this thesis it is necessary to relate this to a high stakes medical selection setting.

From study 2, FOT was established as higher and independent ability, separate from basic level cognition. This higher level and abstract thought ability may in fact be a construct of intelligence assessed by the SJT in medical selection. It may therefore be able to give incremental validity above and beyond the variables described in the McDaniel et al., (2006) model, such as cognitive ability or personality factors when explaining variance in the SJT scores. Furthermore, this study investigated the criterion related validity of the SJT using practical examination performance as a measure of job behaviour alongside the relationship between FOT and job performance.

The aim of this study was to investigate the construct related validity of the SJT and the new construct of interest of FOT as well the criterion related validity of the SJT. Whilst the criterion related validity of the SJT and job performance link has already been established (e.g. Lievens & Patterson, 2012) to replicate this finding in this medical context whilst given the opportunity is useful to develop a larger evidence base. The new FOT construct from intelligence literature (from Piagetian reasoning; Carroll, 1993) was investigated alongside a cognitive ability measure and a job performance measure.

The participants were final year medical students who were sitting the Situational Judgment Test as actual applicants to the Foundation programme (two years post qualification as junior doctors). The ranking of applicants and selection for this programme and location of foundation sites was based upon SJT performance and Educational Performance Measures (EPM).

The variables in this study were: an Educational Performance Measure (EPM) as a measure of cognitive ability, an Observed Structured Clinical Exam (OSCE) as a measure of practical job performance, a Situational Judgment Test (SJT) and a short form version of the FOT (from the study 2 measure). These measures are further discussed below.

Cognitive ability

The EPM is a measure of academic knowledge and achievements from the participant's university career (typically a degree programme lasting for five years). For the purpose of

this study, the EPM is used as a measure of cognitive ability and therefore of declarative knowledge (i.e. facts, rules and principles related to medical and academic knowledge). It is acknowledged that this measure may be affected by numerous variables that influence academic attainment at university, such as motivation, life events and topic of assessment. However, the inclusion of this measure was important as cognitive ability is widely accepted and established as a measure that affects performance upon the SJT (e.g. McDaniel et al., 2006). Therefore the construct validity of the SJT in relation to this could be analysed.

Job performance

Job performance was assessed using a practical exam measure called the Observed Structured Clinical Examination (OSCE). An OSCE requires students to move through a rotation of stations that each ask for the application of clinical skills in a variety of contexts and situations. Furthermore, they involve patient contact and interaction (from actor simulation of a patient encounter/consultation). Therefore, the display of proper and professional conduct as a medical professional is required. Each station lasts approximately 5-10 minutes (Talbot, 2004) and can often be a very high pressure scenario for students.

There has been some suggestion that the performance upon the OSCE is not the same as real life (Ram, van der Vleuten, Rethans, Schouten, Hobma & Grol, 1999) due to a potential lack of validity due to the short station lengths and scoring against a check list (Reznick, Regehr, Yee, Rothman, Blackmore and Dauphinee, 1998). However, the reliability is noted as strong regardless of these points (Waas & van der Vleuten, 2009) and as it is a close approximation of real job performance, and furthermore the only one available due to resources and timing, it will be used as an actual job performance measure for this analysis. It is very difficult to gain actual job performance measures, as noted earlier in chapter 2, (Waas & van der Vleuten, 2009) and hence, this measure is in fact a good substitute.

Situational Judgment Test (SJT)

The students who nationally completed the SJT took one of three papers and this was done as a national scheme for selection/assessment purposes for the gateway to Foundation year training (the two years post qualification as noted). The papers for this year in question, consisted of 70 questions and asked students to say what they *should* do (knowledge tendency instructions) as opposed to what they *would* do (behavioural tendency instructions) in response to a scenario description.

Within this study the SJT was used to investigate firstly, the construct validity of the measure (investigating cognitive ability and formal operational thought as constructs assessed by the SJT) and secondly the criterion validity of the SJT (investigating actual job performance in relation to SJT scores). The study allowed for the investigation of current theory and to develop new modelling integrating existing ideas regarding SJTs and intelligence, along with formal operational thought conceptualisations as an aspect of cognitive intelligence literature stemming from developmental psychology theory.

Formal operational thought (FOT)

A short version of the FOT test used in study two was developed for this study. This was due to time constraints on the length of the session with the medical students and thus it was not possible to administer the complete test to students. Therefore, six questions were chosen from the original measure and these were chosen to ensure each of the hierarchical 'levels of thought' were assessed in the test (as noted by Adey, Shayer & Yates, 2001). To summarise, this short measure was designed and used to allow for the short time period available and to ensure that all of the main stages of higher level thought were measured. The use of this measure allows for the examination of a possible new construct affecting performance on the SJT. Furthermore, it can be investigated to allow for clear conceptualisation of the ability, in relation to other knowledge stores assessed, and actual medical job performance.

Results from this study were used in conjunction with the other two studies to develop a theoretical framework that incorporates current literature on the SJT and intelligence testing. Future directions for research are also considered.

Study Design

The study was a cross sectional within-participants design that involved one data collection day and the obtainment of cognitive ability test performance and practical exam performance from the University of Sheffield.

The students' academic attainments were obtained (i.e. their examination and assessment performance over their entire university courses and how the students were consequently ranked according to this information). Those academic scores were used to rank the students and this is called the Educational Performance Measure (EPM). This is a measure developed in 2011 by the ISFP (Improving Selection to Foundation Programmes) as a composite measure of academic performance. Following piloting, consultation and final alterations, this

measure was agreed on as a measure of academic performance by students, employers and all medical schools (ISFP Final Report by the Medical Schools Council, 2011).

The results from the students' final year practical exams (the Observed Structured Clinical Examinations or OSCEs) were also supplied. This measure (the OSCE) is a set of medical stations that students must each visit once whilst being observed for their clinical skills by expert assessors. The students are faced with hypothetical patients in different scenarios and are asked to perform certain procedures and take certain actions at each station. It is a practical examination of on the job ability.

Alongside the EPM and OSCE results, the results from the nationally administered Foundation training selection SJT were obtained. The students were assessed using this SJT in their final year of university as a selection method for their first Foundation level job roles following graduation from medical school.

The study therefore allowed further investigation into FOT as an intelligence concept as well as investigation of the potential relationships between academic performance (measuring cognitive ability), SJT performance and practical exam performance (measuring on the job performance). In particular, the construct validity of the SJT, using a new variable of FOT, and a well-established and already much investigated variable of cognitive ability was of interest. As a new construct to be investigated in relation to the SJT, it was therefore sensible to investigate both FOT's potential involvement within the construct validity of the SJT, and its potential relationship or association with job performance in this study for completeness.

Hypotheses

Hypothesis 1: Cognitive ability and the SJT

The cognitive ability measure in this study is a composite measure of academic achievements and other educational qualifications (e.g. a master's degree). This score is known as an Educational Performance Measure (EPM). I further discuss this measure and the use of it as an assessment of cognitive ability in the method section of this chapter. Cognitive ability is widely researched in relation to the SJT, and consequently is established as a construct that affects performance upon the SJT and is a construct measured by the SJT (e.g. McDaniel et al., 2006). It is therefore expected that the EPM scores will be directly related to the performance scores upon the SJT.

Hypothesis 1 states that cognitive ability will be positively associated with SJT scores.

Hypothesis 2: Cognitive ability and formal operational thought

The EPM can be referred to as a measure of declarative knowledge, i.e. knowledge of facts, rules and principles. In Carroll's (1993) model of intelligence the 'knowledge and achievement' aspect will encompass the cognitive ability described here. This is a separate conceptual area of intelligence to the abstract intelligence 'fluid intelligence' containing FOT, that is then adapted into the modern CHC theory of intelligence.

It was therefore expected that these two forms of intelligence would be weakly related to each, as correlations between these second-order constructs of intelligence have been identified (Carroll 1993; Ackerman & Heggstad, 1997) and typically there are correlations found across intelligence abilities (e.g. as noted in the literature, this may be classed as an artefact of research that may be mistaken for 'g'). It has also been established in cognitive psychology and the medical selection literature that procedural knowledge is built upon a basis of declarative knowledge (e.g. as noted in Lievens and Patterson, 2011). This further suggests that they should be positively related to at least a weak/moderate extent.

Hypothesis 2 therefore states that cognitive ability and formal operational thought will be positively associated.

Hypothesis 3: Formal operational thought and SJT construct validity

Following study 1, FOT has been investigated as a construct that may be able to further explain variance in SJT scores; the advised direction of focus from past literature and within this thesis is to further investigate the constructs of interest to the SJT and within the intelligence literature, an abstract form of knowledge (FOT) has previously not been investigated and now has been targeted. This higher level thought and reasoning ability may be an aspect of intelligence that the SJT is assessing and the nature of this FOT knowledge is different to that used in typical cognitive ability stores (see chapter 5 for further explanation of this potential link).

Hypothesis 3 therefore states that the abstract knowledge of formal operational thought will further explain variance in SJT scores over the construct of cognitive ability.

Hypothesis 4: SJT criterion validity

The practical job measure of the OSCE can be used in this study as a practical job performance measure. This can then be used to investigate the criterion validity of the SJT; how well does the SJT predict job performance? Meta-analyses present approximate

explanations of 9% variance in job performance (McDaniel et al., 2006). The SJT has been widely established as predicting job performance (for example, as seen in Lievens & Patterson, 2011).

Hypothesis 4 states that the SJT scores will significantly explain variance in actual job performance.

Hypothesis 5: Formal operational thought and job performance

The SJT scores and FOT scores will be used in a hierarchical regression model to establish if the variables can significantly explain variance in actual job performance scores. SJT scores have been noted as significantly explaining variance in job performance (e.g. Lievens and Patterson, 2011; noted above in Hypothesis 4) and hence, hypothesised to recreate this. It is also hypothesised that FOT will add incremental validity over this existing measure. This measure will give further depth and insight into a participant that will then further explain actual job performance. The FOT aspect of intelligence will capture the aspects of intelligence used in job performance that are not assessed by the SJT. For example, this may include the more creative and highly abstract areas of intelligence that are not used for SJT performance, but may be relevant in the on the job behaviour. As this measure has not been investigated in relation to medical job performance or an SJT before, it is necessary and logical to investigate the relationship to both medical job performance and the SJT. It is also suggested that as FOT is an aspect of intelligence from the CHC model, it will show a relationship with job performance; cognitive ability has been noted from empirical studies as by far the best predictor variable for the criterion of job performance (Ones, Dilchert & Viswesvaran, 2012).

It is of interest to investigate FOT as a small aspect of intelligence/cognitive ability and its relationships with job performance. The abilities of higher level thought associated with FOT proficiency are likely linked to levels of general mental ability (i.e. these skills aid higher intelligence, reasoning capabilities and learning abilities). These abilities in theoretical terms are directly linked to job knowledge acquisition of both declarative and procedural knowledge, which produces the empirically shown relationship of cognitive ability predicting job performance (Borman et al., 1993; Borman, White, Pulakos & Oppler, 1991; Ones, Dilchert & Viswesvaran, 2012). It is possible that the elements of FOT and the SJT together will explain job performance more extensively than the SJT alone.

Hypothesis 5 states that formal operational thought will add incremental validity over the SJT scores in explaining actual job performance.

Hypothesis 6: Formal operational thought, cognitive ability and job performance

Patterson et al., (2012) also note that not using cognitive ability in later medical selection following university entry may be contrasting due to high levels of intelligence required for course entry in the first place. Furthermore, cognitive ability has been extensively established as a predictor for job performance, success and effectiveness over any other individual differences predictor (Ones, Dilchert & Viswesvaran, 2012; Sackett, Schmitt, Ellington & Ones, 2004; Schmidt & Hunter, 1998). For completeness of this analysis section, the cognitive ability measure will be used in a regression model to establish if the SJT and the FOT scores can add incremental validity to cognitive ability when explaining variance in job performance. Lievens & Patterson (2011) found the SJT to add incremental validity over knowledge tests when predicting job performance.

Hypothesis 6 states that the SJT and FOT will have incremental validity over cognitive ability in explaining actual job performance.

Method

Study Design

A cross sectional within-participants design with analyses that looked at relationships between FOT proficiency, cognitive ability, actual job performance and SJT scores.

Participants

A sample of 240 final year medical students from the University of Sheffield Bachelor of Medicine Undergraduate Degree course.

The sample was of mixed gender with 53.4% of the sample female (n=132) and 43.7% male (n=108). The age range of the sample was from 22 years old to 44 years old (mean age=24.73, sd=2.56).

The sample was predominantly of white British ethnicity (n=156) with the next largest ethnicity frequency being Asian or Asian British- Indian (n=19).

Ethical approval was obtained from the University of Sheffield's Management School's Ethics committee. The participants themselves were identified through Professor Nigel Bax (the Director of the Medical School at the time in the University of Sheffield) and

subsequently recruited with a letter explaining the study aims, structure and how the data would be used (see Appendix XI). Full informed consent was therefore obtained and the participants were fully aware of the ethical issues involved. By signing the consent form each participant authorised access to their academic history throughout their university careers and as well as their SJT results, the test results from the FOT measure on the data collection day.

Materials

Due to a large amount of the data being readily available from other sources (SJT scores from the Work Psychology Group and the academic history and job performance data from the University of Sheffield) the only data collection necessary was for the FOT test.

The study took place in the Medical School (in the University of Sheffield) with questions displayed via a PowerPoint presentation. It was planned that the participant's response answers would be keyed into personal keypads that would allow for instant viewing of the distribution of data spread, and for each participant to be identified individually and matched to their records via unique identification numbers. However, due to unforeseen technical difficulties on the testing day this was not possible. The questions were still displayed over PowerPoint with one question and all the possible answers per slide. However, students unfortunately could not use the personal keypads and were each asked instead to use a pen and paper to write the answers on the reverse side of their consent forms. No other specialist equipment was required.

Procedures

Academic attainment and performance

The academic ranking of students, along with their demographic information (sex, ethnicity and age) was supplied by the University of Sheffield's Medical School in the form of an Excel spread sheet. This spread sheet included the EPM score and ranking against the rest of the academic year from this score.

Cognitive ability

The EPM was used as a measure of cognitive ability. The EPM is a measure of clinical and non-clinical skills, knowledge and performance up to the point of application to the Foundation Programme (which commences after completion of the five year University course). The measure is made up from three parts assessing the medical school performance, additional degrees and other educational achievements. According to their level of medical

school performance they are assigned to a decile (1-10) across the university year and these qualify into a number of relative points (34-43). Applicants can then earn another 5 points for any additional degrees and/or another 2 points for other achievements such as publications or prizes (maximum score 50).

Whilst this measure is a composite assessment of academic attainment, it is likely that the results used to make up the measure are affected themselves by other variables. For example, personality factors, how hard one worked for certain examinations or other extraneous variables such as family issues or relationship breakdowns at the time of assessment. Nonetheless, the measure was agreed following extensive piloting, consultation and discussion by students, employers and medical schools as a suitable measure of academic attainment (ISFP Report, 2011) and, as the best measure available at the time of this research, will be used within this study as a measure of cognitive ability. The positive association of cognitive ability and exam performance has been well established in previous research (e.g. Kuncel & Hezlett, 2007; Kuncel, Hezlett & Ones, 2004 and Reeve, 2004, as cited in Reeve, Bonaccio & Winford, 2014). The fact that the data for the measure was obtained over a four year period and from numerous different examinations/measures is favourable for outlying occasions of bad performance to be averaged out to some extent.

It is in fact uncommon to use cognitive ability tests in advanced high-level stakes selection (AERA, APA & NCME, 1999; Raymond, Neutzel & Anderson, 2007, as cited in Lievens & Patterson, 2011) and knowledge tests of declarative facts, such as academic examinations, are used as a proxy for cognitive ability. Furthermore, measures of cognitive ability have been found consistently to strongly relate to success, performance and effectiveness in school and at work in general (Jensen, 1980, 1988; Kuncel, Hezlett & Ones, 2004; Sackett, Schmitt, Ellington & Kabin, 2001; Schmidt & Hunter, 1998, as cited in Murphy, 2012). This further supports the idea that an EPM measure, which may be interpreted as a measure of attainment or school/work success, would be related to cognitive ability based on the literature.

Job performance

The breakdown of the students' performance upon the practical exam (the OSCE) was also supplied in the form of an Excel spread sheet by the University of Sheffield's Medical School. Their total marks on the OSCE and grade of performance were disclosed and this was used as measure of job performance.

The OSCE requires students to move through a rotation of stations that each ask for the application of clinical skills in a variety of contexts and situations. Furthermore, they involve patient contact and interaction (from actor simulation of a patient encounter) and therefore the display of proper and professional conduct as a qualified doctor is required. Each station lasts approximately 5-10 minutes (Talbot, 2004).

There has been some suggestion that the performance on the OSCE is not the same as in real life (Ram, van der Vleuten, Rethans, Schouten, Hobma & Grol, 1999) due to a potential lack of validity as a result of the short station lengths and the method of scoring against a check list (Reznick, Regehr, Yee, Rothman, Blackmore and Dauphinee, 1998). However, the reliability is noted as strong regardless of these points (Waas & van der Vleuten 2009) and, as it is a close approximation of real job performance, and furthermore the only one available due to resources and timing, it was used as an actual job performance measure for this analysis.

The OSCE measure may also be viewed as one of training performance. From this viewpoint, this thesis is in line with previous literature (Lievens & Patterson, 2011) that uses supervisor ratings of training performance as a criterion measure of job performance. The rationale for this is rooted in the fact that training performance involves engagement with medicine and patients, as is the OSCE measure which similarly is assessed by professionals making judgments about correct clinical practice and skills.

Situational Judgment Test

The SJT results were also supplied independently through the Medical Schools Council and the Work Psychology Group team in the form of an Excel spread sheet and an SPSS data set. Due to the confidential nature of this data set, analyses only took place on site at the Work Psychology Group offices in Derbyshire where the data was stored.

Those students who nationally completed the SJT took one of three papers. The papers for this academic year consisted of 70 questions (10 pilot and non-operational questions, 40 questions that had a ranking answer format and 20 questions that had a multiple choice answer format). The instructions were knowledge tendency instructions that asked what an individual *should* do in a certain scenario, as opposed to what an individual *would* do (behavioural tendency instructions). The SJT assessed in this study was designed for final year medical students to undertake and is used at this gateway to rank students. Students had previously ranked locations in order of preference for their Foundation year job roles (the

first two years training after degree qualification in medicine) and the allocation of these job roles and locations was based equally upon SJT performance (50%) and the ranking of students from their personal EPM scores at this point (50%).

It was a nationally administered set of SJTs and therefore questions were well structured, researched and unambiguous. Responses are recorded in a paper and pen format and machine marked following completion. This is the same structure of SJT and the same gateway within medical selection where the SJT was developed for in study 1 (although the participants in study 1 only ‘assumed’ this role).

Marking took place mechanically and results were then equated according to the operational items and which paper an individual took out of the three options (this is due to different items within different papers overlapping and also different numbers of questions being based upon different domains across the papers). This equalisation was completed before the data set was disclosed for analysis.

Formal operational thought

The measure

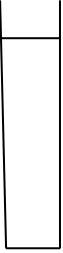
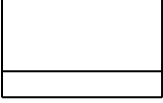
A short form version of the FOT measure used in study 2 was developed for this final study. Firstly, this was due to time constraints on the length of the session with the medical student sample it was not possible to administer the whole test to the students; time constraints indicated a maximum of 6 questions. Secondly, six questions were chosen from the original measure and these were chosen to measure each of the hierarchical ‘levels of thought’ measured in the test in order to assess the different levels of Piagetian attainment (as noted by Adey, Shayer & Yates, 2001). This measure was therefore designed and used to allow for the short time period available and to ensure that all the main stages of higher level thought were measured to the best extent possible. The measure was scored out of 6 marks, and for the final question where there are two parts to the answer, both parts had to be correct to score one mark. The six questions used can be seen below in Table 11 with their ‘level of thought’ measured listed.

The development of this short form scale by choosing items to reflect the hierarchy of FOT levels was similar to the establishment of the short form version of the Medical Outcomes Scales (MOS), which was also derived from a longer scale. The MOS 36-item Short Form is a mental health subscale that contains only 6 items, with 1 item in each of several domains,

such as physical functioning, social functioning, mental health, and global health (McHorney, Ware & Raczek, 1993; as cited in Moore, Halle, Vandivere & Mariner, 2002).

Moore, Halle, Vandivere & Mariner (2002) also note that optimally, when shortening a scale, previously collected data should be available to allow items to be chosen using item analysis or predictive properties of the individual items to inform choice of the best items for the reduced scale. In terms of the statistical analysis and providing evidence for the FOT short form scale it would have been useful to have investigated associations of the measure; it would have been expected that positive correlational relationships could have been established between both the full and short versions had a piloting study with the inclusion of both measures been possible.

Table 11. Short-form version of formal operational thought test questions and level of thought assessed by each question.

Question	Level of thought assessed
<p>Examine the diagrams of the two beakers and tick the correct answer.</p> <p>A is a 100cm³ beaker and X is a 250cm³ beaker. Beaker A is filled with water that is poured into Beaker X. Beaker A is then refilled with water.</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Beaker A</p> </div> <div style="text-align: center;">  <p>Beaker X</p> </div> </div> <p>A has more</p> <p>less</p> <p>the same amount of water compared with X</p>	<p>Early concrete operations</p>
<p>The volume of a block of plasticine is 60cm³.</p> <p>How much water will spill over if this block is put fully underwater in a full beaker of water?</p> <p>a) 60cm³</p> <p>b) 120cm³</p> <p>c) 30cm³</p> <p>d) 0cm³</p>	<p>Mature concrete operations</p>

A brass block is the same size as the block of plasticine. The brass block is heavier than the block of plasticine.

Mature concrete operations/ early formal operations

If the brass block is lowered into the beaker of water will

More.....

Less.....

The same.....amount of water spill out of the beaker as when the plasticine block was lowered

You are given two blocks made of brass.

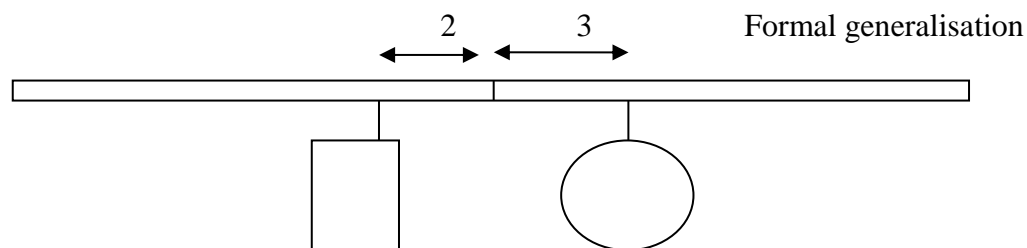
Early formal operations

Block A weighs 60 grams and has a volume of 15cm^3

Block B weighs 160 grams. What is its volume?

- a) 40cm^3
- b) 160cm^3
- c) 30cm^3
- d) 15cm^3

The two weights on the diagram balance.



Which weight is heavier?

How much heavier is it?.....

On a planet in outer space there are two

Brainteaser question (developed for this

species of creatures that have been identified, measure in study two).
these are species x and species y.

If there are no x's that aren't slim and
no y's that aren't x's, then which
statements are always true?

- (a) There is not one slim creature that
isn't an x
 - (b) All y's are slim
 - (c) Any creature that is slim is also a y
 - (d) None of the above
-

Data collection

The data collection involved participants answering the questions on FOT. The study was carried out within a lecture theatre in the University of Sheffield's Medical School.

Before the study itself, as an incentive to the students, a short talk about success performance on the SJT was given to the students. This included tactical advice and psychological points about dealing with the stress and pressure, as well as the psychological research behind the SJT. Participants sat together in the lecture theatre and after the introduction were asked to read the information letters and sign the consent form attached if they understood and were happy to take part in the research. Following this, due to the previously mentioned technical difficulties, participants were instructed to write the answers for the questions on the reverse of their consent form.

The FOT section contained six questions and each question was displayed on the public screen to the students with the possible multiple choice answers. Students then had approximately two minutes to write down their answer before the next question was displayed. After all the questions had been shown participants were asked to leave their forms and answers on their desks as they left the lecture theatre.

Summary of methods

For complete clarity, below is a table showing a breakdown of the measures used as well as a brief description of these methods and the time when the measurement was taken (Table 12).

Table 12. Summary of methods for study 3.

Method	Time of assessment	Brief description of method
Educational Performance Measure (EPM)	Across initial four years of medical degree at the University of Sheffield (2009-2012)	Medical school performance (clinical and non-clinical), other degrees and extra achievements
Observed Structured Clinical Examination (OSCE)	Final year exam (Early 2013)	Practical observed exam of clinical competence
Formal Operational Thought (FOT)	Data collection day (December 2012)	Measure of higher level thought ability through scientific and non-scientific reasoning problems
Situational Judgment Test (SJT)	Final year exam (Late December 2012)	Hypothetical medical scenario questions with ranking (part A) or multiple choice (part B) answer format (60 operational questions). A national test with 3 possible papers.

Results

Sample Size

As a consequence of the technical difficulties during testing, a large number of the participants failed to answer the majority or any of the FOT questions. This was likely due to disruption as a result of the technical problems and the relative inconvenience of having to write answers as opposed to entering them electronically on the personalised keypad as

previously planned. Therefore, a large proportion of the sample was eliminated from the analyses due to this and consequently, $N=153$. For consistency, these incomplete profiles were eliminated for all statistics, even if the analysis in question did not involve the formal operational thought measure specifically.

This final sample was similar in demographic make up to the entire cohort and mean age = 24.42 years old ($sd = 1.67$) with a minimum age of 22 and maximum age of 38 years old. The majority of the sample was female ($N=89, 58.2\%$) with less men reporting in the study ($N=64, 41.8\%$).

Cognitive Ability

Participants were scored with an Educational Performance Measure (EPM) which is composed of three parts assessing the medical school performance, additional degrees and other educational achievements of the students (further breakdown in the method section).

The maximum score possible was 50. The mean score of this sample was 40.51 ($sd=3.98$). This suggests a high range of EPM scores with little variance across the sample. The total EPM scores were the only scores supplied for this research (i.e. each part score was not disclosed) and consequently, internal reliability analyses of the measure as a whole was not conducted.

Job Performance

As part of the students' university career performance information students were examined through their final year Observed Structured Clinical Examination (OSCE) results. This is a measure of how they have performed in 'on the job' tasks whilst being observed. These results were released for analysis as a percentage mark and a grade as a result of the percentage.

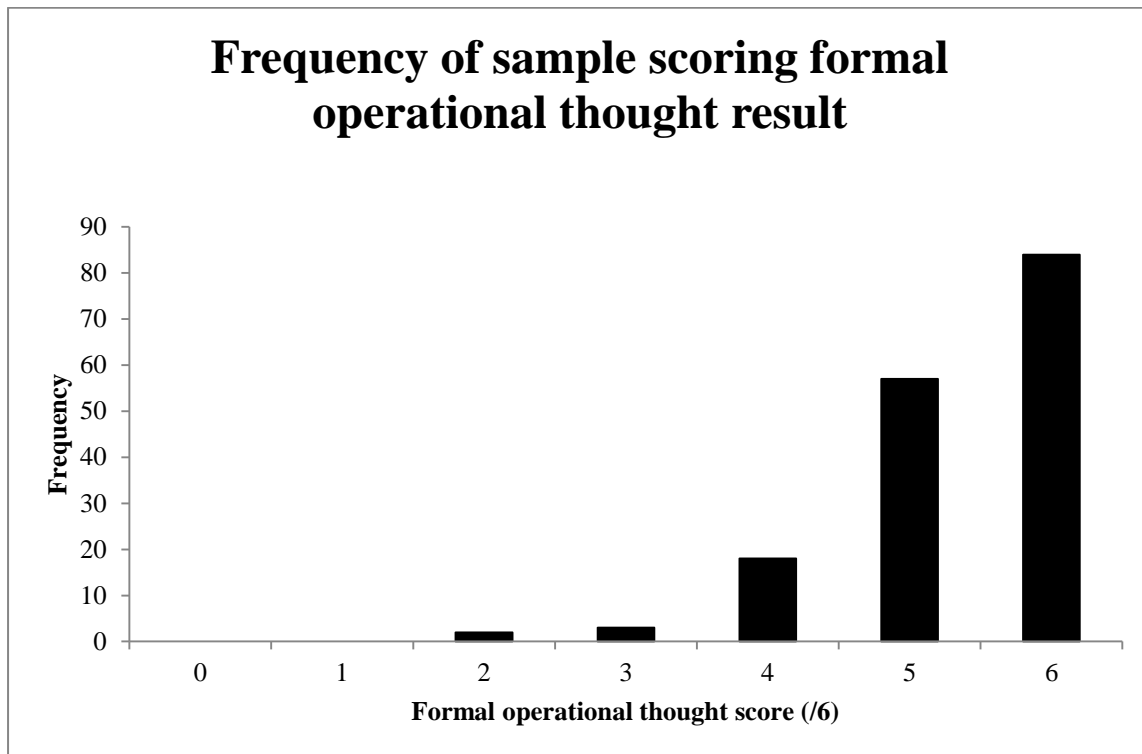
The mean score upon the OSCE was 79.84% ($sd =4.91$) and the mean grade was 3.91 ($sd=.62$). This suggests good performance across the sample upon the OSCE with little variance around the mean value.

The total OSCE scores were the only scores supplied for this research (i.e. individual scores for each station were not disclosed) and consequently, internal reliability analyses of the measure as a whole was not conducted.

Formal Operational Thought

The FOT test consisted of six questions with one mark available for each question. The distribution of the sample according to their score (out of 6) for the FOT questions can be seen below in Figure 8.

Figure 8. Graph to show distribution of sample according to formal operational thought proficiency.



(*N=153)

The maximum score achieved was 6 and the minimum score achieved was 2 (the maximum possible score was 6 and the minimum possible score was 0).

The mean score for the FOT test was 5.33 (sd= .85). Figure 8 suggests that most participants scored highly on the FOT questions. The distribution of scores is negatively skewed (-1.42, stderror= .20) and slightly leptokurtic (2.25, stderror= .39).

The 6 FOT questions as a whole could be further classified as different levels of higher reasoning according to the question's individual content. For example, a correct answer to question one would demonstrate a level of early concrete reasoning, but a correct answer to question two as well would demonstrate a level of mature concrete reasoning. As the questions increased in number they also increased in the level of thought that they could test.

This can be seen in the Table 13 below along with the proportion of participants who scored correctly on each question.

Table 13. Proportion of sample scoring correctly and higher level thought distinctions for each formal operational thought question.

Question number	Classification of higher level thought stage	Proportion of sample answering correctly (%)
1	Early concrete operational thought	100%
2	Mature concrete operational thought	100%
3	Mature concrete operational/ early formal operational thought	97%
4	Early formal operational thought	89%
5	Formal generalisation abilities	65.2%
6	Abstract reasoning capabilities/ hypothetical analysis and reasoning under uncertainty	51.2%

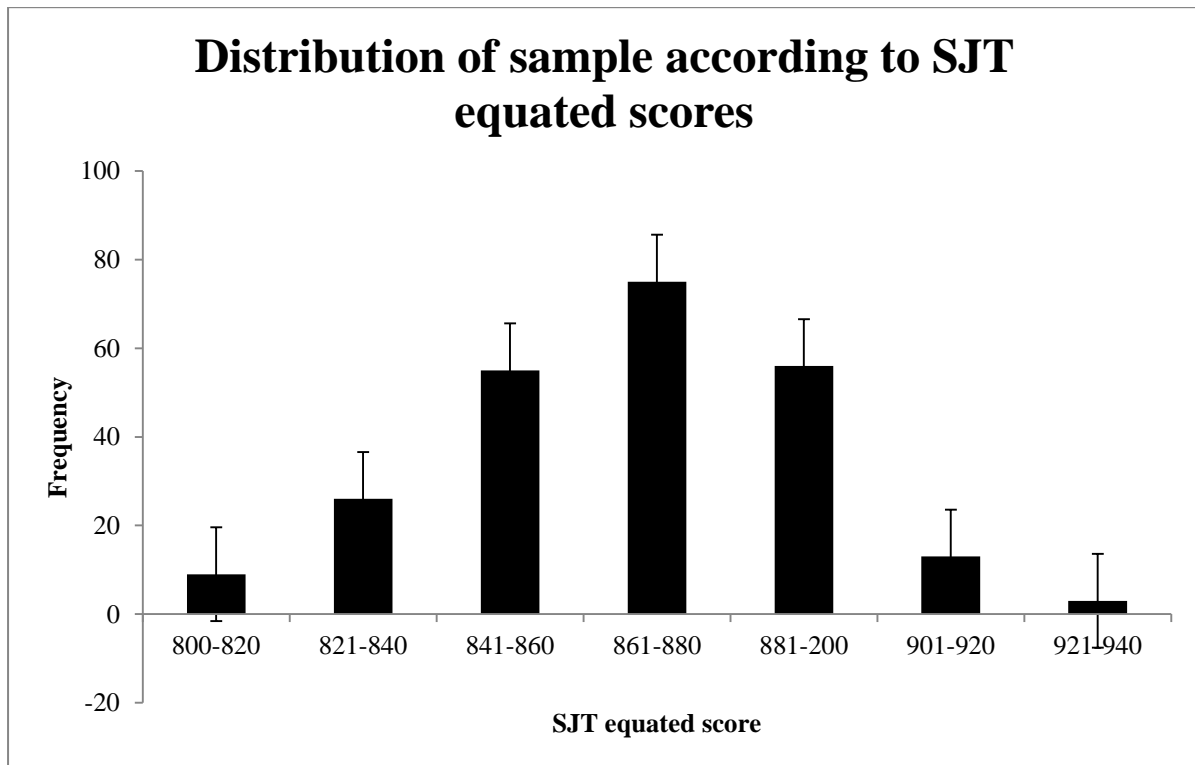
(*N=153)

As the negative skewness of the distribution demonstrates, all those who completed the FOT questions showed the early stages and mature concrete reasoning ability. There are still quite high proportions of the sample showing the later stages of reasoning (51.2%).

Situational Judgment Test General Performance

The distribution of the sample according to performance upon the SJT can be seen in Figure 9 below.

Figure 9. Graph to show distribution of sample according to Situational Judgment Test Performance.



(*N= 153. Error bars showing standard error).

The maximum score was 927 and the minimum score was 802. The mean score for the sample was 870.33 (sd= 22.96).

The graph suggests a spread of results very close to what is considered as a normal distribution of scores. The majority of the participants scored between 861 and 880. (The maximum possible score available after the scores are equated is 1037 points).

The SJT was made up of 70 questions (60 of which were operational and 10 of which were pilot questions). Furthermore, the SJT was split into two sections. Part A was a section consisting of 40 questions that asked the participant to rank multiple choice responses in order of preference, and Part B which asked the participant to choose one of the multiple choice responses as the answer.

The maximum score achieved within the sample for Part A of the SJT was 717 and the minimum score was 625. The mean score was 678.54 (sd=17.86). The maximum possible score for the ranking questions after being equated was 797.

The maximum score achieved within the sample for Part B of the SJT was 220 and the minimum score was 152. The mean score was 184.84 (sd= 11.04). The maximum possible score for the multiple choice questions after being equated was 240.

Linear Correlations

A linear correlation was performed with the FOT test scores, the cognitive ability scores, the job performance scores and the total SJT scores. The results of this analysis can be seen in Table 14 below.

Table 14. Linear correlation of Situational Judgment Test scores (SJT), cognitive ability, formal operational thought (FOT) and job performance.

	Cognitive Ability	FOT	Job Performance
SJT	.256**	.038	.268**
Cognitive Ability		.125	.537**
FOT			.196*

(N= 153, * p<.05 **p<.01)

The SJT was significantly correlated with cognitive ability. The FOT measure failed to correlate significantly with either the SJT or cognitive ability. These results suggest that Hypothesis 1 (cognitive ability and SJT performance) should be accepted and that Hypothesis 2 (cognitive ability and FOT) should be rejected. Cognitive ability is not significantly related to the FOT scores although the relationship was in the expected direction.

This also suggests that Hypothesis 3 (FOT and construct validity) should be rejected as FOT is not significantly related to SJT scores. This is further investigated using regression analyses below. In line with the Hypotheses 4,5 and 6 jobs performance was significantly correlated with the SJT scores, cognitive ability and FOT scores. Higher measures on all these variables are significantly related to higher levels of job performance. This is further investigated using regression analyses below.

Regression Analyses

Hierarchical regression analyses were conducted to investigate the three main hypotheses examining FOT and the relationships with the SJT and job performance.

Situational Judgment Test construct validity

A hierarchical multiple regression analysis was conducted with the SJT scores as the dependent variable and three steps.

The results of this analysis can be seen below in Table 15. Step 1 of the analysis entered the demographic control variables of age and gender which accounted for 3.6% of the variance in SJT scores and this model was non-significant.

The inclusion of cognitive ability in step 2 explained an additional 8.2% of the variance (Total $R^2=.12$, $p<.001$). Cognitive ability provided a unique effect ($B=1.69$, $p<.001$). This gives support for Hypothesis 1; cognitive ability scores significantly explained variance in SJT scores when demographic measures were controlled for.

Thirdly, the FOT scores were entered and this failed to account for any additional variance (Total $R^2=.12$, $p=.830$) and the model was non-significant. This suggests that Hypothesis 3 (formal operational thought and SJT construct validity) is not supported. FOT scores did not significantly predict variance in the SJT scores over the demographic measures and cognitive ability.

Table 15. Hierarchical multiple regression to investigate construct validity of the SJT.

	Step 1 B	Step 2 B	Step 3 B
Age	-2.31*	-3.02*	-3.03*
Gender	-3.83	-3.24	-3.23
Cognitive Ability		1.69*	1.68*
Formal operations			.456
Change R^2	.036	.082*	.000
R^2	.036	.118*	.118
$F(df)$	2.74 (2, 147)	6.49 (3, 146)	4,84 (4, 145)

(* $p<.05$)

Situational Judgment Test criterion validity, formal operational thought and job performance

A hierarchical multiple regression was conducted with job performance as the dependent variable and three steps.

The results of this analysis can be seen in Table 16 below. Step 1 of the analysis entered the demographic control variables of age and gender which accounted for 4.9% of the variance in job performance.

The inclusion of the SJT scores in Step 2 accounted for an additional 6.2% of variance (Total $R^2 = 0.11$, $p = .002$). The SJT had a unique effect ($B = .055$, $p = .002$). This suggests that Hypothesis 4 (SJT criterion validity) should be accepted. The SJT scores did significantly explain actual job performance scores.

The third step further included the FOT variable and this accounted for an additional 3% of variance in a significant model summary (Total $R^2 = .14$, $p = .027$). The SJT ($B = .052$, $p = .003$) and FOT scores ($\beta = .998$, $p = .027$) both had unique effects on the significant model.

This suggests that Hypothesis 5 (FOT and job performance) should be accepted and the null hypothesis rejected. FOT had incremental validity in explaining variance in actual job performance, above the SJT scores.

Table 16. Hierarchical multiple regression to investigate criterion validity of the SJT, formal operational thought and job performance.

	Step 1 B	Step 2 B	Step 3 B
Age	.16	-.04	-.07
Gender	-2.13*	-1.98*	-1.96*
SJT		.055*	.052*
Formal operations			.988*
Change R^2	.049*	.062*	.030*
R^2	.049	.111	.141
$F(df)$	3.68 (2, 143)	5.89 (3, 142)	5.79 (4, 141)

(* $p < .05$)

Situational Judgment Test criterion validity, cognitive ability, formal operational thought and job performance

A hierarchical multiple regression was conducted with job performance as the dependent variable and four steps.

The results of this analysis can be seen in Table 17 below. Step 1 of the analysis entered the demographic control variables of age and gender which accounted for 4.9% of the variance in job performance.

The inclusion of the cognitive ability scores in Step 2 accounted for an additional 28.8% of variance (Total $R^2 = .337$, $p < .001$). Cognitive ability had a unique effect ($B = .69$, $p < .001$).

The inclusion of the SJT scores in Step 3 accounted for an additional 0.9% of variance (Total $R^2 = .347$, $p = .162$). The SJT did not have a unique effect ($B = .022$, $p = .162$). The SJT scores did not significantly explain actual job performance scores when cognitive ability is included in the model. Only cognitive ability ($B = .652$, $p < .001$) and gender ($B = -1.772$, $p = .010$) had unique effects on the significant model.

The fourth step further included the FOT variable and this accounted for an additional 1.4% of variance in a significant model summary (Total $R^2 = .360$, $p = .085$). Only cognitive ability ($B = .634$, $p < .001$) and gender ($B = -1.763$, $p = .010$) had unique effects at the .05 level on the significant model. FOT had a significant main effect at the .1 level. This suggests that Hypothesis 6 (FOT, cognitive ability, SJT and job performance) is partially supported.

Table 17. Hierarchical multiple regression to investigate criterion validity of the SJT, formal operational thought and job performance.

	Step 1 B	Step 2 B	Step 3 B	Step 4 B
Age	-.16	-.44*	-.37	-.38
Gender	-2.13*	-1.82*	-1.77*	-1.76*
Cognitive Ability		.69*	.65*	.63*
SJT			.02	.02
Formal operations				.68 ¹
Change R^2	.049*	.288*	.009	.014
R^2	.049	.337*	.347	.360
F(df)	3.68 (2, 143)	61.80 (3,142)*	1.98 (4,141)*	3.01 (5,140)*

(* $p < .05$)

¹ Significant at the .1 level. To be interpreted with caution based upon the small sample size and lack of power in the measure. These ideas are further discussed in the remainder of this study.

Discussion

Hypotheses Summary

Cognitive ability was significantly associated with SJT scores, however not with formal operational thought scores. This gave support for Hypothesis 1 (cognitive ability and SJT) but not for Hypothesis 2 (cognitive ability and FOT). FOT also failed to significantly explain variance in SJT results which consequently meant Hypothesis 3 (FOT and SJT construct validity) was rejected. The SJT scores did significantly explain job performance but this finding did not withstand the inclusion of cognitive ability into analyses, and furthermore, FOT had incremental validity over and above the SJT when explaining job performance. These results gave partial support to Hypothesis 4 (SJT criterion validity) and Hypothesis 5 (FOT and job performance). The cognitive ability scores did significantly predict job performance. Hypothesis 6 (FOT, cognitive ability, SJT and job performance) was partly supported from the FOT effect significant at the .1 level, although the SJT scores failed to add incremental validity to the final regression model which did not give support for the hypothesis. Taking this with the previous regression model, FOT should be interpreted as a predictor of job performance, albeit with some caution, when cognitive ability is included in analyses. It is possible with a larger participant number to increase statistical power and a more robust measure of formal operational thought that this finding would be significant at the .05 level in a model including cognitive ability.

Discussion of the findings should therefore be interpreted with some caution due to this. Results from both regression models together suggest that cognitive ability fully explains the relationship between the SJT and job performance.

General Results Summary

Educational and practical exam performance

Educational Performance scores were generally high and with a small amount of variance. Participants also received generally high scores on their practical on the job exam during their final year (the OSCE results). This is unsurprising as they are screened thoroughly in terms of intelligence and knowledge capacity even before medical school. After five years of the medical undergraduate course, one would expect high performance across both written and practical examinations and furthermore, those not achieving the necessary standards for progression are forced to re take a year or drop out of the course leaving only the better performers in course. Indeed the OSCE examinations are similar in format each year and are

a recurrent part of the course, hence by final year, and the time of graduation the performance level should be high, although the material included becomes increasingly difficult throughout the course.

Due to time constraints, resources and funding, it was not possible to establish a measure of actual 'on the job' performance (the students had not been working for a year when the period of research ended). Therefore, the OSCE (practical exam in the final year of medical school) was used to represent a high fidelity situation that demonstrates practical job skill and ability, i.e. it is a potential fit for a job performance measure. Justification for this measure was noted in the methods section of this study.

Formal operational thought performance

Participants achieved typically high marks on the FOT measure with all demonstrating levels of the lower stages of higher thought and a large proportion scoring full marks on the test. However, there was still variance evident across the sample. Furthermore, more participants answered correctly for the lower levels of thought on the scale in comparison with the higher levels of thought, which is a logical outcome as it would be expected that not all would have reached higher level thinking stages. The formal generalisation question was only answered correctly by 65.2% of the sample with even less answering the abstract reasoning question successfully (51.2%).

This suggests that the majority of the medical student sample was demonstrating levels of concrete operational thought. This is the stage preceding formal operational/ higher level thought where an individual can reason about concrete objects and situations (i.e. those that in front of the individual and that the individual is physically able to touch) yet not hypothetical scenarios or intangible concepts/emotions. An individual is able to grasp principles of categorisation and conservation yet abstract reasoning is absent. As well, approximately 89% of the sample demonstrated early formal operational thought abilities. However, there is a drop off when we look at formal generalisation abilities, with approximately 40% failing to show these reasoning abilities. Furthermore, there is another, albeit relatively small, drop when looking at abstract reasoning/reasoning under uncertainty (just under half of the sample fail to show evidence of this stage).

This was a final year medical student sample, who were due to shortly be graduating from medical school and consequently beginning their first year Foundation job roles as doctors, with both high responsibility and intense pressure. Notably it is of concern that over

approximately one third of the sample were failing to show formal generalisation abilities and even less showed the ability to reason abstractly and in uncertainty according to the test.

SJT performance

Performance on the SJT as a whole sample showed a generally normal distribution of results. The SJT was made up of two sections (ranking and multiple choice answer formats) and the sample tended to be distributed across a normal distribution for each section with a fairly small amount of variance around the mean. This suggests that in general the sample performed well, although there are outliers that performed particularly poorly and those that performed particularly well, as to be expected from a normal distribution curve.

In agreement with existing research (e.g. Lievens and Patterson, 2011) and hypothesis 1, the SJT scores were significantly related to cognitive ability showing links between increased cognitive ability and better performance upon the SJT. Furthermore, the SJT significantly explained 11.1% variance in actual job performance in the regression analysis of criterion validity of the SJT giving support for hypothesis 4. This is in line with previous research where meta analyses indicated an explanation of 8% variance in job performance (McDaniel et al., 2006). However, when cognitive ability and SJT were examined as predictors of job performance together, the SJT failed to significantly explain variance in job performance suggesting that the SJT did not add anything further in terms of predicting medical job performance, over cognitive ability measures. Cognitive ability fully explains the relationship between the SJT performance and job performance.

This is inconsistent with the previous research noted from Lievens & Patterson (2011) who used a clinical problem solving test as a measure of declarative knowledge which acted as a proxy for 'cognitive ability' and found the direct effect of knowledge upon job performance did not remain significant when the SJT was included in analysis. This inconsistency and the strength of the cognitive ability predictor in this study may be due to the strength of the measure used (i.e. the EPM encapsulated all four years of university performance as well as other achievements such as prior degrees or publications). This may also be due to the inter correlation between the SJT in this study and the measure of cognitive ability; although only reported as $r=.256$, $p=.001$, which may be interpreted as a relatively low positive correlation) it is possible that the variance explained by the SJT in job performance (without the inclusion of cognitive ability in analysis) is fully explained by the aspect of the SJT correlating with the intelligence measure.

Formal Operational Thought Discussion

Formal operational thought as a concept

The FOT scores did not significantly correlate with cognitive ability or the SJT scores. However, FOT scores were significantly positively correlated with job performance. This indicated that FOT was not related to cognitive ability. It suggests a separate cognitive entity to declarative knowledge (as assumed to be the main component of academic attainment in the cognitive ability measure) and as expected was not used for cognitive or intelligence test success. This is somewhat surprising as FOT is conceptualised as an aspect of intelligence from the CHC model, and specific intelligence abilities are known to correlate with more general measures of intelligence (i.e. 'g') and hence a relationship between these two variables would have been likely. The FOT scores were not significantly related to the SJT scores suggesting that, for this medical selection SJT, the FOT knowledge is not involved in the typical 'procedural knowledge' that is widely held to be assessed by the SJT. Cognitive ability was significantly related to SJT performance and could explain 8.2% of the variance in these scores suggesting that this SJT assessed the construct of cognitive ability (declarative knowledge). This is in line with cognitive ability fully explaining the SJT as a predictor variable for job performance due to the overlap in the constructs assessed. FOT however, conceptually, appears to be separate to both declarative and procedural knowledge stores.

These results indicate that FOT is an ability that is nevertheless used in actual job performance for medicine; higher levels of FOT proficiency were significantly related to better job performance. Formal operational thought was not significantly related to SJT performance, but when examining actual job performance, the formal operational thought aided explanation above and beyond the information from the SJT score. This finding was somewhat upheld even when investigated alongside cognitive ability as a predictor for job performance. Whilst cognitive ability significantly and consistently predicted job performance (28.8% variance explained) the FOT scores had incremental validity over this that was significant at the .1 level ($p=.085$). This suggests that the finding should be interpreted with caution due to the small amount of variance explained and the significance value, but that there is some merit in further investigating this relationship. With a more rigorous methodology (i.e. a scale that was designed and shortened in a systematic manner) and a larger sample size this may very well be found significant at a .05 level. In fact the FOT measure lacks some power in that the initial lower levels of the hierarchy assessed by the early questions (q1-3) were answered correctly by almost the entire sample. Any variance or

effect from the measure is therefore rooted in the final three questions where the variance in performance is seen. It is noteworthy that an effect was found with only three questions contributing to the variance in performance examined.

A finding of this study holds from the regression model that does not include cognitive ability as a predictor. Medical students have been rigorously screened for above average intelligence on entry to university, and hence cognitive ability measures are rarely used in further medical selection; they may be rendered less applicable (Patterson et al., 2012). Furthermore, it is very strongly grounded in past literature that cognitive ability tests are valid predictors of performance across numerous jobs and professions, and in fact they consistently produce the highest validities of all individual differences predictors for the largest variety of jobs and settings (Ones, Dilchert & Viswesvaran, 2012). The large percentage of variance in job performance explained by cognitive ability here may be somewhat due to the complexity of the medical job role. The validity of cognitive ability as a predictor increases with the complexity of a job and examples of these roles include medical doctors, pilots and attorneys (Ones, Dilchert & Viswesvaran, 2012).

Formal operational thought added an additional 3% in variance over the SJT's 6.2% variance explained. Whilst the figures may appear to be trivial or small, it should be noted that the sample size used was also relatively small and hence some statistical power was lost. Establishing incremental validity in this study suggests that in larger scale studies with a longer assessment of FOT, there could potentially be very meaningful and rich data revealed, i.e. larger percentages of variance explained and further significance, even adding incremental validity over cognitive ability. Schmidt & Hunter (1998) note that decades of applied psychological research has deemed cognitive ability as the best predictor of job performance, and that these tests are incomparable to other measures in terms of their validities and generalizability. Hence, it is noteworthy that FOT showed incremental validity over cognitive ability, albeit at the .1 level as opposed to the .05 level of significance. As suggested, it is likely with a larger sample size and a more robust measure of FOT the statistical power would increase for the relationship. Following such larger replications, the links between better job performance within medicine and FOT proficiency may lead to assessment, measurement and potentially teaching to enhance the FOT ability. The foundations for such ideas lie within the noted Cognitive Accelerations and coaching programmes. Furthermore, there is scope to develop more intricate measurements of the

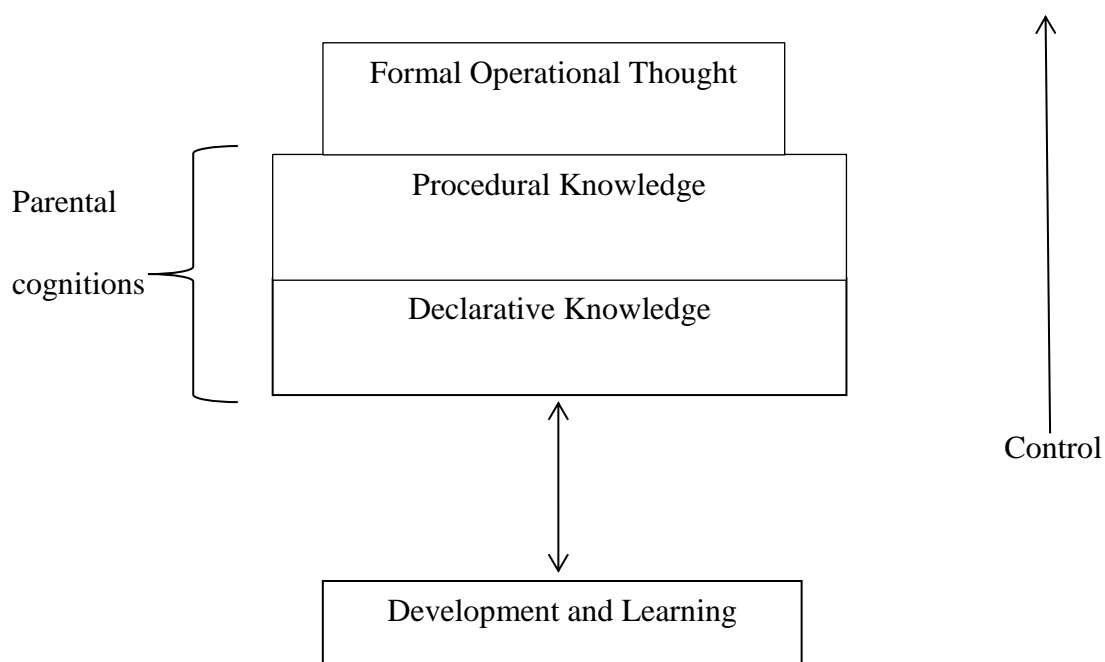
concept. Whilst the main stages of the thought levels are satisfactorily assessed in this study and measure used, there may be more to investigate.

To summarise the FOT concept, from the results of this study, the ability of formal operations can be conceptualised as an abstract thought store, independent of declarative and procedural knowledge basis that holds skills, information and abilities necessary and utilised for ‘on the job’ behaviour in medicine.

Theoretically future researchers may consider that the declarative and procedural knowledge stores are integrated and built up over the early stages of development and may act as parent tools that instruct, utilise and nurture the FOT abilities when the task is complex, uncertain and/ or one that requires abstract reasoning. As noted, procedural knowledge is widely held to be built upon declarative knowledge, and it is proposed that the abstract thought of formal operations is then built on this in turn. From this study, this abstract and creative thought appears to be utilised in medical job performance but not in this SJT performance. In certain situations, for example medical job behaviour, the parental tools may call on FOT to tackle the relevant challenge and cognitively deal with the situation in hand. Due to the significant relationships between cognitive ability and FOT established, solid bases of knowledge in both declarative and procedural knowledge stores may therefore be a pre requisite for FOT development and then utilisation.

See Figure 10 below for an illustration of this proposed idea.

Figure 10. Hierarchical build-up of cognition levels resulting in formal operational thought.



The three knowledge stores are seen as a route from development and learning and consequently knowledge is viewed from an ontogeny point of view, much as it was by Piaget himself. Knowledge can be seen as ever changing, evolving and improving as a result of experience, interactions and the environment. The procedural knowledge store is built on a pre-requisite of a foundation of declarative knowledge. The FOT ability is further built on both of these stores as an add-on for certain behaviours requiring the abstract intelligence contained in formal operations.

It is proposed that the basic stores of declarative and procedural knowledge have some form of control over the utilisation of FOT, i.e. they do not lead to its use for tasks that the two stores can handle alone, for example, basic rudimentary tasks such as reading and writing. The control direction is seen to flow from the development and learning direction and through declarative, procedural and then formal operations. This is due to the idea that knowledge is ever changing from the environment and experiences through development and learning hence, an individual's experiences and memory can alter when the two parental cognitions decide to utilise, or not as the case may be, the higher level thought of formal operations. In turn, FOT proficiency may then be seen to improve either as a result of solid knowledge stores resulting in a solid base of formal operations, or less established bases of thought yet the possession of relevant experience and internal knowledge of employment of the ability. This would explain how performance across non-job specific tasks can be similar across people with different specific declarative knowledge bases from different job experiences.

This model is a proposal of conceptual relationships that are based on the results of this study and the SJT used. Further research and investigation into these proposed cognitive links is necessary. The further development of these ideas into a model containing the SJT is included in the final discussion section of this thesis.

Formal operational thought and the SJT

FOT was not found to be a construct of interest when examining the construct validity of the SJT; it did not significantly correlate with the SJT nor add any incremental validity over cognitive ability when explaining SJT performance. These scores did not produce a significant regression model or main effect on SJT performance. This suggests that FOT does not directly affect performance on the SJT. This may in fact be a result of the particular SJT used and there is still potential for FOT to be investigated in relation to other SJTs within

different fields, with different instruction formats, different administration and/or different development procedures behind them. The versatility of the SJT also means that theory has to be adaptable and relationships between underlying constructs may change according to the specific SJT or field.

An important finding has nonetheless evolved from this study. FOT scores significantly added incremental validity over demographic controls and the SJT itself when looking at actual job performance. The FOT measure may therefore be of interest as a measure *alongside* the medical selection SJT, and not as a construct that the SJT measures within itself, as was hypothesised. This study is the first of its kind to examine the relationships between FOT, this medical selection SJT and job performance. It is a first step in this area and should results hold up in future studies, then there is scope for FOT measures to be used alongside an SJT in selection processes at the major gateway points in medical careers, and/or potentially education, to give further indicators of non-academic measures that can offer more information about potential future job performance of candidates.

To conclude, the ability to reason hypothetically and about future, intangible situations with complex reasoning abilities, as measured by the FOT test, seems logically to tie in with an individual's procedural knowledge. However, from this analysis, it is suggested that in fact the FOT capabilities are within a separate cognitive entity. This may explain the lack of relationship between FOT and the SJT itself. Since this study indicates that FOT is a factor in job performance within medicine, this could inform the development of SJTs to include this construct as one of assessment.

It is suggested that future research may attempt to investigate that in fact, the higher level reasoning abilities may lie so implicitly that they fall into another knowledge store, beyond that of procedural knowledge, as suggested in Figure 10. For example, the FOT fails to show correlational links with procedural knowledge measures such as the SJT itself, yet it does give additional explanation to job performance results. This suggests that FOT is not assessed by the SJT but could be an important factor or measure in personnel selection nonetheless. Hence, the FOT abilities do exist, are of importance in the medical occupation and do affect job performance, and this may be done through an alternative cognitive route aside from procedural and declarative knowledge stores.

Formal Operational Thought and SJT theory

The current SJT theoretical model (McDaniel et al., 2006) suggests that cognitive ability and personality factors (agreeableness, conscientiousness and emotional stability) are the four main constructs measured by the SJT. McDaniel et al., (2006) also support relationships between the SJT and job performance from their meta-analyses work. Other empirical studies (e.g. Lievens & Patterson, 2011) find the SJT to significantly predict job performance, over and above measures of declarative knowledge. The results of this study give support for cognitive ability as a construct assessed by the SJT, as well as a relationship between the SJT and job performance (although this appeared to be explained by the construct of cognitive ability). The findings from this study suggest that there is an overlap in measurement constructs between the cognitive ability and the SJT measures. This is in line with past research suggesting that knowledge tendency instruction SJTs will show correlations with cognitive ability measures (noted in Schmitt & Chan, 2006).

The SJT predicts job performance, but this finding does not withstand the inclusion of cognitive ability as a predictor variable and hence does not show incremental validity over measures of declarative knowledge. This may be due to the encapsulation of many different types of knowledge within the cognitive ability measure used in this study (i.e. it is likely it incorporates both procedural and declarative knowledge to some extent due to the nature of it covering academic attainment across four years as well as other achievements). Hence, the procedural element of the SJT that may show incremental validity over a cognitive ability measure may have been lost in this study.

Cognitive ability fully explains the SJT and job performance relationship. Cognitive ability is seen as the main predictor of job performance, and this is present in the SJT-job performance relationship due to an element of this SJT measuring cognitive ability. Furthermore, cognitive ability has extensively been reported as a predictor for job performance (Ones, Dilchert & Viswesvaran, 2012), and this link has been found to increase in strength with more complex jobs. Hence, finding cognitive ability as a strong predictor for job performance in the medical field is in line with these findings.

Conclusions

Formal operational thought appears to potentially be another piece in the jigsaw of cognition examined and used by professionals to inform personnel selection generally and specifically in medicine.

The higher level reasoning abilities contributed additional and unique variance above the well-established and widely used SJT when examining actual job performance. There is uncertainty about the theoretical location of this reasoning epicentre although there is scope to suggest that FOT is an important cognitive store built on a foundation of declarative and procedural knowledge stores that have been established through development and learning using the environment and interactions.

The results do not support existing theory. The results show links between the SJT assessing cognitive ability and predicting job performance, although this appears to be explained by overlapping in construct assessment of cognitive ability in the SJT from the findings.

The measurement and assessment of FOT could potentially provide further conceptual modelling of the factors affecting job performance in medicine. Furthermore, it can give incremental validity over the SJT itself when explaining job performance, although it does not appear to be a construct of intelligence that is assessed by the SJT measure itself. This study is the first step towards further understanding the theory surrounding and within FOT, the SJT and job performance. Future research is needed to further investigate these findings.

Limitations

The structure of the study was adequate and allowed for good participant and researcher interaction, yet interaction between participants was difficult to limit at times and there was potential conferring between participants at times.

The technical fault on the data collection day was a major obstacle and consequently, the study could not be run electronically as planned and this may have acted as a confounding variable. This may have put some participants off taking part and furthermore, increased the amount of effort needed to complete the task for them and hence, some participants did not answer the questions fully. Furthermore, the resulting lack of participation has greatly reduced the power of statistical testing in some areas, for example, the FOT scores. Hence, there may be a lot more power in the measure and concept that has gone overlooked in this study. This is an important point to note and future larger scale studies with a more rigorous method during data collection are needed regarding FOT, the SJT and resultant job performance.

It would have been preferable to have been able to include an actual on-the-job performance measure from actual workplace ratings collected within the subsequent Foundation training

years, or even later in their careers. The OSCE measure itself was used as an actual job performance measure and, as noted, there may be some criticism for the use of it as an on-the-job measure (Ram et al., 1999). However, this was the best option for this study and the closest obtainable score of job-like behaviour. It is very difficult to obtain on the job measures within medicine (Waas & van der Vleuten, 2009) and hence, information available during this thesis research period was adapted as a job performance measure and used as necessary. Similarly, the EPM measure was used to represent levels of cognitive ability. The limitations for using this have been noted already in the method section which considers that the EPM may include numerous factors such as motivation and life events, as opposed to just cognitive ability. However, the EPM was used due to this data being the only scores upon academic examinations available to the researcher. Furthermore, the fact that the data for the measure was obtained over a four year period and from numerous different examinations/measures is favourable for outlying occasions of bad performance to be averaged out to some extent.

Furthermore, in terms of relating the results to supporting the existing theory of the SJT (McDaniel et al., 2006) the inclusion of the personality variables from the stability factor (emotional stability, agreeableness and conscientiousness) would have given further scope to compare this research to the previous meta-analyses of empirical studies.

Finally, the construction of the short-form version of the FOT measure was not done under ideal circumstances and had this been the case, full evidence for the equivalence of this version would have been presented. For example, investigating correlations between the full and short form scales. The typical scale development process literature and application to the FOT scale is also discussed in study 2.

Future Research
There is potential to replicate the study and include additional longitudinal measures by assessing performance over the two year Foundation job roles, and furthermore, even into later job roles and speciality training years. The inclusion of an actual job performance measure would be a robust addition to this study and supervisor and/or patient ratings would be a suitable way to do this.

Additionally, the inclusion of other established constructs that the SJT assesses (e.g. personality factors) alongside the measure of cognitive ability would be useful and could consequently encourage the development of a complete theoretical model. Researchers should be encouraged to create similar studies to this one, using larger samples and taking

advantage of the nationally run medical selection SJT where there is scope for large scale research into this phenomena with a nationally administered SJT taking place at least annually. The SJT is also becoming an increasingly popular tool for selection into medical school itself, and this new use/tool at this particular gateway for selection is noted as an important addition to areas of potential research within medical selection and assessment.

The findings from this study may also be applicable to other career fields and the FOT ability itself may be important in other job roles that are not related to medicine. For example, within the police force - a high pressure job that requires reasoning of hypothetical scenarios, outcomes and judgments under time pressure as well as decision making that requires complex thought processes and can result in major human and/or financial costs if they are incorrect. Further research should target other careers and the applicability of FOT as an important skill within these and/or their selection procedures.

CHAPTER 7: DISCUSSION

Summary of Thesis

In study 1, a current model of the construct validity of the Situational Judgment Test (SJT) was examined (McDaniel et al., 2006) with the addition of two new non-cognitively oriented constructs that have not been examined before in relation to the SJT assessing other individual differences; the Need for Cognition (mapped onto the interests domain from Murphy, 2012) and Occupational Self-Efficacy (mapped onto the self-evaluation domain from Murphy, 2012) constructs. A cognitive ability measure unfortunately could not be assessed but the personality factors as a whole did explain variance in the scores, but only to a small extent. The conscientiousness factor and the openness factors were the only ones to show significant effects, and the openness to SJT performance link was a negative one. Results did not support only the inclusion of the three personality traits from the stability broad personality factor (α ; agreeableness, conscientiousness and emotional stability) but suggest that elements of both main broad factors (the plasticity factor, β) were also included in personality constructs assessed by this SJT. Furthermore the newly introduced variables assessing the interests and self-evaluation domains of individual differences (Need for Cognition and Occupational Self-Efficacy) did not provide any further explanation of these scores either. These two additional individual differences variables were in fact more strongly positively related to the personality traits not hypothesised to be constructs of interest from the McDaniel et al., (2006) model (i.e. the broad plasticity factor (β) consisting of openness to experience and extraversion).

It was concluded that the existing theory of construct validity was inadequate for this medical selection SJT and could not consistently explain variance in the SJT used in this study. The absence of strong links with the personality factors may be a result of the knowledge tendency questions used in the SJT and had a cognitive ability variable been included, this link would likely have been stronger. Results suggested that variance in SJT performance could be explained from inclusion of the five personality traits, as opposed to the three from the McDaniel et al., (2006) model and the stability (α) broad personality factor.

Consequently, it was concluded that following this research and other noted encouragement from researchers within literature in the field, that other constructs should be examined in an attempt to further explain the construct validity of the medical selection SJT used at this Foundation training gateway, and as a result, the surrounding theoretical ideas.

Two major strands of the research that needed development were discussed. Firstly, the need to bring theory into line with the large amount of empirical research surrounding the SJT was noted. It became apparent that the SJT as a tool, whilst widely established and acknowledged as a successful low fidelity measure, lacks theoretical explanation. The developed theory basis appears thin and mainly based around the McDaniel et al., (2006) ideas and the Motowidlo & Beier (2010) model with the newly developed ITP concept. At the time of writing this thesis, there were only these two major cognitive psychology theories of the SJT that conceptualised the construct and criterion validity of the measure.

Secondly, it was noted that fully understanding the SJT as a tool, and in particular within medicine as a field, was necessary and that this would in turn inform developing the theory as described above. In such a high risk procedure, such as high stakes testing in medical selection, decisions are made based upon momentary judgements and assessments. The SJT is used as one of these methods in selection and it is of utmost importance that we fully understand and can explain the method. Furthermore, from a psychological stance in particular, it is therefore necessary to explain what we are measuring, what we are asking of applicants, and additionally what skills or cognitive abilities we can say they are demonstrating through successful performance in the selection and assessment methods used. Investigation of the literature led to the examination of reviews of the SJT as a tool in terms of validity and practical applications. Despite the acceptance of the usefulness of the method, there appeared to be much more left to investigate with regards to the 'judgment' aspect of the SJT.

These gaps in the literature led to further inspection of already established theory and constructs that could aid explanation of SJT performance and further theoretical ideas. Consideration of the research surrounding judgment and intelligence (CHC Theory and Carroll, 1993) led to the targeting of the psychological phenomena of higher level thinking abilities and namely Jean Piaget's developmental theory of cognitive development (1952). Piagetian reasoning was identified as an area of the abstract intelligence concept of fluid reasoning within contemporary intelligence modelling. Whilst tacit knowledge and underlying factors of the SJT had been investigated in the past, formal operational thought (FOT) had not been investigated in relation to the SJT. This aspect of intelligence appeared to be a new potential construct that could be being measured by the SJT.

Derived from the intelligence literature, in study 2, an existing measure of FOT was extended and then piloted using two samples differing in basic cognitive abilities and the study focused upon validating this extended measure as best possible and in a suitable way for a scale of this nature (cumulative scores and based on the FOT concept stages). The study sample consisted of two groups; dyslexic and typical, and therefore analysis targeted FOT proficiency, performance on basic cognition tests from the DST-S and how these two measures were affected, or unaffected, by dyslexia group membership. The study aimed to establish and validate the new measure of FOT as a test showing a variance in scores across and within the samples used. The well-established theory surrounding the developmental disorder of dyslexia allowed for a measure of higher level thought (FOT) to be validated. As hypothesised, there was a lack of group differences across the FOT measure, yet a markedly poorer performance for the dyslexia sample on the primary cognition tasks. Hence, FOT as a higher level skill that is used for tasks requiring thought abilities above and beyond those used in primary cognitive tasks, was seen to be similar in both groups.

The results suggested that the FOT test was a valid measure that showed variance across the sample. The measure also managed to establish significant differences between the two samples where an existing measure of IQ did not (i.e. there were group differences across the different levels of FOT). The lack of a relationship between FOT and the DST-S tests, and in particular the non-verbal reasoning test (which is held to assess general cognitive ability) further suggests that this higher thought ability is used for problems and complex cognition where reasoning, thinking abstractly and malleable creative thought is required as opposed to standard cognitive tasks (e.g. reading/writing).

After the establishment of the measure and a basic theoretical conceptualisation of the cognitive entity of 'FOT', the measure was administered in a short form version to a medical sample alongside an SJT in the final study.

In study 3, SJT performance, cognitive ability, job performance and FOT were assessed in a sample of final year medical students. Results indicated that FOT was a separate cognitive base to declarative and procedural knowledge stores, and furthermore, that whilst FOT was not a construct useful in explaining SJT score variance, it added additional and unique incremental validity over SJT scores and demographic controls, when investigating actual job performance. Current SJT theory (McDaniel et al., 2006) that suggests cognitive ability as a construct assessed by the SJT and the SJT as a predictor of job performance was supported.

However, inter-correlations and further analysis between the SJT and cognitive ability suggested that the variance explained in job performance by the SJT was accounted for by cognitive ability.

Cognitive ability as a predictor of job performance furthermore dampened the effect of FOT as a predictor variable. It was suggested that with a larger sample size and a more robust measure of FOT that the finding would gain power. Furthermore, it was suggested that the strength of this cognitive ability measure (i.e. encapsulation of four years of university career progress and other achievements) may have given substantial weighting to this variable as a predictor. This may have encapsulated the procedural knowledge element of an SJT that may have given the SJT incremental validity over cognitive ability, had it been a pure cognitive ability assessment/measure of declarative knowledge. These results were represented in a model of the FOT, cognitive ability and the SJT predicting job performance.

Interpretation of this study led to further discussion of the results and a suggestion of a hierarchical model of knowledge stores. The higher level thought abilities are suggested to be built, in a hierarchical nature, on top of solid foundations of firstly, declarative knowledge and secondly, procedural knowledge. It is suggested that potential gaps or omissions in these two knowledge parent stores may be responsible for either a lack of or faulty/unsuccessful use of FOT abilities in some cognitive profiles. This may explain why not all adults develop the complex levels of thought (used in FOT) and additionally why, in study 2, the disorder dyslexia that impairs basic cognitions, did not impair FOT. The scope for the use of a FOT test, alongside an SJT, for selection processes in medicine was also discussed. This inclusion of another measure could offer further detail and information regarding non-academic abilities in job applicants regarding future job performance.

Final Discussion

From the three studies as a whole this thesis suggests that the construct validity of the medical selection SJT used at the Foundation selection gateway requires further research to establish more meaningful explanations of variance in SJT scores. Exploration of the literature has led to investigation into FOT as an abstract area of intelligence that may be measured by the SJT. The piloting of this measure allowed for the establishment of FOT as a conceptual knowledge base that is separate to rudimentary basic cognitions and used when complex challenges or reasoning is required. Furthermore, when the measure was used alongside an SJT, and both academic and practical exam measures in a medical sample, FOT

was found to be a separate entity to procedural and declarative knowledge, yet showed incremental validity over the SJT when examining actual job performance. This important finding came from the final study which investigated these variables within a medical sample for the first time. The study found comparable amounts of variance to previous research between the SJT and job performance (e.g. Lievens & Patterson, 2011) which allows for confidence in the results and a useful replication of this existing finding. The SJT is still a useful predictor and measure of job behaviour when used in selection without cognitive ability as a predictor.

Practical Implications for Personnel Selection

The above results have been reported and discussed as relevant based upon literature that states that elements of cognitive ability would not be used for medical selection past university entry (Patterson et al., 2012). Hence, without assessing cognitive ability, the SJT, and FOT together would significantly explain job performance in medicine from these findings. From a practical perspective, if in reality medical graduates are not classified or assessed according to their cognitive ability at all, then the measures used for selection should directly explain job performance. The SJT did explain job performance, and furthermore, the FOT measure had incremental validity over this suggesting that when used together these potential personnel selection measures could help predict job performance in medical applicants. Nonetheless, the main predictor of job performance was cognitive ability and hence the practical implications of FOTs small incremental validity may be limited.

Furthermore, the SJT failed to predict job performance over and above cognitive ability. This suggests that for this SJT the prediction of job performance was explained by the cognitive ability element of the simulation measure. This may have been a reflection of the cognitive ability measure used, and it is suggested that the study be replicated using a pure declarative knowledge assessment as opposed to the 'Educational Performance Measure' which may encapsulate both declarative and procedural knowledge elements. It is likely that such a replication will allow for the SJTs incremental validity over declarative knowledge measures shown in past research to become evident again.

This research suggests that in fact that the SJT used in this study adds no value to cognitive ability in terms of predicting medical job performance. It may be unnecessary and expensive to administer potentially irrelevant measures such as the SJT and FOT alongside cognitive ability measures or scores, if cognitive ability can explain future job performance alone and

to such a large extent as demonstrated in study 3. This practical implication is suggested with caution due to the methodology and measures used within this research.

Consequential Themes and Theory of Research

Whilst the thesis originally aimed to investigate the theories of the SJT and the constructs involved in these (i.e. exploration of the models and hence, the targeting of the McDaniel et al., 2006, model in study 1) there have been no conclusions drawn about further constructs in the SJT as the three additional constructs investigated in this research did not significantly explain the SJT performance (Need for Cognition and Occupational Self-Efficacy in study 1, and FOT in study 3). This suggests that the SJT does not target other ‘individual differences’ such as interests or self-evaluations, as from the Murphy (2012) framework. The main construct assessed by the SJT from this research is cognitive ability. It predicts the largest amount of variance and no other variables showed incremental validity over it when predicting job performance.

Research Question 1 asked ‘to what extent can variables from existing SJT theory explain the construct validity of the SJT within a medical selection context?’ From examining the McDaniel et al., (2006) theory in study 1 (the three personality variables of agreeableness, conscientiousness and emotional stability) and the cognitive ability variable in study 3 the findings suggest that the inclusion of only three personality factors from the ‘stability’ factor was not adequate to explain SJT performance- main effects were found from the ‘plasticity’ broad factor as well suggesting that all five personality factors should be included to significantly explain variance in the medical SJT used. Furthermore, these 5 factors together only explained an additional 3% of SJT performance variance over the demographic control measures, suggesting that there are likely numerous other constructs and factors being assessed by the SJT. In study 3, the cognitive ability variable explained a larger 8.2% of the variance in SJT scores in another SJT used for medical selection. This suggests that cognitive ability is a weightier construct being assessed by the medical SJT, and this is in line with research suggesting that knowledge tendency SJTs correlate more highly with cognitive ability measures in comparison with personality factors.

A major shortcoming in tackling this research question was the omission of a study investigating all four of the McDaniel et al., (2006) construct validity variables simultaneously. Future research should investigate these in a similar medical sample alongside an SJT used for medical selection (e.g. for Foundation applications as here). These

four predictor variables could be assessed in final year medical students and used within regression analyses to examine predictors of SJT performance in the final year of medical school for Foundation programme selection. This study could be extended to include further prediction analyses assessing the criterion validity of the SJT related to job performance longitudinally using supervisor ratings of on the job performance across the two following Foundation training years in various hospitals.

Research Question 2 asked ‘to what extent can the inclusion of additional non-cognitive variables assessing individual differences further enhance the construct validity of the SJT over and above variables from existing SJT theory?’ It was noteworthy that based upon past empirical work, McDaniel et al., (2006) in their meta-analytic based theory suggested that the three main personality factors being assessed by the SJT are agreeableness, conscientiousness and emotional stability, and in fact, the two new additional variables of Need for Cognition and Occupational Self-Efficacy correlated more strongly with the other remaining two personality factors from the broad ‘plasticity’ factor. This suggests that these additional non-cognitive measures of individual differences may be related to this broad aspect of personality, as opposed to the ‘stability’ factor. If the medical SJT had been designed to measure the three factors within the ‘stability’ factor (as suggested by existing theory) then the lack of relationships found between the other individual difference variables may be logical. Relationships may be found in future research if interest and self-evaluation measures are investigated in relation to personnel selection measures designed to assess the plasticity factor of personality as opposed to focusing upon the ‘stability’ factor of personality.

It is again a short coming of this research project that not all four main constructs of interest in relation to the SJT from the McDaniel et al., (2006) model could be included in a study together. Had this been possible, there would have been scope to look at the additional measures of interests and self-evaluations in relation to cognitive ability, personality traits and the SJT simultaneously.

Nonetheless, the thesis has resulted in three novel studies with useful findings regarding the SJT in medical selection and cognitive concepts. The thesis has importantly produced a meaningful finding that FOT gives incremental validity above and beyond that of an SJT when looking at actual job performance scores in a medical sample. To summarise, FOT does not appear to explain the construct validity of the SJT, but it may a predictor variable when job performance is an outcome criterion variable.

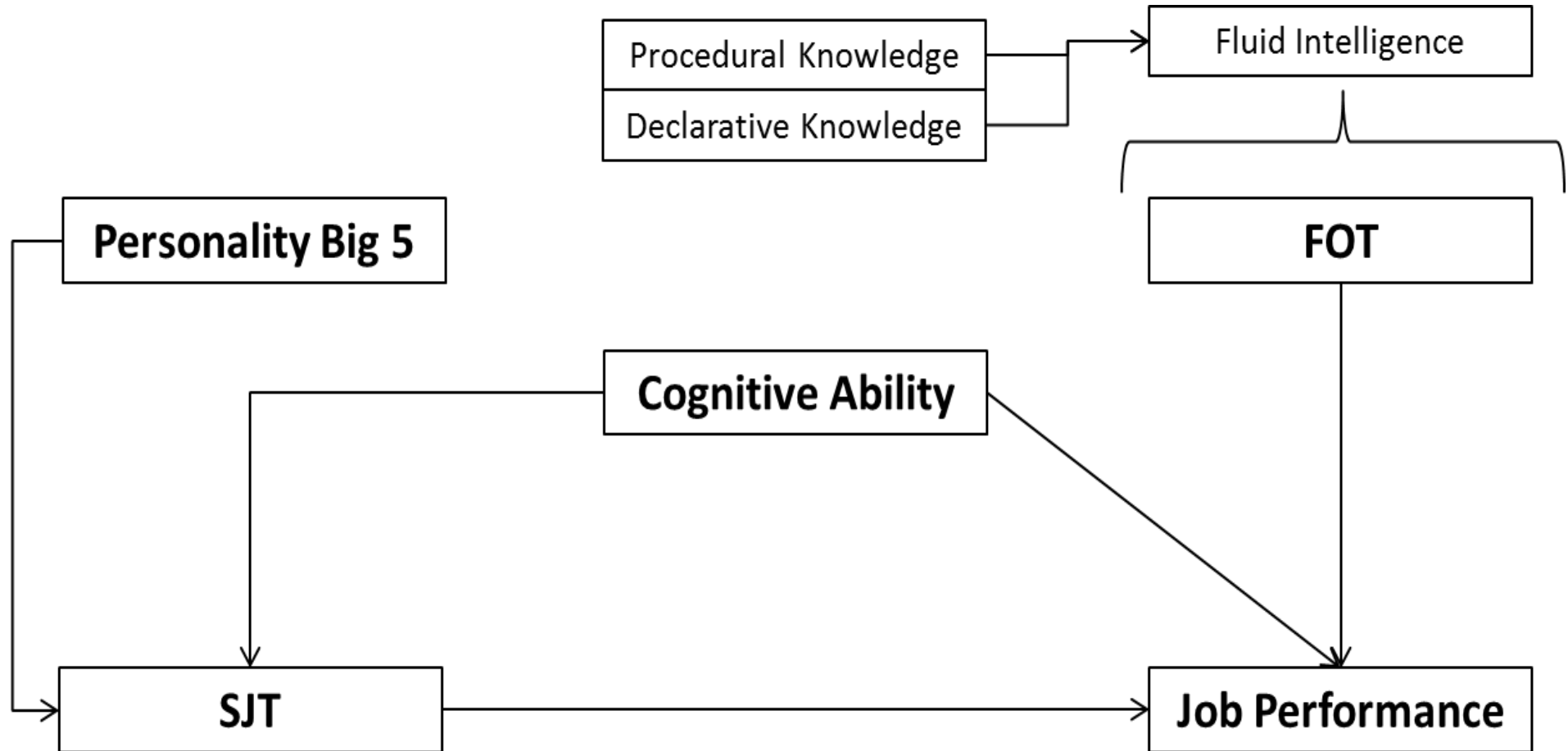
Research Question 3 asked ‘can the inclusion of an abstract intelligence variable assessing higher level thought further explain the construct validity of the SJT in a medical selection context over and above variables from existing SJT theory?’ and Research Question 4 asked ‘can the inclusion of an abstract intelligence variable assessing higher level thought further explain medical job performance independent of the SJT in a medical selection context? FOT failed to predict performance upon the SJT as hypothesised. This is potentially due to measurement issues (for example, the failure to develop and shorten the FOT measure in an ideal way, or potentially the measurement failures on the data collection day). This lack of finding may also be due to a lack of phenomenon; the hypothesis may have been incorrect, and post-hoc interpretation of the data suggests that in fact FOT may be encapsulated in personnel selection as a job performance predictor, as opposed to an SJT predictor. Due to a lack of previous research into FOT and personnel selection or any outcome criteria that were not school based achievements (e.g. Adey, Shayer & Yates extensive CASE work) the predicted theoretical location and effect of FOT may in fact fall as part of job performance, as opposed to SJT performance.

As noted in this discussion and the study 3 discussion above, the practical implications for such a finding may result in changes/additions of measures to personnel selection procedures within medicine if these findings can be supported in future studies. If cognitive ability is not assessed for past entry to university, then for later selection (e.g. speciality selection procedures) using FOT as well as an SJT for personnel selection will give incremental validity when explaining future job performance, based on these findings. However, when cognitive ability was included in analysis, it appeared to account for a large and noteworthy 28.8% of the variance in job performance and the SJT failed to show incremental validity over it. FOT showed incremental validity over cognitive ability, albeit only at the .1 level of significance. This is nonetheless a noteworthy finding due to the extensive research relating cognitive ability to job performance as the strongest predictor over and above all other measures, and of an increasingly large validity for more complex jobs such as medicine (e.g. Ones, Dilchert & Viswesvaran, 2012). Practically this may mean that cognitive ability is the only necessary selection measure needed for post graduate medical selection. Other possible explanations for this finding were considered in the study 3 discussion.

From a theoretical point of view there has been a consistent theme throughout the literature to bring current theory up to a point of agreement and balance with the current dominance of empirical and practical research regarding the SJT. It is still recommended in recent reviews

of the SJT that the relevant theory should be strengthened and used more often in SJT development and furthermore, SJTs grounded in theory tend to report higher reliabilities (Campion, Ployhart & MacKenzie, 2014). Research Question 5 asked ‘how can existing theoretical literature surrounding the SJT’s construct and criterion validity be expanded and linked to develop an integrative conceptual model of the SJT?’ The remainder of this final discussion will focus upon this research question and attempting to use this research project and existing theory to give an integrative and theoretical model and explanation of the SJT construct and criterion validity. In response to this research question, the results of study three were summarised in a model of the hierarchical stores of knowledge and a model representing cognitive ability as a mediator of the cognitive ability and job performance relationship (see Figure 10 previously). To incorporate the ideas from all three studies within this thesis, the existing intelligence literature and the SJT literature that has been examined (e.g. Carroll, 1993; CHC theory and McDaniel et al., 2006 respectively) an integrated model of the SJT and job performance has been proposed. The model incorporates findings from this thesis, and from intelligence and SJT literature into a proposed integrative explanation of the processes/relationships at play. This can be seen below in Figure 11. The studies using the SJT within this thesis have provided the empirical results to base this model on from within the medical field. This model derived from the data is therefore applicable to the medical selection SJT at the Foundation Year selection gateway. It is likely to be applicable across other careers and points of selection, and future research should investigate the proposed model and FOT.

Figure 11. An integrative theoretical model of intelligence concepts, knowledge stores, personality, the SJT and job performance.



The model represents formal operational thought within the fluid intelligence second order construct as in Carroll's (1993) model of intelligence. As suggested in this thesis, by Carroll and the more modern CHC theory, Piagetian reasoning has been noted within 'fluid intelligence'. Both this and the 'crystallised intelligence' second order factors are referred to as abstract intelligence capacities.

The separate stores of declarative and procedural knowledge are noted, and as proposed in study 3, procedural knowledge is built upon a basis of declarative knowledge, with fluid intelligence (FOT within this from Carroll's model) as a separate store that can be called upon by both of these when needed. It has been suggested in this thesis that these lower stores may act as parental cognitions that develop and provide foundations for, as well as target the capacity when needed, of FOT. These stores are placed above 'cognitive ability' as they are linked within the constructs of intelligence as well.

As established from study 3, FOT is shown to directly affect job performance, and yet not SJT performance. Study 1 established that personality factors significantly explained a small amount of variance in the SJT scores. Study 3 further established the links between SJT and job performance and the link of cognitive ability as a predictor of the SJT and job performance. The SJT and job performance finding did not withstand the comparison of cognitive ability as a predictor of job performance. All of these relationships can be seen in the model proposed. As noted in the introductory chapter this model is based upon the specific empirical data from the studies reported in this thesis, i.e. the two SJTs designed for medical selection at the Foundation year training gateway. However this model should be used as a guide for further SJT research into other specific SJTs designed for different purposes. It is noted that the relationships between personality and cognitive ability were not assessed in this research and neither were the relationships between SJT explaining cognitive ability performance. As noted in the discussion of study 3, future research should include studies that investigate FOT, personality, cognitive ability, SJT performance and job performance inclusively.

Furthermore, from Motowidlo & Beier (2010) there are other influences that may be investigated. For example, the effect of personality on the SJT could potentially go through an Implicit Trait Policy (ITP) to alter the expression of a personality trait. This has not been included on the model as it was not possible to examine the concept in this thesis. However, it is strongly recommended that this is investigated in future research as an addition and

another influential factor within the proposed model. Specific job knowledge and job experience may also be a route of effect upon job performance stemming from personality trait influences (Motowidlo & Beier, 2010).

Future Research and Practice Recommendations

This proposed model is a step towards a theoretical view of the medical selection SJT used in this research. It follows that FOT should be further investigated in different contexts. It is also suggested that the SJT would potentially show links with other outcome variables, not purely job performance. The diverse nature of the tool and the constructs assessed by the SJT suggest that it may be able to predict other criteria. For example, future criminal or deviant behaviour. The numerous constructs that the SJT can be designed to measure suggests that the criterion related validity outcome variable could also vary. The criterion related validity of the SJT should be examined in relation to other outcome variables and not purely job performance. There is therefore scope for the model to be developed by additional construct inclusion (e.g. emotional intelligence) or further investigation of the criterion related validity of the SJT (by exploring alternative outcome criteria for the measure). The relationships shown should also be considered in relation to susceptibility to coaching or faking on the measure used, although this would be expected to be low due to the knowledge tendency instructions used.

Further examination of this proposed model and the described relationships is encouraged. In particular, the key finding that FOT had incremental validity over and above the SJT when examining actual job performance. There is a need for a similar or close representation of study 3 with a larger sample and longitudinal job performance follow-up measures during medical training years. The nationally run SJT is an ideal opportunity for empirical research upon an annual basis in a real life high stakes selection setting. There are large year groups of medical students (e.g. approximately 300+) that can be recruited and asked for consent to access their academic history, SJT scores and job performance ratings over time. As noted it would be interesting to assess the main predictors of personality and cognitive ability simultaneously in such a sample using the available knowledge tendency SJT. This would allow for full examination of the existing McDaniel et al., (2006) model of SJT construct validity within the medical selection field. This study design could also be used to examine the proposed theoretical model above. In study 2 the full desired process for scale development and validation was discussed. Following this, the FOT measure could be implemented in future research. This would allow for further investigation into both the

construct and criterion related validity of the SJT and FOT, as a pair and individually through stepwise regression modelling.

The researcher recommends that alternative measures of cognitive ability could be used alongside the EPM measure that was used in this research project. This would allow for further investigation into the suggestion from these results that the EPM measure, used as an approximation of cognitive ability, may in fact be the only necessary and the best predictor measure for future job performance in medicine. Furthermore, the researcher recommends that the actual job performance measure follows the Foundation doctors through their two year training programme and uses objective supervisor ratings of performance during this period, as opposed to the practical examination of the OSCE.

The fact that cognitive ability appeared to fully explain the relationship between the SJT and job performance is interesting and suggests that future research should consider this, both within the medical selection field, but as well across other occupations that are currently using the SJT as a selection or assessment tool.

Further focusing upon FOT as a concept, it may be interesting to use an assessment of FOT and a validated tool for the measurement of fluid intelligence to examine the conceptual location of the FOT ability in relation to this as proposed from the CHC model and the diagram above. Further research development surrounding FOT outside of personnel selection may consider investigating the ability and differing strengths or weaknesses within it. For example, alterations in proficiency over age and/or across professions or gender. Although the Piagetian concept dates back to the 1950's there is relatively limited amount of recent research into this aspect of intelligence. There is a need to fully validate the developed measure from this thesis and consequently this could be a useful tool for such future research across numerous different samples.

The further development of the study and investigation of this relationship could inform selection procedures within medicine and what variables/ skills are seen as important for the career path and success. There is also scope to investigate the FOT variable within other careers and fields to see if the link remains/ exists. FOT would be expected to be associated with creative problem solving and therefore certain job roles over other job roles. For example, researchers may question whether FOT is important for jobs that involve problem solving and creative solutions, but not for other roles that only involve performing routine tasks repeatedly.

The expansion of theory is still an area of the SJT literature that needs work and time invested into it. Future research should attempt to incorporate other ideas from established psychological theory into models. For example, personality theories, social psychology and judgment making research (as described in the literature review of this thesis). Brooks & Highhouse (2006) distinguish the SJT from other commonly used ability tests as a measurement that uses both intuition and analytical thought together, as opposed to just one branch of thought alone. This view of 'good judgment' could be expanded upon and may provide another angle to view SJT performance from. Furthermore, using ideas such as this can lead to examination of new formats. For example, the scoring upon the ability to predict others behaviours in situations, as opposed to agreement with others (how most SJTs are marked at present).

From using such theories regarding personal traits, the situational aspect and the decision making aspect of the SJT, the reasons behind the links between SJTs and job performance may become clearer. Previous research shows numerous studies in the field and that report empirical data. The focus should now therefore fall to the examination and development of theory to inform SJT use and investigation. Rather than using SJTs in the field for empirical data it may be useful to develop experimental SJTs that hold no real consequences for poor performance, and therefore can be manipulated and altered to investigate the SJT further. It is important for the measure that solid and justifiable theory is developed to ensure full acceptance of the tool by practitioners and test takers. This point is reiterated by Campion, Ployhart & MacKenzie (2014) who state that

'it is not enough that there is an empirical relationship; we need to know why the relationship exists... for 20 years the field has progressed using SJTs for high-stakes decisions without fully knowing how or why those scores are indicative of future job performance'. (p.308, Campion, Ployhart & MacKenzie, 2014).

It follows that this relationship and theory needs to be better explained, and the ways to do this are through further explanation of the construct and the criterion validity of the SJT. When construct validity was examined the FOT construct was not a key variable in the SJT performance, hence there is still need for other constructs to be defined. This thesis originally intended to develop a new theory of the SJT using FOT as a new construct of assessment by the SJT, and not to establish an additional variable affecting job performance. However, this was not supported from the study findings and the evolution of the thesis has revealed

another interesting and key relationship for investigation. Importantly the thesis has established a meaningful finding regarding FOT and its incremental validity over the SJT in explaining actual job performance. Research should target new variables to investigate in relation to the construct and criterion related validity of the SJT.

Through the identification of these variables and measures, the explanation of the construct and criterion validity of the SJT can be improved and further explained. A recent paper by Ingold et al., (2014) attempted to explain the criterion validity of the Situational Interview (SI) method which is similar to the SJT. Both measures are known to predict job performance, although the reasons for these links are less evident. The noted paper found that the ability to identify the criteria used in the interview was significantly related to both measure performance and job performance. Due to the situational nature of this measure, and the SJT, research may take direction and ideas from such papers. This construct may be one of many that could be investigated in relation to the SJT or other situational measures.

For example, one may look to organisational culture; the job role and the 'fit' of an individual into a role or workplace. The diversity of jobs, fields and the skills helpful or necessary for success can vastly differ; different professions have particular cultures. This form of 'job culture' and organisational fit may be a factor of relevance within the SJT and, in particular, within medicine. How an individual behaves within the job place or setting, within the National Health Service (NHS) and how one interacts with other professionals and patients whilst ensuring the goals of the patient, doctor and organisation are met, are all at the forefront of modern medical care. These ideas may be constructs that are already being implicitly assessed by the SJT, or there may be scope to develop SJTs that are designed to assess these aspects of personnel selection. The context that an SJT is set for is therefore highly important and it is necessary to investigate constructs related to how one fits into a specific job culture through an SJT.

Furthermore, there is scope to investigate other implicit levels of thought or cognitive phenomena that may be relevant in the SJT as potential constructs of interest. For example, fuzzy reasoning abilities (Kosko, 1993) can be conceptualised as another higher level thought ability that may act above the level of formal operational thought. Fuzzy reasoning is a type of reasoning that deals with the non-exacts as opposed to the specifics in the world. This concerns the thinking about the 'grayness' as opposed to the black and whites, as described by Kosko himself. Non-academic aspects of behaviour and organisational fit are likely to be

assessed by the SJT and may lead to further explanation of performance upon the SJT measure. The versatility and numerous different types of the SJT measure makes the development of construct validity theory a difficult challenge for researchers.

Recent reviews of the SJT suggest that research should move towards systematically reporting the different SJTs used in the field and how these differ from one another (i.e. the attributes of the SJT) in order to establish which versions and formats of the tool need further investigation (Campion, Ployhart & MacKenzie, 2014). This review also notes the lack of clarity in reporting the reliability of the SJT and how a Cronbach's alpha may not be the optimum way to assess this.

It is also important to recognise the necessity for more research on the topic of FOT in general (i.e. outside the SJT research area). There are large cognitive challenges and barriers that FOT itself appears to bring with it. For example, relatively large proportions in both study 2 and 3 failed to display formal generalisation abilities. This is unsurprising and follows on from research such as Shayer and Ginsburg (2009) who found that levels of FOT are radically reduced from what they were in 1976. Superimposed on this decline in FOT, Shayer and Ginsburg describe the situation as one where school pupils appear to be developing concrete operational thought abilities very competently by the ages of approximately 14 years old, yet are failing to go on to develop the abstract reasoning and evaluative abilities of FOT. A possible explanation for this is that teachers have altered the way they teach in order to produce success in examinations rather than within problem solving or life situations etc. Therefore in terms of life success, adaptability and potential job success for those in high profile and high pressure jobs such as in medicine where reasoning is essential, this is surely problematic, not only for the applicants or job holders/employees themselves but also for organisations, service users and patients. There are potential directions for future research into this phenomena and its apparent decline.

For practitioners using, or contemplating using, the SJT within organisations and/or job selection and assessment, there are various considerations to be made that have been drawn from the topics within the literature. This is due to the numerous aspects of an SJT that can vary from one to another. Importantly, these aspects need to be chosen by practitioners to fit the task, aim, organisation and applicant in question. The main issues to consider concern the purpose of the SJT (i.e. for selection, assessment or purely formative) and therefore what information might be gained from it (for example information regarding a certain topic or a

more general field). SJTs can be manipulated and designed to fit a requirement, for example, through their response instructions (knowledge vs behavioural), how they are developed (for example, through job analysis or not) and/or the method of administration (video vs interview vs written). As noted in the literature, these can all affect how a candidate performs and what is revealed through the data. For example, behavioural tendency instructions are generally more likely to reflect personality traits as opposed to the use of knowledge tendency instructions. However, as a positive point, knowledge instructions may be faking resistant (Whetzel & McDaniel, 2009) and less susceptible to coaching. McDaniel et al., (2001) found that the design of an SJT through job analysis was a moderator of the validity of the measure with higher validities being reported for SJTs not based upon analysis. Even the scoring of an SJT may affect validity, and as shown in Motowidlo & Beier (2010), whilst numerous keys may be valid, only some (i.e. expert-based keys) allowed for incremental validity over and above other methods.

Final Comment

This thesis has brought together aspects of organisational, cognitive and developmental psychology to investigate the topic of medical selection and assessment, and in particular the SJT measure used in medical selection. This thesis evolved from the aim of investigation into new potential constructs of interest to the SJT, to result in a proposed theoretical model involving FOT as a variable giving incremental validity over the SJT when looking at actual job performance, with a framework of general SJT constructs of interest and contemporary intelligence models.

To summarise, through a set of three novel studies this thesis has investigated a main model of the SJT and following a lack of clear support for existing theory of the constructs involved in the SJT, investigated a new abstract aspect of knowledge from the intelligence literature: FOT. This higher level thought was conceptualised as a separate cognitive entity for abstract reasoning that through the final study was found not to be a construct assessed by the SJT. However, FOT did significantly explain job performance above and beyond the established selection method of the SJT. Cognitive ability was found to fully explain the SJT and job performance relationship. Practical and theoretical conclusions are discussed in relation to firstly, the inclusion of a formal operational thought measure in medical selection, and secondly a newly developed and integrative model of the SJT (construct and criterion validity). This finding gives a first step to consideration of other measures that may be useful

or insightful within medical selection, alongside existing and established ones such as the SJT.

Future research directions regarding examination of the newly developed model, potential new constructs of interest to the SJT to be investigated, general FOT research and advice for practitioners using the SJT have all been discussed to conclude the thesis.

Since the dramatic increase of SJT research in the 1990s there have been 59 empirical studies reported but there are still numerous unanswered questions regarding the SJT. Research has moved from themes of investigating criterion related validity of the SJT in the 90's, to subgroup differences in the 2000s and more recently the construct validity of the tool (Campion, Ployhart & MacKenzie, 2014). This thesis has attempted to provide a new theoretical perspective on the measure from studies within medical selection and by investigating new constructs that may be related to SJT performance. It is suggested now that the focus of research turns to the theoretical underpinning of the SJT and how this can inform best use of the tool. This shift in view could potentially give way to research to greatly improve the understanding, application and evaluation of the method in the future.

References

- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.
- Ackermann, H., Mathiak, K., & Ivry, R. B. (2004). Temporal organisation of ‘internal speech’ as a basis for cerebellar modulation of cognitive functions. *Behavioural and Cognitive Neuroscience Reviews*, *3*, 14-2.
- Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum. *Curriculum and Instruction*, *11(1)*, 1-29.
- Adey, P., & Shayer, M. (1994). *Really raising standards*. London, UK: Routledge.
- Adey, P., Shayer, M., & Yates, C. (2001). *Thinking Science: Student and Teachers’ Materials for the CASE Intervention*, 3rd ed. London, UK: Nelson.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice Hall.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge/Reading, MA: Addison-Wesley.
- Allport, G.W. & Odbert, H.S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47(211)*.
- Alon, A., Bray, D. W., Byham, W. C., Crooks, L. A., Grant, D. L., Hellervik, L. W. et al. (2009). Guidelines and ethical considerations for assessment centre operations. *International Journal of Selection and Assessment*, *17*, 243-253.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC; American Educational Research Association.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 389-406.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22, 261-295.

Anderson, M. (2008). What can autism and dyslexia tell us about intelligence? *The Quarterly Journal of Experimental Psychology*, 61, 116-128.

Anderson, N., & Cunningham-Snell, N. (2000). Personnel selection. In N. Chmiel (Ed.) *Introduction to work and organizational psychology*. Oxford: Blackwell.

Andrade, J., & May, J. (2004). *Cognitive psychology*. New York, NY: Taylor & Francis.

Armstrong, M. (2009). *Armstrong's handbook of human resource management practice*. London, United Kingdom.

Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion related validity of assessment centre dimensions. *Personnel Psychology*, 56, 125-154.

Ashton, M. C., Jackson, D. N., Helmes, E., & Paunonen, S. V. (1998). Joint factor analysis of the Personality Research Form and the Jackson Personality Inventory: Comparisons with the Big Five. *Journal of Research in Personality*, 32, 342-250.

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the Big-Five, and the five factor model. *Journal of Personality*, 73, 1321-1353.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., De Blas, L., et al. (2004). A six-factor structure of personality- descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86, 356- 366.

Baddeley, A. D., & Hitch, G. J. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, 28, 603-621.

Badian, N. A. (1984). Reading disability in an epidemiological contest: Incidence and environmental correlates. *Journal of Learning Disabilities*, 17, 129-136.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review* 84(2): 191-215.

Bandura, A., (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.

Bandura, A. (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Barker, R. G. (1963). *The stream of behavior*. New York, NY: Appleton-Century-Crofts.

Barker, R. G., & Wright, H. F. (1954). *Midwest and its' children: The psychological ecology of an American town*. Evanston, IL: Row, Peterson.

Barkley, R. A. (1997). Behavioural inhibition, sustained attention and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121, 65-94.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Barrick, M. R., Mitchell, T. R., & Stewart, G. L. (2003). Situational and motivational influences on trait-behaviour relationships. In M.R. Barrick & A.M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp.60-82). San Francisco: Jossey-Bass.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology*, 44, 1-26.

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters; Moving onto more important issues. *Human Performance*, 18, 359-372.

Barrick, M. R., & Mount, M. K. (2012). Nature and use of personality in selection. In Schmitt, N. (Ed.), *The Oxford handbook of personnel selection and assessment*. New York, NY: Oxford University Press.

Bauer, T. N., & Truxillo, D. M. (2006). Applicant reactions to situational judgment tests: research and related practical issues. In J.A. Weekley & R.E. Ployhart, (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real world learning environment. *Psychology and Aging, 20*, 341-355.

Berk, L. E., & Landau, S. (1993). Private speech in learning disabled and normally achieving children in classroom academic and laboratory contexts. *Child Development, 64*, 556-571.

Berk, L. E., & Potts, M. K. (1991). Development and functional significance of private speech among attention-deficit hyperactivity disorder and normal boys. *Journal of Abnormal Child Psychology, 19*, 357- 377.

Binet, A. (1905). New Methods for the Diagnosis of the Intellectual Level of Subnormals. *L'Année Psychologique, 12*, 191-244.

Bitner, B. L. (1991). Formal operational reasoning modes: Predictors of critical thinking abilities and grades assigned by teachers in science and mathematics for students in grades nine through twelve. *Journal of Research in Science Teaching, 28*(3), 265-274.

Blattberg, R. C., & Hoch, S. J. (1990). Database models and managerial intuition: 50% model+50% manager. *Management Science 36* (8), 887–897.

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology, 62*, 229–58.

Bond, M. H. (1979). Dimensions used in perceiving peers: Cross cultural comparisons of Hong-Kong, Japanese, American and Filipino university students. *International Journal of Psychology, 14*, 47-56.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L.A. (1993). Role of early supervisor performance. *Journal of Applied Psychology, 78*, 443-449.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76*, 863-872.

Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15*, 234–238.

Brooks, M. E., & Highhouse, S. (2006). Can good judgment be measured? In J.A. Weekley & R.E. Ployhart, (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Brunswik, E. (1952). *The Conceptual Framework of Psychology, International Encyclopaedia of Unified Science*. Chicago, US: University of Chicago Press.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131

Cacioppo, J. T., Petty, R. E., Feinstein, J, A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119, 197–253.

Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson and J. Smith (Eds.), *Toward a general theory of expertise: Prospects and Limits*. Cambridge, MA: Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Campbell, D. P., & Johansson, C. B. (1966). Academic interests, scholastic achievements and eventual occupations. *Journal of Counseling Psychology*, 13, 416-424.

Campion, M. C., Ployhart, R. E., & MacKenzie Jr., W. I. (2014). The state of research on Situational Judgment Tests: A content analysis and directions for future research. *Human Performance*, 27, 283-310.

Cantillon, P., & Glynn, L. (2009). Peer-assisted learning. In Carter, Y. & Jackson, N. (Eds.), *Medical education and training, from theory to delivery*. New York, NY: Oxford University Press.

Carrington, D. H. (1949). Note on the Cardall Practical Judgment Test. *Journal of Applied Psychology*, 33(1), 29-30.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, England: Cambridge University Press.

- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behaviour*. New York, NY: Springer Verlag.
- Cattell, J.McK., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, 3, 618-648.
- Cattell, R. N. (1957). *Personality and motivation structure and measurement*. New York, NY: World Book.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15(3), 233-54.
- Chan, D., & Schmitt, N. (2005). Situational judgment tests. In N. Anderson, A. Evers & O. Vosjuijil (Eds.), *Blackwell Handbook of Selection*. (pp219-242). Oxford, England: Blackwell.
- Choudrhy, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: the relationship between experience and the quality of health care. *Ann Intern Med*, 142 (4), 260-273
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational Judgment Tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1), 83-117. Clevenger, J., Pereira, G.M., Wiechmann, D., Schmitt, N. and Schmidt-Harvey, V.S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86(3), 410-7.
- Cohen, A.R. (1957). Need for Cognition and Order of Communication as Determinants of Opinion Change. In Hovland, C.I. (Ed.), *The Order of Presentation in Persuasion*. New Haven: Yale University Press.
- Cok, F. (1990). Body image satisfaction in Turkish adolescents. *Adolescence*, 25, 409-413.
- Connell, R. W. (1972). Political socialization in the American family: The evidence re-examined. *Public Opinion Quarterly*, 36(3), 323-333.

Corkum, P., Humphries, K., Mullane, J. C., & Theriault, F. (2008). Private speech in children with ADHD and their typically developing peers during problem solving and inhibition tasks. *Contemporary Educational Psychology, 33*, 97-115.

Costa, P. T., Zonderman, A. B., Williams, R. R., & McCrae, R. C. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology, 48*, 925-933.

Coughlin, L., & Patel, V.L. (1987). Processing of critical information by physicians and medical students. *Journal of Medical Education, 62*, 818-828.

Cowan, P. (1978). *Piaget: With feeling*. New York: Holt, Rineheart, and Winston.

Crain, W. (2005). *Theories of development Concepts and applications (5th ed.)*. Upper Saddle River, NJ: Pearson Education Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 6*, 671-684. Danner, F.W. and Day, M.C. (1977). Eliciting formal operations. *Child Development, 48*, 1600-1606.

De Guerrero, M. C. M. (2005). *Inner speech – L2: Thinking words in a second language*. Boston: Springer.

DeVellis, R. F. (2003). *Scale development: theory and applications (2nd ed. Vol. 26)*. Thousand Oaks, CA: Sage Publications.

De Young, C. G. (2006). Higher-order factors of the Big-Five in a multi informant sample. *Journal of Personality and Social Psychology, 91*, 1138-1151.

Digman, J. M. (1997). Higher-order factors of the Big-Five. *Journal of Personality and Social Psychology, 73*, 1246-1256.

Donaldson, T. (1996). Values in Tension: Ethics Away from Home. *Harvard Business Review, 48-62*.

DuBois, P. H. (1970). *A history of psychological testing*. University of Michigan: Allyn & Bacon.

Dyer, L. & Holder, G. Strategic human resource management and planning. In: Dyer, L., (ed) (1998) *Human resource management: evolving roles and responsibilities*. Washington DC: Bureau of National Affairs.

Eide, B. L. & Eide, F. F. (2011). *The Dyslexic Advantage: Unlocking the hidden potential of the dyslexic brain*. United Kingdom: Hay House.

Elkind, D. (1967). Egocentrism in adolescence. *Child Development*, 38(4), 1025–1034.

Elkind, D., & Bowen, R. (1979). Imaginary audience behaviour in children and adolescence. *Developmental Psychology*, 15, 33-44.

Ellis, A. (1973). *Reason and emotion in psychotherapy*. New York, NY: Julian Press.

Emick, J., & Welsh, M. (2004). Association between formal operational thought and executive function as measured by the Tower of Hanoi-revised. *Learning and Individual Differences*, 15(3), 177-188.

Fawcett, A. J., & Nicolson, R. I. (1992). Automatisation Deficits in Balance for Dyslexic Children. *Perceptual and Motor Skills*, 75, 507-529.

Fawcett, A. J., & Nicolson, R. I. (1994). Naming Speed In Children With Dyslexia. *Journal of Learning Disabilities*, 27(10), 641-646.

Fawcett, A. J., & Nicolson, R. I. (1995). Persistence of phonological awareness deficits in older children with dyslexia. *Reading and Writing*, 7, 361-376.

Fawcett, A. J., & Nicolson, R. I. (1998). Dyslexia: The role of the cerebellum. *The electronic journal of research in educational psychology*, 2(2), 35-58

Fawcett, J. & Nicolson, R. (2004). *The Dyslexia Screening Test- Secondary (DST-S)*. London, UK: Pearson Assessment.

Fejfar, A. J. (2007). A short test for formal operations. Found at <http://www.scribd.com/doc/95753/A-Short-Test-for-Formal-Operations-Thinking>

Fiske, S. T. & Taylor, S. E. (1991). *Social cognition (2nd ed.)*. New York, NY: McGraw-Hill.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn & Bacon.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment: Third edition*. Hoboken, New Jersey: Wiley & Sons.

Frank, L. K. (1939). Projective methods for the study of personality, *The Journal of Psychology: Interdisciplinary and Applied*, 8, 389-413.

Frawley, W. (2008). Private speech, cognitive computational control, and the autism-psychois continuum. *Behavioral and Brain Sciences*, 31, 269 +.

Frink, D. & Klimoski, R. J. (1998). Toward a theory of accountability in organizations and human resources management, In G. R. Ferris (Ed.), *Research in personnel and human resources management*. Greenwich, CT: JAI Press.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological bulletin*, 101(1), 75-90.

Funke, U. & Schuler, H. (1998). Validity of stimulus- and response components in a video test of social competence. *International Journal of Selection and Assessment*, 6, 115-123.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*, New York, NY: Basic Books.

Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice*, New York, NY: Basic Books.

Gardner, H. (1995). *Reflections on multiple intelligences: Myths and messages*. Phi Delta Kappan, 77, 200–209.

Gardner, H. (1998). A reply to Perry D. Klein's "Multiplying the problems of intelligence by eight." *Canadian Journal of Education/Revue canadienne de l'éducation*, 23, 96–102.

Gellatly, A. (1987). Acquisition of a concept of logical necessity. *Human Development*, 39, 32-47

Gessner, T. L., & Klimoski, R. J. (2006). Making sense of situations. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Lawrence Erlbaum Associates.

Goldberg, L. R. (1992). An alternative "Description of Personality.": The Big Five-Factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.

Goldberg, L.R. (1993). The structure of phenotypic personality traits: Authors' reactions to the six comments. *American Psychologist*, 48, 1303-1304.

Gottfredson, L.S (1981). Circumscription and compromise: a developmental theory of occupational aspirations. *Journal of Counselling Psychology*, 28, 545-79.

Gottfredson, L. S. (2002). Gottfredson's theory of circumscription and compromise, and self-creation. In Brown, D. (Eds.), *Career choice and development*. San Francisco, CA: Jossey-Bass.

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its' claims and evidence. *Intelligence*, 31(4), 343-397.

Grandey, A. (2000). Emotion regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, 5(1), 95-110.

Gray, W. (1973). *Development of a written test based upon the model of Piaget*. Dayton, OH: University of Dayton.

Greenwald, G. A. (1968). On Defining Attitude and Attitude Theory, in A. G. Greenwald, T. C. Brock, and T. M. Ostrom, (Eds.), *Psychological Foundation of Attitude*, New York: Academic Press.

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition*, 31, 902-917.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44, 339-387.

- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255-262.
- Hastie, R., & Dawes, R. M. (2001). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Thousand Oaks, CA: Sage Publications.
- Hausknecht, J. P., Day, D. V., & Thomas, S. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683.
- Hays, R. (2009). Measuring professionalism. In Carter, Y. & Jackson, N. (Eds.), *Medical education and training, from theory to delivery*. New York, NY: Oxford University Press.
- Hayes, S. C. (1997). Behavioral epistemology includes nonverbal knowing. In L. J. Hayes & P. M. Ghezzi (Eds.), *Investigations in behavioral epistemology* (pp. 35-43). Reno, NV: Context Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.
- Hogan, R. (2006). *Personality and the fate of organizations*. Mahwah, NJ: Erlbaum.
- Hogarth, R. M. (2001). *Educating Intuition*. Chicago and London: The University of Chicago Press.
- Holland, J. L. (1973). *Making vocational choices: A theory of careers*. New York, NY: Prentice Hall.
- Hooper, F. H., Hooper, J. O., & Colbert, K. K. (1985). Personality and memory correlates of intellectual functioning in adulthood: Piagetian and psychometric assessments. *Human Development*, 28, 101-107.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. PhD dissertation, University of Illinois.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & W. Woodcock, *Woodcock- Johnson technical manual* (pp. 197-232). Chicago, IL: Riverside.
- Hough, L. M. (1992). The 'Big-Five' personality variables- construct confusion: Description versus prediction. *Human Performance*, 5, 139-155.

Hudak, M., & Anderson, D. (1990). Formal operations and learning style predict success in statistics and computer science courses. *Teaching of Psychology, 17*, 231-234.

Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. New York: Macmillan.

Huitt, W., & Hummel, J. (2003). *Piaget's theory of cognitive development*. Educational Psychology Interactive. Valdosta, GA: Valdosta State University.

Hunter, D. R. (2003), Measuring general aviation pilot judgment using a situational judgment technique, *International Journal of Aviation Psychology, 13*(4), 373-86.

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

Ingold, P. V., Kleinmann, M., Konig, C. J., Melchers, K. G., & Van Iddekinge, C. H. (2014). Why do Situational Interviews predict job performance? The role of interviewees' ability to identify criteria. *Journal of Business Psychology*, DOI 10.1007/s10869-014-9368-3

James, D., Yates, J., Nicholson, S. (2010). Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *The British Medical Journal, 340*-478.

Jensen, A. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. (1998). *The g factor*. Westport, CN: Praeger.

Johnson, E. J. (1988). Expertise and decision under uncertainty: Performance and process. In M. R. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Joint Committee on Testing Practices. (1998). *The rights and responsibilities of test takers: Guidelines and expectations*. Washington, DC: American Psychological Association.

Judge, T. A. (2009). Core self-evaluations and work success. *Current Directions in Psychological Science, 18*, 58-62.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The Core Self-Evaluations Scale (CSES): Development of a measure. *Personnel Psychology, 56*, 303-331.

Judge, T. A., Higgins, C., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability and career success across the life span. *Personnel Psychology, 52*, 621-652.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology, 92*, 107-127.

Kail, R. V., Cavanaugh, J. C., & Ateah, C. A. (2006). *Human development: A life-span view*. (First Canadian ed.). Toronto, ON: Nelson.

Kanfer, R., Ackerman, P. L., & Cudeck, R. (Eds.) (1989). *Abilities, motivation, and methodology: The Minnesota Symposium on Learning and Individual Differences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kasser, T., Koestner, R., & Lekes, N. (2002). Early family experiences and adult values: A 26-year, prospective longitudinal study. *Personality and Social Psychology, 37*, 1966-1981.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children-Second edition*. Circle Pines, MN: AGS.

Kaufman, J. C., Kaufman, S. B., & Plucker, J. A. (2013). Contemporary theories of intelligence. In J. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 811-822). New York, NY: Oxford University Press.

Keating, D. (1979). Adolescent thinking. In J. Adelson (Ed.) *Handbook of Adolescent Psychology*. New York, NY: Wiley.

Kenny, A. (1978). *The development of a diagnostic instrument based on a Piagetian model*. Nova, FL: Nova University.

Kibby, M. Y. (2009). There are multiple contributors to the verbal short term memory deficits in children with developmental reading disabilities. *Child Neuropsychology, 15*, 485-506.

Kim, N. S., & Ahn, W. (2002). The influence of naïve causal theories and lay concepts of mental illness. *American Journal of Psychology, 115*, 33-65.

Kline, P. (1993). *The handbook of psychological testing*, London: Routledge.

Kosko, B. (1993). *Fuzzy Thinking: the new science of fuzzy logic*. New York: Hyperion.

Krokos, K. J., Meade, A.W., Cantwell, A., Pond, S. B., & Wilson, M. A. (2004). *Empirical Keying of Situational Judgment Tests: Rationale and Some Examples*. Paper presented at the 19th Annual Conference of the Society for Industrial/Organizational Psychology.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate student's success. *Science*, *315*, 1080–1081

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations: Implications for Graduate Student Selection and Performance. *Psychological Bulletin*, *127*, 62-81.

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

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

Lambert, T. W., Goldacre, M. J., & Parkhouse, J. (1997). Reasons for considering leaving UK medicine: Questionnaire study of junior doctors' comments. *BMJ*, *329*, 1263-5.

Larson, L. M., Rottinghaus, P. J., & Borgen, F. (2002). Meta-analyses of Big Six interests and Big Five personality factors. *Journal of Vocational Behaviour*, *61*, 217-239.

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes*, *82*, 171-193.

Libbrecht, N., Lievens, F., Carette, B., & Côté, S. (2014). Emotional intelligence predicts success in medical school. *Emotion*, *14*, 64-73.

Lievens, F., Buyse, T., & Sackett, P. R. (2005a). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442–452.

Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (2012). The effects of coaching on situational judgement tests in high-stakes selection. *International Journal of Selection and Assessment, 20* (3), 272–82.

4 Medical Schools Council. UK

Lievens, F., & De Soete, B. (2012). An update on the diversity-validity dilemma in personnel selection: A review. *Psychological topics, 21*(3), 399-424.

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96*(5), 927-940.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*(4), 426-441.

Lievens, F., & Sackett, P. R. (2006). Video based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181-1188.

Lievens, F., & Sackett, P. R. (2007). Situation judgment tests in high stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology, 92*, 1043-1055.

Lievens, F., & Sackett, P.R., (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*(2), 460-468.

Lilienfeld, S. O., Wood, J. M. & Garb, H. N. (2001). What's wrong with this picture? *Scientific American, 81–87*.

Little, P., Everitt, H., Williamson, I., Warner, G., Moore, M., Gould, C., Ferrier, K., & Payne, S. (2001). Preferences of patients for patients centred approach to consultation in primary care: observational study. *The British Medical Journal, 322*, 468-72.

Liu, C. J., & Shen, M. H. (2011). Influence of different representations on solving concentration problems at elementary school. *Journal of Science Education and Technology*, 20, 621-629.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694.

Lohman, D. F. (2001). Fluid intelligence, inductive reasoning, and working memory: Where the theory of multiple intelligences falls short. In N. Colangelo & S. G. Assouline (Eds.), *Talent development IV: Proceedings from the 1998 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* (pp. 219–227). Scottsdale, AZ: Gifted Psychology Press.

Luria, A. R. (1966). *Higher cortical functions in man*. New York, NY: Basic Books.

Luria, A. R. (1970). *Traumatic aphasia*. The Hague: Mouton.

Luria, A. R. (1973). *The working brain*. New York, NY: Basic Books.

MacLane, C. N., Barton, M. G., Holloway-Lundy, A. E. and Nickels, B. J. (2001). *Keeping score: Expert weights on situational judgment responses*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Madrid, H., & Patterson, M. G. (in press). Creativity at work as a joint function between openness to experience, need for cognition and organizational fairness. *Learning and Individual Differences*. (Accepted for publication December 2014).

Magai, C., & McFadden, S. H. (1995). *The role of emotions in social and personality development: History, theory, and research*. New York, NY: Plenum Press.

Markon, K. E., Krueger, R.E., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, 88, 139-157.

Martin, G. N., Carlson, N. R., & Buskist, W. (2007). *Psychology (3rd ed.)*. Harlow, Essex, England: Pearson Education.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist*, *63*, 503–517.

McCrae, R. R., & Costa, P. T., Jr. (1985). Comparison of EPI and Psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, *6*, 587-597.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81-90.

McCrae, R. R., & Costa, P. T. Jr. (1990). *Personality in adulthood*. New York: Guilford.

McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five-Factor Model of personality: Theoretical perspectives*. New York, NY: Guilford Press.

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O.P. John, R. W. Robins & L. A. Pervin (Eds.), *Handbook of research in personality* (3rd ed., 159-181). New York; Guildford Press.

McDaniel, M.A., Hartman, N. S. & Grubb III, W. L. (2003). *Situational Judgment Tests, Knowledge, Behavioral Tendency, and Validity: A Meta-Analysis*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology. Orlando.

McDaniel, M.A., Hartman, N.S., Whetzel, D.L. & Grubb, W.L., III (2007). Situational judgment tests, response instructions and validity: A meta-analysis. *Personnel Psychology*, *60*, 63-91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*(1-2), 103-113.

McDaniel, M.A. & Whetzel, D.L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, *33*, 515-525.

McDaniel, M.A., Whetzel, D.L., Hartman, N. S., Nguyen, N. & Grubb, W. L. (2006). Situational judgment tests: Validity and an integrative model. In R. Ployhart & J. Weekley (Eds). *Situational judgment tests: Theory, measurement, and application*. Jossey Bass. 183-204.

McDaniel, M. A., Whetzel, D., Schmidt, F. L., Maurer, S. (1994). The validity of the employment interview: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*, 599-616.

McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive CHC framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests and issues* (pp. 151-180). New York, NY: Guilford Press.

McHorney, C. A., Ware, J. E. Jr., & Raczek, A. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and Clinical Tests of Validity in Measuring Physical and Mental Health Constructs *Medical Care*, *31*, 247-63. McKay, N. J., & Covell, K. (1997). The impact of women in advertisements on attitudes toward women. *Sex Roles*, *36*, 573– 583.

McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Journal of Applied Cognitive Psychology*, *14*, 527–541.

McManus, C., & Goldacre, M. J. (2009). Predicting career destinations. In Carter, Y. & Jackson, N. (Eds.), *Medical education and training, from theory to delivery*. New York, NY: Oxford University Press.

Medical Schools Council, (2011). ISFP Final Report. Found at:
<http://www.isfp.org.uk/ISFPDocuments/Pages/FinalreportofPilots.aspx>

Medin, D. L., Altorn, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 333-352.

Milgram, S. (1963) Behavioural Study of Obedience. *Journal of Abnormal and Social Psychology*, 67, 371-78.

Millard, K. A. (1952). Is 'How Supervise?' an intelligence test? *Journal of Applied Psychology*, 36(4), 221-224.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65, S63-7.

Moore, K. A., Halle, T. G., Vandivere, S., & Mariner, C. L. (2002). Scaling Back Survey Scales How Short is Too Short? *Sociological Methods & Research*, 30(4), 530-567.

Moores, E., Nicolson, R. I. and Fawcett, A. J. (2003). Attentional Deficits in Dyslexia: Evidence for an Automatisation Deficit? *European Journal of Cognitive Psychology*, 15, 321-348.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683-729.

Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12(2), 125-135.

Moss, F. A. (1926). Do you know how to get along with people? Why some people get ahead in the world while others do not. *Scientific American*, 135, 26-27.

Motowidlo, S. J., & Beier, M. E. (2010). Differentiating Specific Job Knowledge From Implicit Trait Policies in Procedural Knowledge Measured by a Situational Judgment Test. *Journal of Applied Psychology*, 95(2), 321-333.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology, 91*(4), 749-761.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher order dimensions of the Big Five personality traits and the Big Six vocational interest types. *Personnel Psychology, 58*, 447-478.

Mumford, T. V., Van Iddekinge C. H., Morgeson, F. P., & Campion, M. A. (2008). The team role test: development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology, 93*, 250-67.

Murphy, K. R. (2012). Individual differences. In Schmitt, N. (Ed.), *The Oxford handbook of personnel selection and assessment*. New York, NY: Oxford University Press.

Muuss, R. E. (1996). *Theories of adolescence (6th ed.)*. New York, NY: McGraw-Hill.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Itasca, IL: Riverside Publishing.

Neimark, E. D. (1979). Current status of formal operations research. *Human Development, 22*, 60-67.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on failing a Situational Judgment Test. *International Journal of Selection and Assessment, 13*(4), 250-260.

Nickerson, R., Perkins, D., & Smith, E. (1985). *The teaching of thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). "Dyslexia, development and the cerebellum - Discussion." *Trends in Neurosciences, 24*(9), 515-516.

Nicolson, R. I., & Fawcett, A. J. (2006). Do cerebellar deficits underlie phonological problems in dyslexia? *Developmental Science, 9*(3), 259-262.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders? *Trends Neuroscience, 30*(4), 135-141.

- Nicolson, R. I., & Fawcett, A. J. (2008). Learning from the Science of Learning: Implications for the Classroom. *Dyslexia in Context: Research, Policy and Practice*, 257-277.
- Njiokiktjien, C. (1993). Developmental dysphasia- Concept and clinical aspects. *Zhurnal Neuropatologic i psikhiatrii*, 4, 88-92.
- Noelting, G. (1980a). The development of proportional reasoning and the ratio concept Part I—Differentiation of stages. *Educational studies in Mathematics*, 11(2), 217-253.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.
- Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population*. Department of Psychology, University of Michigan: Ann Arbor.
- Northrop, L C (1989). *The psychometric history of selected ability constructs*. Washington, DC: U.S.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Oatley, K. (2004). Emotional intelligence and the intelligence of emotions. *Psychological Inquiry*, 15, 216–221.
- Olson-Buchanan, J.B., Drasgow, F., Moberg, P.J., Mead, A.D., Keenan, P.A., Donovan, M.A. (1998), Interactive video assessment of conflict resolution skills, *Personnel Psychology*, 51(1), 1-24.
- Ones, D. S., Dilchert, S., & Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995-1027.
- Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive Abilities. In Schmitt, N. (Ed.), *The Oxford handbook of personnel selection and assessment*. New York, NY: Oxford University Press.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a bio data measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187-207.

Patterson, F. (2009). *Evaluating methods for high stakes selection: Operational validity and reactions to CVs, knowledge tests, SJTs and assessment centres*. Proceedings of the 14th congress of the European Association of Work and Organisational Psychology conference 2009. Galicia, Spain.

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgment tests to assess non-academic attributes in selection. *Medical Education, 46*(9), 850-868.

Patterson, F., & Ferguson, E. (2007). *Selection into medical education and training*. Understanding Medical Education Series, ASME Publications.

Patterson, F., Ferguson, E., Lane, P., Farrell, K., Martlew, J., & Wells, A. (2000). A competency model for General Practice: Implications for selection and development. *British Journal of General Practice, 50*, 188-193.

Patterson, F., Ferguson, E., & Thomas, S. (2008). Using job analyses to identify core and specific competencies for three secondary care specialties: Implications for selection and recruitment. *Medical Education, 42*, 1195-204.

Patterson, F., Lane, P., & Carr, V. (2009). Selection methods for medicine: Core concepts and future issues. In Carter, Y. & Jackson, N. (Eds.), *Medical education and training, from theory to delivery*. New York, NY: Oxford University Press.

Patterson, F., Lievens, F., Kerrin, M., Zibarras, L., & Carette, B. (2012). Designing selection systems for medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal of Selection and Assessment, 20*, 486-496.

Paullin, C. and Hanson, M. A. (2001). *Comparing the validity of rationally-derived and empirically-derived scoring keys for a situational judgment inventory*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Penfield, W., & Roberts, V. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Pervin, L.A. (1980), *Personality, Theory, assessment and Research (3rd Ed)*. New York, NY: Wiley & Sons.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cooks trans.). New York, International Universities Press.

Piaget, J., & Inhelder, B. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York, NY: Basic Books.

Piaget, J., & Inhelder, B. (1974). *The Child's Construction of Quantities*. London, UK: Routledge.

Ployhart, R. E. (2006). In J.A. Weekley & R.E. Ployhart, (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of Situational Judgment Tests. *International Journal of Selection and Assessment*, *11*(1), 1-16.

Plucker, J. A. (2000). Flip sides of the same coin or marching to the beat of different drummers? A response to Pyryt. *Gifted Child Quarterly*, *44*, 193–195.

Ram, P., Grol, R., Rethans, J. J., Schouten, B., Van Der Vleuten, C., & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility, *Medical Education Oxford*, *33*, 447-454.

Ram, P., Van der Vleuten, C., Rethans, J. J., Schouten, B., Hobma, S., Grol, R. (1999). Assessment in general practice: The predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. *Medical Education*. *33*(3), 197-203.

Raymond, M. R., Neutsel, S., & Anderson, D. (2007). Retest effects on identical and parallel forms in certification and licensure testing. *Personnel Psychology*, *60*, 367-396.

Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than g. *Intelligence*, *32*, 621–652.

Reeve, C. L., Bonaccio, S., & Winford, E. C. (2014). Cognitive ability, exam-related emotions and exam performance: A field study in a college setting. *Contemporary Educational Psychology*, *39*, 124–133.

Resnick, L. B. (1987). *Education and Learning to Think*. Washington, DC: National Academy Press.

Rethans, J., Norcini, J., Baron-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T., Lew, S., Page, G. G., & Southgate, L. H. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical Education*, *36*, 901-9.

Reznick, R. K., Regehr, G., Yee, G., Rothman, A., Blackmore, D., & Dauphinee, D. (1998). High-stakes examinations: What do we know about measurement? *Academic Medicine*, *75*(10), S97-S99.

Rigotti, T., Schyns, B. & Mohr, G. (2008). A short version of the Occupational self-efficacy scale. Structural and construct validity across five countries. *Journal of Career Assessment*, *16*, 238-255.

Rogers, C. (1951). *Client-centered Therapy: Its Current Practice, Implications and Theory*. London, UK: Constable.

Rosch, E. H., & Mervis C. B. (1976). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Rosenthal, R. & Jacobson, L. (1966). Teacher's expectancies: Determinants of Pupils' IQ gains. *Psychological Reports*, *19*, 115-118.

Roth, P., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*, 1009-1037.

Ryan, A. M., & Ployhart, R. E. (2013). A century of selection. *The Annual Review of Psychology*, *65*, 693-717.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, *56*, 471-499.

Sadowski, C. J. (1993). An examination of the short need for cognition scale. *Journal of Psychology*, *127*, 451-454.

Sadowski, C. J., & Gulgoz, S. (1992b). Internal consistency and test-retest reliability of the need for cognition scale. *Perceptual and Motor Skills*, *74*, 610.

Sadowski, C. J., & Cogburn, H.E. (1997). Need for cognition in the big five factor structure. *Journal of Psychology: Interdisciplinary and Applied*, *131*, 307-312.

Sadowski, C. J., & Gulgoz, S. (1996). Elaborative processing mediates the relationship between need for cognition and academic performance. *The Journal of Psychology*, *130*, 303-307.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. Imagination, *Cognition and Personality*, *9*, 185-211.

Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. San Diego: Jerome M. Sattler.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Schmitt, N. & Chan, D. (2006). Situational judgment tests: method or construct? In J.A. Weekley & R.E. Ployhart, (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., p99-144). New York, NY: Guilford Press.

Schyns, B. & von Collani, G. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology* *11*(2), 219-241.

- Scott, W. D., (1920). Changes in some of our conceptions and practices of personnel. *Psychological Review*, 27, 81-94.
- Sharp, S.E. (1899). Individual psychology: A study in psychological method. *American Journal of Psychology*, 10, 329–391.
- Shayer, M. (1979). Has Piaget’s construct of formal operational thinking any utility? *British Journal of Educational Psychology*, 49, 256-276.
- Shayer, M. (1999a). Cognitive acceleration through science education II: its’ effect and scope. *International Journal of Science Education*, 21(8), 883-902.
- Shayer, M. (1999). *GCSE 1999: Added-value from schools adopting the CASE Intervention*. London: Centre for the Advancement of Thinking.
- Shayer, M., & Adey, P. (1993). Accelerating the development of formal operational thinking in high school pupils, IV: Three years on after a two-year intervention. *Journal of Research in Science Teaching*, 30(4), 351-366.
- Shayer, M. & Adey, P. (2002). *Learning Intelligence*. Buckingham, Philadelphia: Open University Press.
- Shayer, M., & Ginsburg, D. (2009). Thirty-years on- a large anti-Flynn effect? (II): 13-14-year olds . Piagetian tests of formal operations norms 1976-2006-7. *British Journal of Educational Psychology*, 79, 409-418.
- Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities*, 22, 469-478. Cited in Fawcett AJ, Nicolson RI & Maclagan F (2001) Cerebellar tests differentiate between groups of poor readers with and without IQ discrepancy. *Journal of Learning Disabilities*, 34(2), 119-135.
- Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, 414-433.
- Slaughter, J.E., Christian, M.S., Podsakoff, N.P., Sinar, E.F., & Lievens, F. (2014) On the limitations of using Situational Judgment Tests to measure interpersonal skills: The moderating influence of employee anger. *Personnel Psychology*, 4, 847-885.
- Spearman, C. (1927). *The abilities of man*. London, England: Macmillian.

Stein, J., & Walsh, V. (1997). To see but not to read: the magnocellular theory of dyslexia. *Trends Neuroscience*, *20*, 147–52.

Stemler, S. E., & Sternberg, R. J. (2006). Using situational judgment tests to measure practical intelligence. In J. Weekley & R. Ployhart (Eds.) *Situational Judgment Tests*, (p107-131). Mahwah, NJ: Laurence Erlbaum Associates.

Sternberg, R. J. (1985). *Beyond IQ :A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (1997). *Successful intelligence*. New York, USA: Plume.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

Sternberg, R. J., & Horvath, J. A. (Eds.) (1999). *Tacit knowledge in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R. J., Nokes, K., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*, 401–418.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacit knowledge in work and at school. In H. Reese & J. Puckett (Eds.), *Advances in lifespan and development*. (pp. 205-227). Hillsdale, NJ: Erlbaum.

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, *20*(2), 503-530.

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*(2), 207-228

Storey, J. (1989). Personnel management to human resource management. In (ed) Storey, J. *New Perspectives on Human Resource Management*. Routledge, London.

Strong, E. K. (1943). *Vocational interests of men and women*. Stanford, CA: Stanford University Press.

- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology*, *54*, 101-114.
- Talbot, M. (2004). Monkey see, monkey do: A critique of the competency model in graduate medical education. *Medical Education*, *38*, 587-92.
- Taroyan, N. A., Nicolson, R. I., & Fawcett, A. J. (2007). Behavioural and neurophysiological correlates of dyslexia in the continuous performance task. *Clinical Neurophysiology*, *118*(4), 845-855.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analysis. *Personnel Psychology*, *46*, 259-293.
- Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin*, *34*, 275-284.
- Thornton, G. C., & Rupp, D. (2006). *Assessment centres in human resource management: Strategies for prediction, diagnosis and development*. Mahwah, NJ: Erlbaum.
- Thurstone, L. L. (1919). Mental test for college entrance. *Journal of Educational Psychology*, *10*, 129-142.
- Tuckett, D., Boulton, M., Olson, C., & Williams, A. (1985). *Meeting between experts: An approach to sharing ideas in medical consultations*. London/New York: Tavistock Publications.
- Tupes, E. C. & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings*. Lackland, TX: U.S. Air Force.
- UKCAT Consortium. (2015). *What is the UKCAT?* Retrieved January 7, 2015 from <http://www.ukcat.ac.uk/about-the-test/what-is-the-ukcat/>
- Umnikrishnan Nair, K., & Ramnarayanb, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality*, *34*(3), 305-328.
- Valanides, N. (1997). Formal reasoning abilities and school achievement. *Studies in Educational Evaluation*, *23*(2), 169-185. Van der Vleuten, C. (1996). The assessment of professional competence: developments, research and practical implications. *Advanced Health Science Education*, *1*, 41-67.

Van der Vleuten, C. (2000). Validity of final examinations in undergraduate medical training. *British Medical Journal*, *321*, 1217-19.

Vinchur, A. J. (2007). A history of psychology applied to employee selection. In L. L. Koppes (Ed.), *Historical perspectives in industrial and organizational psychology* (pp. 193-218). Mahwah, NJ: Lawrence Erlbaum Associates.

Vygotsky, L. S. (1986). *Thought and language*, A. Kozuli (Eds. and Trans.), Cambridge, MA: MIT Press.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology*, *52*, 1236-1247.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *The Journal of Personality and Social Psychology*, *49*, 436-458.

Wagner, R. K., & Sternberg, R. J. (1990). Street smarts. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership*. West Orange, NJ: Leadership Library of America.

Wallace, G. L., Silvers, J. A., Martin, A., & Kenworthy, L. E. (2009). Brief report: Further evidence for inner speech deficits in autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 1735-1739.

Wass, V., & van der Vleuten, C. (2009). Assessment in medical education and training 1. In Carter, Y. & Jackson, N. (Eds.), *Medical education and training, from theory to delivery*. New York, NY: Oxford University Press.

Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, *50*, 25-49.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, *52*, 679-700.

Weekley, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, *18*, 81-104.

Weekley, J.A., & Ployhart, R. E. (2006). (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Weekley, J.A., Ployhart, R.E., & Holtz, B.C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J.A. Weekley & R.E. Ployhart, (Eds.) *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ: Erlbaum.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples and criteria. *Journal of Applied Psychology*, 52, 372-376.

Wechsler, D. (1949). *The Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.

Wechsler, D. (1955) *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale (4th edition)*. San Antonio, Texas: Pearson.

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

Weekley, J. A. and Ployhart, R. E. (2006), *Situational Judgment Tests: Theory, Measurement and Application*, San Francisco, CA: Jossey-Bass.

Whetzel, D. L. & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19, 188-202.

White, J. (2008). Illusory Intelligences. *Journal of Philosophy of Education*, 42(3-4), 611–630.

Whitecotton, S. M., Sanders, D. E., & Norris, K. B. (1998). Improving predictive accuracy with a combination of human intuition and mechanical decision aids. *Organisational behaviour and human decision processes*, 76(3), 325-348.

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry*, 47, 857-865.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), 181-192.

Winsler, A., Diaz, R. M., Atencio, D. J., McCarthy, E. M., & Chabay, L. A. (2000). Verbal self-regulation over time in preschool children at risk for attention and behaviour problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41, 875-886.

Wissler, C. (1901) The correlation of mental and physical tests. *Psychological Review Monograph Supplements*, 3(6).

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock- Johnson III normative update tests of cognitive abilities*. Itasca, IL: Riverside.

World Federation of Neurology (1968). Report of research group on developmental dyslexia and world illiteracy. *Bulletin of the Orton Society*, 18, 21-22

Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A point scale for measuring mental ability*. Baltimore: Warwick & York.

Zalesny, M. D. & Highhouse, S. (1992). Accuracy in performance evaluations. *Organisational Behaviour and Human Decision Processes*, 51, 22-30.

Zeidner, M., Roberts, R. D., & Matthews, G. (2004). The emotional intelligence bandwagon: Too fast to live, too young to die? *Psychological Inquiry*, 15, 239–248.

Zimbardo, P. G. (1974). On "obedience to authority." *American Psychologist*, 29 (7), 566-567.

Appendix

Appendix I. Explanation of thesis terminology.

Term	Definition
Face validity	<p>Does the test appear to be a good measure of the variable it claims to measure?</p> <p>E.g. To the participant, does the personality test appear to assess one's personality?</p>
Concurrent validity	<p>How well does the independent variable correlate with the outcome criterion when data collected at the same time?</p> <p>E.g. How does a measure of personality correlate with a simultaneous job performance measure?</p>
Criterion validity	<p>How well does the variable predict the outcome of another variable?</p> <p>E.g. How much does a personality trait affect performance on a behavioral criterion taken at a later time?</p>
Construct validity	<p>Does the test accurately measure what it is supposed to measure?</p> <p>E.g. Does the personality measure actually measure levels of the trait it specifies?</p>
Incremental validity	<p>Does one test add explanation of variance, and therefore predictive ability, over another test?</p> <p>E.g. Does the SJT explain additional variance in actual job performance scores, above and beyond the Assessment Center?</p>
Reliability (Inter-case)	<p>An assessment of the consistency of performance of a participant across cases in a measure. This can be expressed as a coefficient (Cronbach, 1951) using Classical Test Theory and in high stakes testing cases (such as medical selection) a coefficient greater than 0.8 is usually aimed for (Wass & van der Vleuten, 2009).</p>

Reliability (Test-retest)	How well do scores at time1 testing match up with scores at time2 testing? E.g. Do scores testing on Monday correlate with scores from testing on the following Monday?
Moderating variable	A moderator ‘affects the direction and/or strength of the relation between an independent variable and a dependent variable’ (Baron & Kenny, 1986, p.1175). E.g. different levels of a moderator can cause different outcomes; one form of B may cause depression, but another form of B may not (as seen in model B1 predicts $A \rightarrow C$ Whereas, B2 predicts A does not $\rightarrow C$).
Mediating variable	A variable can be classed as a mediator ‘to the extent that it accounts for the relation between the predictor and the criterion’ (Baron & Kenny, 1986, p1181). E.g. A factor or mechanism that A goes through to get to C (as seen in model $A \rightarrow B \rightarrow C$).

Appendix II. Breakdown of five factors of personality*I. Surgency:*

Introverted- Extroverted

Unenergetic- Energetic

Silent- Talkative

Unenthusiastic- Enthusiastic

Timid- Bold

Inactive- Active

Inhibited- Spontaneous

Unassertive- Assertive

Unadventurous- Adventurous

Unsociable- Sociable

II. Agreeableness

Cold- Warm

Unkind- Kind

Uncooperative- Cooperative

Selfish- Unselfish

Rude- Polite

Disagreeable- Agreeable

Distrustful- Trustful

Stingy- Generous

Inflexible- Flexible

Unfair- Fair

III. Conscientiousness

Disorganised- Organised

Irresponsible- Responsible

Undependable- Reliable

Negligent- Conscientious

Impractical- Practical

Careless- Thorough

Lazy- Hardworking

Extravagant- Thorough

Rash- Cautious

Frivolous- Serious

IV. Emotional Stability

Angry- Calm

Tense- Relaxed

Nervous- At ease

Envious- Not envious

Unstable- Stable

Discontented- Contented

Insecure- Secure

Emotional- Unemotional

Guilt ridden- Guilt free

Moody- Steady

V. Intellect

Unintelligent- Intelligent

Imperceptive- Perceptive

Unanalytical- Analytical

Unreflective- Reflective

Uninquisitive- Curious

Unimaginative- Imaginative

Uncreative- Creative

Uncultured- Cultured

Unrefined- Refined

Unsophisticated- Sophisticated

Appendix III: Need for Cognition Scale (18 items)

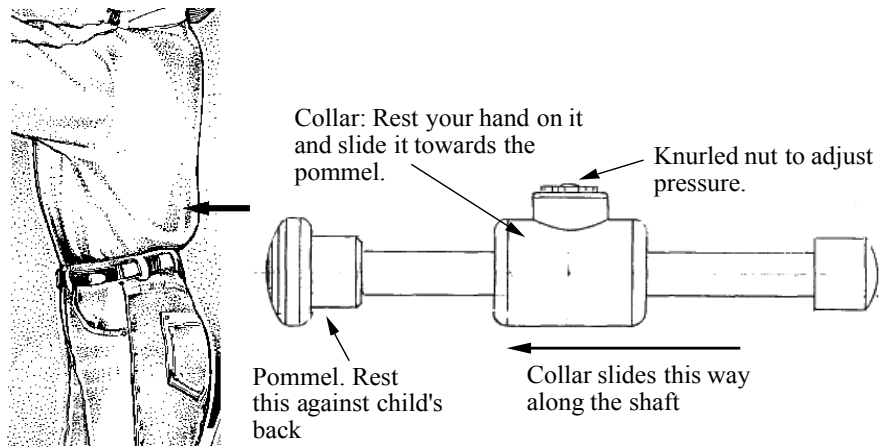
Reverse coded items are marked by an asterisk.

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.*
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.*
5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.*
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to.*
8. I prefer to think about small, daily projects to long-term ones.*
9. I like tasks that require little thought once I've learned them.*
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn't excite me very much.*
13. I prefer my life to be filled with puzzles that I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.*
17. It's enough for me that something gets the job done; I don't care how or why it works.*
18. I usually end up deliberating about issues even when they do not affect me personally.

Appendix IV: Short form version of Schyns & von Collani Occupational Self-Efficacy Scale

1. Thanks to my resourcefulness, I know how to handle unforeseen situations in my work
2. If I am in trouble in my work, I can usually think of something to do
3. I can remain calm when facing difficulties in my work because I can rely on my abilities
4. When I am confronted with a problem in my work, I can usually find several solutions
5. No matter what comes my way in work, I am usually able to handle it
6. My past experiences in work have prepared me well for my occupational future
7. I meet the goals that I have set myself in my work

I feel prepared to meet most of the demands in my work

Appendix V. Diagram of balance tester for postural stability test.

Appendix VI. Phonemic segmentation stimuli.

A Say <i>football</i>	Say it again, but without <i>ball</i>
B Say <i>breakfast</i>	Say it again, but without <i>/brek/</i>
C Say <i>cat</i>	Say it again, but just say the first sound. (prompt with c-c-c-c-cat if necessary)
<i>Main test</i>	
1 Say <i>rainbow</i>	Say it again, but without <i>bow</i>
2 Say <i>wigwam</i>	Say it again, but without <i>wig</i>
3 Say <i>marmalade</i>	Say it again, but without <i>mar</i>
4 Say <i>dog</i>	Say it again, but just say the first sound
5 Say <i>boat</i>	Say it again, but without the <i>/b/</i>
6 Say <i>stake</i>	Say it again, but without the <i>/st/</i>
7 Say <i>stake</i>	Say it again, but this time without the <i>/s/</i>
8 Say <i>stake</i>	Say it again, but this time without the <i>/k/</i>
9 Say <i>snail</i>	Say it again, but without the <i>/l/</i>
10 Say <i>flag</i>	Say it again, but without the <i>/f/</i>
11 Say <i>glow</i>	Say it again, but without the <i>/l/</i>
12 Say <i>igloo</i>	Say it again, but without the <i>/l/</i>

Appendix VII. Spoonerisms stimuli.

Spoonerisms:

1. *Fat pig*
2. *Teddy bear*
3. *Packed Lunch*
4. *Car park*

Main test:

5. *Zoe ball*
6. *David Beckham*
7. *Darth Vader*
8. *Frank Lampard*

Appendix VIII. Backwards span stimuli.

Practice

<i>Stimuli</i>	<i>Correct answer</i>
9 2	2 9
1 3 4	4 3 1
1 5 6 7	7 6 5 1
{2 6}	6 2 only use this if the child has not got the idea
{4 3 1}	1 3 4 only use this if the child has not got the idea

Main test

4 2	2 4
9 6	6 9
5 3 8	8 3 5
6 7 1	1 7 6
4 3 9 6	6 9 3 4
7 1 8 3	3 8 1 7
3 2 6 1 4	4 1 6 2 3
8 6 4 7 2	2 7 4 6 8
9 6 5 1 7 8	8 7 1 5 6 9
8 7 2 5 1 4	4 1 5 2 7 8
5 7 1 4 6 8 3	3 8 6 4 1 7 5
1 6 9 3 2 8 5	5 8 2 3 9 6 1
7 1 2 3 5 4 8 6	6 8 4 5 3 2 1 7
2 6 7 8 3 5 1 4	4 1 5 3 8 7 6 2

Appendix VIII. Formal operational thought test questions for study 2.

FORMAL OPERATIONAL THOUGHT

NAME TODAY'S DATE

BOY OR GIRL..... CLASS

SCHOOL DATE OF BIRTH.....

The test consists of four sections, please attempt all questions but if you are unable to do any then go on to the next question.

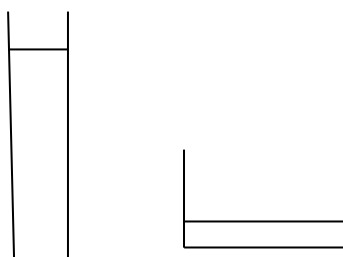
Diagrams are shown to demonstrate the questions but these are not shown to scale.

Please write your answers on this paper in the spaces provided.

SECTION A: VOLUME AND HEAVINESS

1. Examine the diagrams of the two beakers and tick the correct answer.

A is a 100cm³ beaker and X is a 250cm³ beaker. Beaker A is filled with water that is poured into Beaker X. Beaker A is then refilled with water.



Beaker A

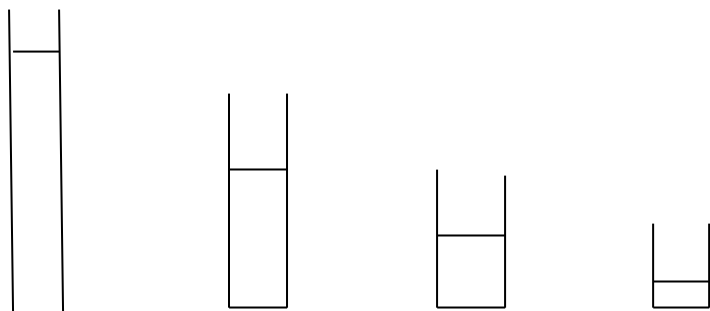
Beaker X

A has more

less

the same amount of water compared with X

2. Examine the diagrams of the beakers and tick the correct answer. Beaker A is filled with water which is then poured into Beaker B. Beaker A is then refilled and poured into Beaker C. Beaker A is then refilled and poured into Beaker D. Beaker A is then refilled with water.



Beaker D Beaker C (500) Beaker B (250) Beaker A (100)

Do these cylinders all have the same amount of water?

YES

NO

If you answered "NO" write down which has most water..... (A/B/C/D)

- (c) The volume of a block of plasticine is 60cm^3 .

How much water will spill over if this block is put fully underwater in a full beaker of water?

.....

- (d) The container shown is full of water.



Level A

Level B

Level C

The block of plasticine is lowered into the container below to level A. Water spills out of the beaker.

- If the block of plasticine is lowered to level B will

More.....

Less.....

The same..... amount of water spill out of the beaker

- If the block of plasticine is lowered to level C will

More.....

Less.....

The same.....amount of water spill out of the beaker

(e) A brass block is the same size as the block of plasticine. The brass block is heavier than the block of plasticine.

- If the brass block is lowered into the beaker of water will

More.....

Less.....

The same.....amount of water spill out of the beaker as when the plasticine block was lowered

SECTION B: BRAINTEASER CHALLENGES

- (a) Archimedes is famous for determining a way to calculate the volume of an object with an irregular shape, for example, a crown. How do you think Archimedes could use large measuring cylinders to compare the volume of two different crowns?



-
-
-
- (b) Archimedes then weighed the two crowns and found that the new, bigger crown weighed more than the old one. Nevertheless he said that the new crown had some lighter metal in it.

How do you think he worked this out?

.....

.....

.....

SECTION C: THOUGHT PROBLEMS

1. 'Sing Right' is a competition to find a new successful recording artist who will sell records and earn money for themselves and the record company that they are signed for. There are two judges, Mark and Ian. Together they have to choose a winner from two remaining acts.

Darren is a 25 year old male singer who Mark thinks has a lot of potential to be successful in the public eye but is unlikely to sell many records once the competition is over. He also thinks that Darren is a handsome man with a good sense of humour. Ian also likes Darren as an act and thinks that he has a lot of charisma.

The second act is a 32 year old female singer called Daisy. Ian prefers Daisy to Darren and thinks that she will sell many records in the future. Mark believes she will sell many records and be successful both during and after the competition. He does not get along

with Daisy though and finds her to be rude and quite arrogant. Ian thinks that Daisy is pretty but that she does not have a sense of humour.

Which do you think is the most likely outcome of the competition?

- (d) The two judges decide that Darren will win the competition
- (e) The two judges decide that Daisy will win the competition
- (f) The two judges will not be able to agree upon a winner

Why have you chosen this answer?

.....

.....

.....

2. An audience watching a ballroom dancing competition final are asked to vote for a winner out of three finalists, Jane, Jamie and Henry. A panel of judges have passed comments on the finalists' performances, although the final decision of who will win is entirely decided by the audiences' vote.

The judges have described Jane as a very talented dancer who technically performs the dance moves almost perfectly nearly all of the time. Her performances have been described as captivating and a joy to watch. She has not been unpopular with the audience, although she has failed to win their hearts fully as she has been consistently good throughout the competition and not shown any improvement or development.

Jamie has been described as the joker of the competition who entertains the audience, although his dances are not so enjoyable to watch as they are technically poor and lack finesse. He is a high spirited and handsome young man who has proven popular in particular with the female members of the judging panel and the audience.

Henry is older than Jamie but younger than Jane. His dances have been described as challenging and technically quite accurate. He has shown improvement over the competitions dances from the start of the competition, although his technical ability is still not up to the same standard as seen in Jane's dances. Henry has somewhat charmed audiences and included elements of comedy in his dances, although judges have described these additions as juvenile and as undermining to the competition itself.

Who do you think is most likely to go on to win the dancing competition?

- d) Jane
- e) Jamie
- f) Henry

Why do you think this person will win the competition?

.....

.....

.....

Who do you think is most likely to come second in the competition?

.....

Why have you chosen this answer?

.....

.....

.....

3. On a planet in outer space there are two species of creatures that have been identified, these are species x and species y.

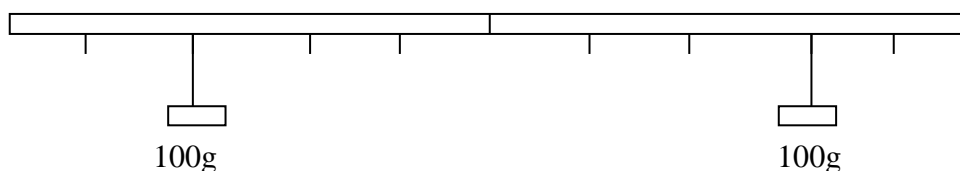
If there are no x's that aren't slim and no y's that aren't x's, then which statements are always true?

- 1. There is not one slim creature that isn't an x
- 2. All y's are slim
- 3. Any creature that is slim is also a y
- 4. None of the above

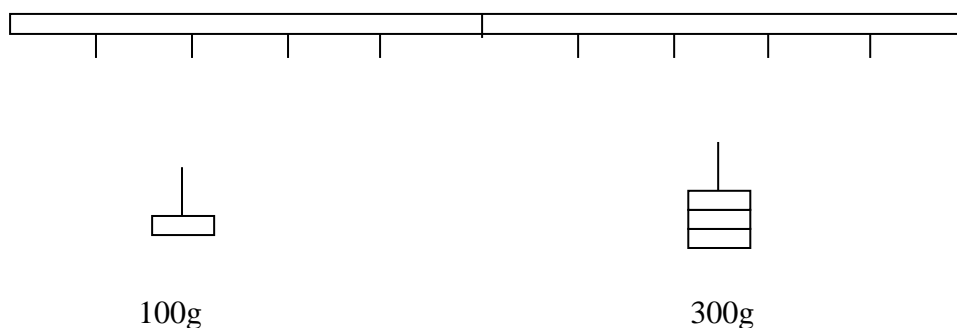
SECTION D: EQUILIBRIUM AND BALANCE

The following questions are based upon the principles of balance and location relating to balancing objects, i.e. much the same ideas as those used when explaining how a see saw works.

For example, on the balance beam below the weights are balanced:



(a) Make the diagram balance by placing the weights on the bar as necessary. Draw on the line below.



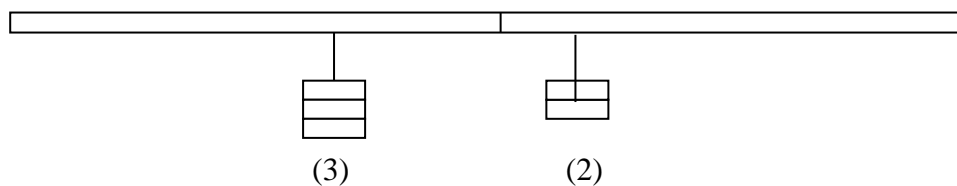
(b) You are given two blocks that are both made of brass.

Block A weighs 60 grams and has volume of 15cm^3 .

Block B weighs 160 grams. What is its volume?

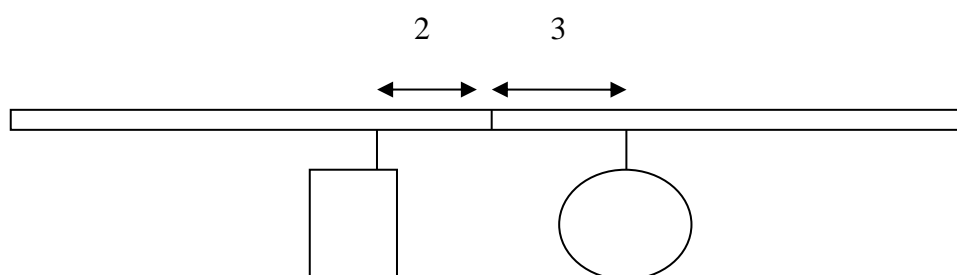
..... cm^3

4. The weights shown are in balance. If the (3) is moved out one unit then how much does the (2) weight have to be moved to retain the balance?



.....units

5. The two weights on the diagram balance.



Which weight is heavier?

How much heavier is it?.....

This is the end of the test.

Thank you for your time. For any further information regarding the test or the results please ask the researcher or contact Sally Mumford (ecq10srm@sheffield.ac.uk).

Section A, B & D tests adapted from 'Thinking Science: The materials of the CASE project' (3rd ed) by Philip Adey, Michael Shayer & C

Appendix X. Information and consent form for study 2.**From: Miss Sally Mumford****Telephone:** 0114 2755921**Email:** ecq10srm@sheffield.ac.uk**Dean of School:** Prof. K. Glaister

Institute of Work Psychology

Management School

University of Sheffield

January 2012

Dear

I am a PhD student at the University of Sheffield (within the Institute of Work Psychology) and I am interested in how individuals 'speak in their heads' and how they think abstractly and logically when solving problems. I think this internal speech and high level thinking can affect how people perform upon tests and am also interested in how these abilities relate to and are affected by the developmental disorder dyslexia. The aim of the research is to establish and investigate links between dyslexia, internal speech and high level thought. In the future this information will be helpful when looking at how people perform upon assessment tasks, for both education and/or careers.

I have developed a simple written test to investigate internal speech and higher level thought abilities. Alongside this I would like participants to complete a short set of tests assessing dyslexia. The session takes around 60 minutes.

I would be very grateful if you could consider the attached form and – if you are willing to participate – email me with your contact information so we can arrange a session.

I will then need you bring the completed form along to the research session with you. It is important for ethical grounds that participants are aware of the study and its aims. Participants should also be aware that they have the right to withdraw at any point during or after the study and that all information will remain confidential. Ethical approval for the study has been gained from the ethics committee within the Management School of the University of Sheffield.

I would like to reassure you that these tests will not cause you any distress and you will be able to withdraw at any time if you wish. I will write up my thesis using the results but all information will remain anonymous and your participation will be much appreciated.

Please feel free to email me (see header above) or my supervisor (Prof Rod Nicolson, r.nicolson@sheffield.ac.uk) if you would like more information.

Yours sincerely

Miss Sally Mumford

Consent form

University of Sheffield project:

Name _____

Preferred contact (email or mobile) _____

.....I have read the attached letter concerning the study investigating internal speech, high level thought and dyslexia and (please tick one!):

I am willing to take part _____

I am not willing to take part _____

Signed _____

Date _____

You will not be able to take part without providing this consent form.



THE QUEEN'S
ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
1998 2000 2002

Appendix XI. Information letter and consent form for study 3.

The
Department
Of
Psychology

Dear Student,

I am a PhD student of Psychology at the University of Sheffield, and I am interested in how cognitive factors and academic profiles are measured through medical selection and assessment techniques. I think that cognitive abilities may be a key measure and may measure a new dimension alongside existing selection methods. I have developed some simple ways of investigating these cognitive abilities and hope that these will aid us with the information I need.

I am working with both the University of Sheffield's Medical School and the Psychology department, and with your consent, I would be very grateful if you could take part in our study. The study involves you answering some multiple choice questions relating to healthcare scenarios; these are very similar to the questions used and the set-up of the Situational Judgement Test (SJT). They take around 20 minutes altogether. Your responses will be automatically received through a computer system as the responses will be entered using the zappers in the lecture theatre.

I would like to reassure you that these questions will not cause you any distress and you will be able to withdraw at any time if they wish. With your consent, I would also like to have permission to gain access to your rank order scores (your academic history of medical school), your UKCAT scores and your actual Situational Judgement Test (SJT) scores. You

will not have to do anything extra for me to access these scores. I will use these scores to look for relationships across university years and within academic years. I will write up the results of the study as part of my thesis, but in all the work the results will be completely anonymous and your answers will be completely confidential.

Please fill in the attached form whether or not you are willing to take part. Thank you very much for your time and your participation would be greatly appreciated! If you would like further information about the study then please feel free to contact me on ecq10srm@sheffield.ac.uk.

Yours sincerely,

Sally Mumford

Consent form:

University of Sheffield project:

Name _____

Please tick only one!

..... I am **not** willing to take part

or

.....I am willing to take part

Signed _____

Date _____