The use of psychometric and other assessment centre measures in predicting performance on a naval command course

Ian William Beadle

Doctor of Education

The University of Sussex

April 2011

DECLARATION

WORK NOT SUBMITTED ELSEWHERE FOR EXAMINATION

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature

I W Beadle

SUMMARY

UNIVERSITY OF SUSSEX

IAN WILLIAM BEADLE   DOCTOR OF EDUCATION

The use of psychometric and other assessment centre measures in predicting performance on a naval command course

SUMMARY

The Admiralty Interview Board (AIB) is the Royal Navy's assessment centre whose role is to select young people for officer training. The two aims of the study were (1) to investigate the relative value of psychometric versus other assessment centre selection measures and (2) the value of these and other approaches for selecting naval commanding officers for practitioners.

The AIB selection data was used to investigate the long-term prediction of some of the selection measures, particularly the psychometric tests, in predicting the outcome for students attending the Submarine Command Course. Few pieces of research have looked at the long-term prediction of a real command situation. This research examines the prediction of a practical naval command situation where the student has to make rapid decisions under pressure and where failure to make the correct decision could be costly. A literature review showed that whilst cognitive tests, personality inventories and other assessment measures can predict job performance and training successes, the meta-analytical techniques used to pool research studies have produced inconsistent findings that could confuse practitioners.

The students attended the command course, on average, thirteen years after the initial AIB selection process. Selection scores were available for 93 students, 57 of whom also completed a 'Big-Five' personality inventory and an Occupational Stress Indicator (OSI) at the start of the 24-week course. The average age of the students starting the course was 32. The students were assessed throughout the course and were graded as pass or fail. They were also given an A to F Course Grade. In addition, 88 students were graded on twenty aspects of performance covering eight tactical grades, three administrative grades and nine personality grades.

The results showed that the means of the pass and fail groups on the AIB Non-verbal test were statistically significantly different with a moderate effect size. The correlation for this relationship (with the A to F Course Grade) was 0.20. The Non-verbal test score also correlated with the course instructor's grades on tactical performance at 0.30 and the Verbal test correlated 0.23 with the administration grades. None of the other AIB selection measures showed significant results.

While this is a disappointing result, the students were a very homogeneous group and to obtain these findings for the Non-verbal and Verbal test after thirteen years shows the predictive power of these tests. Although the findings may be of theoretical interest the low correlations mean that not much variance in performance is explained. The tests would not be a useful screening device to reduce the failure rate on the course because there would be too much misclassification.

None of the Big Five personality scales predicted success on the course or the other course grades but statistically significant differences were found for the means of two the OSI scales: these were for 'Ambition' and a Type A Behaviour measure. Ambition was the only scale which correlated significantly with the A to F Course Grade at 0.43. Ambition was also found to correlate with the total score for the twenty performance grades, the tactical grades and the personality grades and several individual performance grades including Practical Ability, Leadership and Command Presence with correlations approaching 0.4. Further research on this aspect of behaviour may be worthwhile.

However, there are lessons to be learned. The literature review shows that practitioners need to scrutinize journal articles and book chapters on the validity of selection measures extremely carefully. It may be that measures which have been shown to predict the performance of junior staff are inappropriate for the selection of more senior staff with similar job experience.

[594 words]

TABLE OF CONTENTS

LIST OF TABLES Page

Chapter 5

Chapter 6

LIST OF FIGURES   Page

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENT OF THIRD PARTY COPYRIGHT

The author is grateful to the following publishers and other organizations for allowing the reproduction or adaption of their material in this electronic version of the thesis. The specific material is acknowledged in the list below and also at the appropriate points in the main text.

Organization	*Thesis Material*	*Thesis Page*
Academy of Management Review	Figure 1.1	3
American Psychological Association	Table 3.1	34
	Table 3.1	34
British Psychological Society	Table 2.3	14
	Table 2.4	15
	Table 2.5	17
	Figure 3.1	39
	Appendix C	128
Defence Science and Technology Laboratory	Material from: Beadle (1997) and Collis (1997)	
	Table 4.7	58
	Table 4.8	58
	Table F.1	135
	Table F.2	135
	Table F.3	136
Taylor and Francis	Table 2.1	13
	Table 2.7*	28
Wiley-Blackwell, UK	Table 1.1	6
	Table 2.7*	28
John Wiley and Sons, Inc, USA	Table 2.2	13
	Table 2.6	20

* Amalgamated into the same Table.

PREFACE

No part of this thesis results from joint work with any other persons. The design and the methods used in both phases of the research were developed by the author, though these ideas were discussed with the author's line manager, the late Dr Vic Schmit, who was Senior Principal Psychologist and Head of Senior Psychologist (Naval) Branch of the Ministry of Defence, London and subsequently the Technical Manager for Selection and Assessment at the Centre for Human Sciences, Defence Evaluation and Research Agency, Farnborough, Hants.

The sources from which the information presented in this thesis is derived are:

- Books and journals in the University of London Senate House Research Library Psychology Collection (which includes the British Psychological Society Library).
- Electronic journals and books from the Open University Library website.
- Books in the British Library.
- The author of the thesis collected data from the participants using the OCEAN personality inventory and the Occupational Stress Indictor. Data collection took place on Royal Naval premises first in Portsmouth and then later in Plymouth from 1993 to 2002.
- Historic selection data on the participants were extracted by the author from the Royal Navy officer candidates' database in 2006. The author was responsible for the management, upgrading and updating of this database from 1992 to 2000.

1 Introduction

1.1 Background

This chapter starts with the aims of my research, and then explains the background to the assessment and selection process and some of the features and components of the assessment centre process. The use of meta-analysis in researching the predictive validity of assessment and selection measures will be introduced, as well as some of the problems faced by practitioners. The chapter will finish with an outline of the rest of the thesis.

The selection measures discussed in this thesis can be used as individual measures or in combination, but for this research most of the selection measures investigated were those of the selection procedures used at the Admiralty Interview Board. This is the Royal Navy's assessment centre for assessing and selecting young people suitable to be junior officers in the Royal Navy and Royal Marines. Internal research carried out for the Royal Navy (DERA, 1999) into the predictive validity of the Admiralty Interview Board selection methods showed that these assessment centre methods were a valid method of selecting junior officers and can predict job performance and training marks in the first two or three years after selection.

There is no direct entry to senior officer levels in the armed services. All officers must start as junior officers and be promoted from this level. The qualities which make a good junior manager (or junior officer) may not be sufficient to guarantee success at a more senior level. The management writer Drucker (1955) noted that one of the major differences people faced in organizations as they progressed was the change from being someone who had to answer questions to someone who asked the questions. This is a shift from more day-to-day tactical management to a longer-term strategic management approach. If this is the case, then assessment centre methods which predict the short term success of new entrants in training may not be able to predict the longer-term success of senior commanders. The research presented in this thesis will look at how useful these initial assessment centre measures are in predicting success at a later stage in an officer's career in a command situation.

1.2 Aims of the Research

The two aims of the study were (1) to investigate the relative value of psychometric versus other assessment centre selection measures and (2) the value of these and other approaches for selecting naval commanding officers for practitioners.

The research investigated the long-term predictive validity of the procedures of an assessment centre, the Admiralty Interview Board, in predicting the success of students on the Royal Navy Submarine Command Course (SMCC), with a particular interest on the four psychometric tests taken over ten years previously. In addition, the study investigated two other assessment measures, a personality inventory and an occupational stress indicator, which were completed by the students at the beginning of the SMCC Course.

1.3 Assessment and Selection

The research is set within the assessment and selection area of professional practice in occupational psychology which includes the monitoring and validating of selection procedures based on the psychological discipline of psychometrics, the study and application of quantitative methods to measure human attributes such as intelligence, aptitude and personality. The word "psychometric" means psychological measurement, whose analysis and interpretation has been developing for over 130 years.

Viswesvaran and Ones (2010) describe personnel selection as 'one way of ensuring that employees have the requisite characteristics, knowledge and skills to perform the work they are hired to do' (p. 170). Assessment and selection are parts of the human resource management process and assessment and selection link to other human resource processes. Figure 1.1 adapted from Von Glinow *et al.* (1983) shows the major linkages between the two boxes labelled 'assessment' and 'selection and placement' and the other parts of the system. Von Glinow *et al.*'s diagram also emphasises how these processes follow on from, and also feed back into, both the human resource strategies and the overall strategy of the organization through 'evaluation'. Legge (1975) has suggested that the integration and internal consistency of the human resource systems is important for the success of an organization and that assessment and selection can be seen as an important first stage in this process.

Iles (1999) has outlined four approaches which selection and assessment can take: the strategic management approach, the psychometric approach, the social approach and the critical discourse perspective. Although this research project is set within the psychometric approach, the other three approaches will be used when appropriate since the work carried out by occupational psychologists and other professionals goes beyond an academic approach to help managers and organizations to function better. In

summarising the strategic approach, Iles pointed out the important role of assessment and selection in organizational development:

> … leading organizations in the private and public sector have realised the critical strategic role of selection and assessment processes and revamped their strategies and practices in the light of environmental changes. In this way, it is argued, assessment and selection processes can not only assist in the selection decisions but can also assist in the selection of development activities, and help in the appraisal of potential (p. 1).

In the armed services where the training of junior officers is expensive, and where senior officers cannot be bought in from outside the organization, the function of initial assessment and selection is extremely important.

Figure 1.1: Internal integration of human resource systems (adapted from Von Glinow et al., 1983, p. 26). [Adapted and reproduced with permission from the Academy of Management]
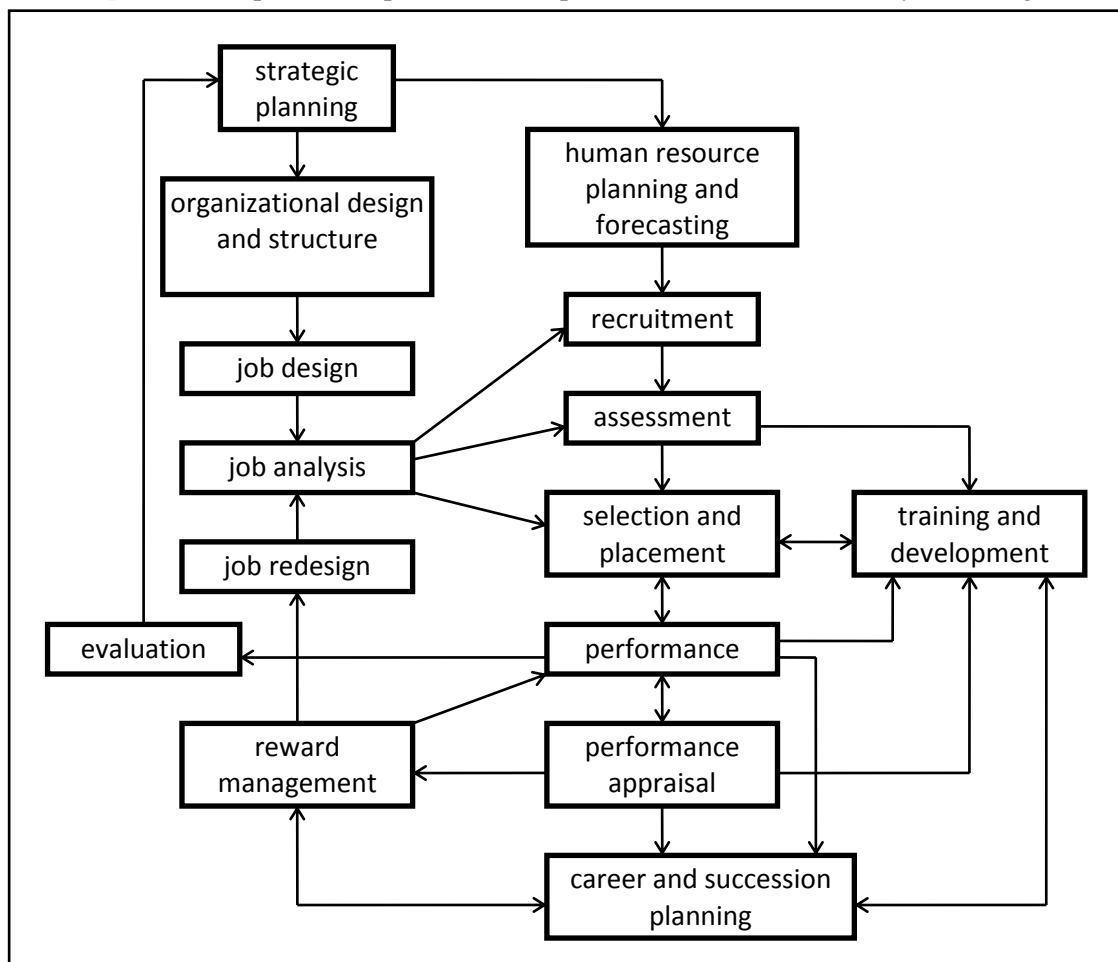
1.4 Assessment Centres

Arthur and Day (2011) note that assessment centres are a method; there is no single assessment centre. Assessment centres are adapted by a particular organization for the specific purpose of assessing and selecting their employees. Assessment centres are

usually held inside buildings[1] but the term does not apply to the building (like leisure centre or arts centre) but to the assessment and selection process itself. Jansen and de Jongh (1997, p. xiii) note that: 'An Assessment Centre is an evaluation process which can be used to identify the potential of employees and job candidates for a broad range of functions'. Seegers (1997) identified six characteristics which typify the range of assessment centres. These are:

1. That behaviour exhibited by the candidates at an assessment centre can be used to predict future behaviour.
2. Carefully developed criteria, based on a thorough job analysis, are used to assess candidates.
3. The exercises candidates undertake are geared towards demands of the future job they will do if successful.
4. Group exercises can be used to observe and record the way candidates deal with each other.
5. Two or more assessors are employed in the assessment process, who are (preferably) managers senior in position to the candidates.
6. The final result of the assessment centre is based on the outcome of the various exercises undertaken by the candidate.

Whilst the basic ideas outlined by Seegers (1997) apply to many assessment centres, those assessment centres looking for potential in candidates, say for management or leadership, would probably not use criteria developed through a specific job analysis, because in large organizations there would be many jobs involved. Instead, the criteria would be developed with the range of jobs in mind, but the criteria would be focussed on higher order factors such as effective intelligence, problem solving skills, communication skills and leadership potential. This is the start of an iterative process in which the data gathered on the particular assessment centre is examined to see what the predictive validity of the various procedures are on subsequent outcomes for the candidate such as job performance and training success. From this analysis the assessment centre methods are modified or replaced. This process can only be achieved with reasonably large numbers of candidates and entrants, so people running small, infrequent or one-off assessment centres rely on the literature to design and develop their assessment centre processes.

Seegers (1997) summarises his article on assessment centres by writing that:

> … the Assessment Centre method is both scientifically justified and practically applicable. The method is not typically American, nor is it a passing trend, but a very useful method of bringing long-awaited changes into the personnel arena (p. 17).

---

[1] The UK army's officer selection centre, The Regular Commissions Board, at Westbury in Wiltshire has many practical tasks which take place outside in the grounds in all weathers.

Both public and private sector organizations have adopted the use of assessment centres to select people.

Assessment centres are seen by organizations as a good way to try to ensure that the selection system is as fair as possible. Whilst Employment Tribunals have found cases of individual selection methods which have been considered to be unfair, the author is not aware of any which have tested the whole concept of the assessment centre. In addition, research by Anderson and Goltsi (2006) on the negative psychological effects of selection did not find any evidence of negative psychological effects for candidates who fail the assessment procedure.

On the other hand, Roe (2005) notes that, while the psychometric approach has a lot to offer, it has its limitations. This approach can ignore the context in which the assessment and selection takes place and the different organizational stakeholders who have in interest in the assessment centre outcomes. Additionally, these authors note, the whole design and operation of the assessment centre can affect the outcomes. Woodruffe (2005) notes that emotional factors can play a part for candidates attending assessment centres who are supposed to be on their best behaviour and their reaction to the artificial circumstances can be wrongly interpreted.

1.5 Typical Components of an Assessment Centre

The features of the assessment centre today are not far removed from the original assessment centres of the early forties, though exercises which have proved to be less valid have been replaced by others. There has also been a growth in the use of personality tests. The typical assessment centre format can be illustrated by looking at The Admiralty Interview Board (AIB) which will be described later. The exception to this is the practical group Gym task at the AIB where candidates use equipment like planks and ropes to tackle several practical scenarios. For a comparison, the details of the procedures at the Civil Service Selection Board can be found in Fletcher (2005) who also summarises selection practices in the private sector. Silvester and Dykes (2005) describe the recent use of an assessment centre procedure to select political party candidates in the United Kingdom.

Robertson and Smith (1989) report on a survey they conducted on what components make up an assessment centre. Their list, which includes both individual and group components, is given in Table 1.1. Those which will feature in the research presented in this thesis are asterisked. Note that not all these methods would be used, but a selection

made from these components. A typical assessment centre would use some of the appropriate items numbered 1 to 3 and References (numbered 8).

Table 1.1: Components of an assessment centre. [Adapted and reproduced with permission from Wiley-Blackwell]

1. Interviews • unstructured • structured • situational • behaviour description 2. Tests (analytical or signs) • cognitive ability* • perceptual motor • personality* • interest 3. Tests (analogous or samples) • work samples • situational* (in-trays, role plays, simulations) • trainability tests	4. Computer Aided Tests 5. Repertory grid 6. Biodata* and Accomplishment records 7. Future Autobiography 8. References* 9. Graphology 10. Astrology 11. Self-assessment 12. Supervisor/Peer assessment
Source: Robertson and Smith (1989) p. 90. Note: * = types of assessment featured in the research presented in this thesis.	

Robertson and Smith note that, except for the repertory grid technique and future autobiography ('Where do you see yourself in ten years' time?'), little had changed in the last twenty years. A review of the recent articles on assessment centres in the occupational psychology practitioner publication *Selection and Development Review* shows that little has changed since Robertson and Smith's article was written more than twenty years ago. In practice, these individual components can be combined. For example, individual candidates may carry out an in-tray exercise before being brought together as a group to discuss their work and produce a joint proposal; or after some group discussion task the individuals may be required to write up the group solution to the problem on their own or give an individual presentation of the group's decision to some assessors. Hough and Dilchert (2010) explain that high administrative costs have limited the use of assessment centres to occupations in which the performance variation in monetary terms is large, for example, managerial positions and high risk jobs, but the building blocks of the assessment centre, the exercises, can still be used individually to assess personal characteristics.

The literature review in the next chapter shows how psychometric tests can help select the right people for jobs since the tests can predict the level of performance in the job a

couple of years after selection, as well as indicate success in initial training. The research presented here gives a unique opportunity to examine if these psychometric measures can predict performance in a command situation ten or more years after the tests were taken. In addition, the investigation of whether the scores on a personality inventory and a workplace stress indicator can tell us about the likely performance of students is a second worthwhile investigation. Most research on this subject has been conducted using static performance measures like salary or annual reports. Some research has used simulations devised to train or assess candidates rather than real pressurised decision-making tasks.

1.6 The Use of Meta-Analysis in Personnel Selection Research

Many of the research studies presented later in this thesis rely on the technique of meta-analysis which in turn relies on techniques to correct for restriction of range. Murphy (2003) notes that: 'The term *meta-analysis* refers to a wide array of statistical methods that are applied to the outcomes of multiple studies to describe in some sensible fashion what these studies have typically found, and draw inferences about what those findings might mean' (p. 3; original emphasis). Meta-analysis became a popular tool for researchers in the late 1970s and early 1980s and a flurry of articles using this technique appeared in journals. Schmidt and Hunter (1977) pioneered the meta-analytic methods of combining the predictive validity coefficients from multiple studies to estimate the overall validity for a wide variety of tests and selection procedures. They noted that most of the observed differences across cognitive ability studies were due to sampling error where the average size of samples was about 70. This 'validity generalization' is a special case of meta-analysis applied to validity studies. These methods can be seen to be analogous to the reduction in variance, often achieved in research which uses stratified random sampling rather than pure random sampling. The combination of many studies, each with a homogenous sample, ought to reduce the variance in the combined sample and so give a better estimate of the likely predictive ability of the particular method.

Schmidt and Hunter (2002) note that small samples give contradictory results due to the distorting effects of sampling and measurement error, which occur in all studies, even if attempts have been made to control these factors. As a result 'meta-analysis is needed to integrate the findings across studies to reveal the simpler patterns of relations that underlie research literatures through providing a basis for theory development' (p. 51). Drasgow (2003) notes that 'This literature was so vast and the effects of sampling variability so pernicious that the findings were essentially incomprehensible until

statistical methods for aggregation were introduced by Frank Schmidt and John Hunter in 1977' (p. 122).

Hunter and Hunter (1984) noted that information gathered using meta-analytic techniques can be used to produce accurate utility analysis figures to show the costs and benefits of using cognitive tests or alternative methods. Their analysis shows that these tests save many billions of dollars each year and even validity coefficients of 0.1 are worth considering because of the savings these tests can make.

1.7 Practical Aspects of Personnel Selection

Most practitioners only assess a small number of people at a time. Whilst the practitioner may attempt to validate their assessment methods, the small size of their sample may not give them any certainty about the results. Additionally, entrants who leave the organization before they have been assessed make this validation an even more difficult task. The practitioner who is asked to devise a new assessment procedure for an unfamiliar job would carry out a job analysis and produce a matrix of job components against suitable assessment methods. Arthur and Day (2011) note the importance of devising this type of matrix. The ability to produce a matrix and suggest the most appropriate methods of assessment relies on the use of the literature to guide the practitioner about which selection methods would be best suited to predict which criteria. An example of a job criteria and exercise matrix is given in Table 1.2. Here an estimate is made of the likely strength of prediction for each cell in the matrix. Organizations have limited funds and a complete assessment of each candidate using all the methods may be expensive, not only for the organization but also for the candidate if they have to take several days off work. So in practice, a few of these methods would be selected, based on their contribution to assessing the particular knowledge, skills and attitudes required for the job.

Table 1.2: Job criteria and exercise matrix.

	Assessment exercises				
Job criteria	*Cognitive tests*	*Personality inventory*	*Group exercise*	*In-tray exercise*	*Structured interview*
Reasoning ability	***		**	***	*
Team-working		**	***		*
Effective communication	*	**	*** (oral)	*** (written)	***
Organizational skills			**	***	*
Leadership skills		**	***	*	*
Note: *** strong predictor ** moderate predictor * weak predictor					

The meta-analytic literature helps the practitioner to decide how much reliance to place on the alternative methods. For example, will a group exercise provide better evidence of a candidate's team-working ability than a personality inventory? The practitioner also has to advise the assessors and selectors on how to weigh the various assessment information which is produced by the different methods when they come to make a final decision about a candidate. Here, again, individual studies in the literature may be based on a small number of candidates and not cover appropriate groups, and the results of these separate studies may be contradictory. Murphy (1997) notes that 'There are substantial gaps between research and application in areas such as personnel selection and assessment ...; meta-analysis provides one set of tools that can be used in closing these gaps' (p. 32). The results of meta-analytic research indicate the sort of correlations that practitioners might expect to find. These findings are particularly helpful to staff running smaller, infrequent or one-off assessment procedures. They help the practitioner to know which of the processes are most likely to be best at predicting success in a particular job.

Searle (2003) notes that in the past twenty-five years there has been an expansion in the use of psychometric tests by organizations for recruitment and selection. She gives two main reasons for this growth. The first reason is that these methods of assessment and selection were adopted by large, well-known companies and public sector organizations. This gave the methods credibility with other organizations who adopted them. The second, and associated reason, is that organizations have felt more confident about using these methods because of improved information on their predictive validity. After a spate of legal challenges, particularly in the United States, and criticisms about the fairness and robustness of tests and other aspects of the assessment process, the use of meta-analytic research has shown that the picture was more positive. This has resulted in the conformation and enhanced standing of many of the tools as valid and reliable selection procedures. Searle comments on this: '... validity studies, particularly meta-analysis-based ones, have played an important role in improving the credibility and professionalism of human resource practices and applied psychology' (p. 65). As Searle notes, meta-analytic studies also give the practitioner some protection, for instance in Employment Tribunals, if the person being assessed is not happy with a particular method used for selection.

1.8 Thesis Structure

This introductory chapter is followed by the conventional chapters expected in this type of thesis.

Chapter 2: Psychometrics and other Assessment Centre Methods in the Selection of Personnel. In this chapter a review of some of the published research which underpins both the theoretical and practical aspects of selection and assessment will be presented.

Chapter 3: Selection in the Military. In this chapter the focus will be on selection research in the military context. The end result of military training is to enable personnel to work in stressful situations so the issue of stress will be discussed.

Chapter 4: Context and Methods. This chapter will look at how this study addresses the theoretical questions. Then the settings for the data collection, the Admiralty Interview Board and the Submarine Command Course, will be described. Details of the methods used to gather the data, the assessment centre methods such as the psychometric tests, as well as the concurrent measures, the personality measure and Occupational Stress Indicator will be described. An explanation of the type of analysis to be undertaken will also be given. The limitations of the research will also be presented.

Chapter 5: Results. The results of the analysis will be presented here.

Chapter 6: Discussion. The results of the analysis will be discussed in terms of the two aims of the study followed by a conclusion. This is followed by a list of References and Appendices.

1.9 Chapter Summary

This chapter has set out the aims of the research which was to investigate the relative value of psychometric versus other assessment centre selection measures, and the value of these and other approaches for selecting naval commanding officers from a practitioner's point of view. The research is set in the psychometric tradition of assessment and selection which is a central starting point to many human resource management functions. The assessment centre process is regarded as a fair method of selection, but the practitioner needs to rely on validity studies based on meta-analysis to aid their understanding of which assessment methods to use. The next chapter presents the literature on the validity of psychometric and other assessment centre measures in predicting job performance and training success.

2 Psychometric and other Assessment Centre Methods

2.1 Introduction

This chapter will report the literature on the relative value of psychometric and other assessment centre measures used for assessing and selecting personnel. In particular, the predictive validity of psychometric tests and personality inventories will be discussed. The research presented in these sections comes mainly from those studies that are most often quoted in summaries of the literature in respected journals, book chapters or books read by practitioners, who have to devise assessment and selection procedures using these methods. Most of the predictive validity studies outlined below examine prediction of both job performance and training success which are the most common way to validate these procedures.

The Submarine Command Course is a qualifying (training) course for those who wish to command submarines and carry out other duties of senior officers before taking command. The course is used to deselect those *not considered suitable* for these roles, so this literature review covers both job performance and training outcomes. Although comments will be made about particular components of an assessment centre, they are not intended to be comprehensive, but rather to investigate the methods most relevant to the procedures at the Admiralty Interview Board which form the basis for this research.

2.2 Psychometric Tests which Measure Cognitive Ability

Appendix A (p. 121) reviews the background to tests of cognitive ability. Some recent summaries note the value of these tests for selection purposes. For example, Drasgow (2003) reviewed the role of intelligence and workplace performance, particularly the two important measures of job performance and training proficiency, and concluded that 'A large and compelling literature shows that intelligence predicts these two important classes of criterion variables' (p. 108). Ones *et al.* (2005) in another influential review of the literature noted that 'The overwhelming evidence suggests that CA [cognitive ability] tests are predictive of job performance across jobs and cultures' [online]. In the case of the research presented in this thesis, where job knowledge is likely to be important, Ones *et al.* (2010) note that the acquisition of job knowledge is linked to cognitive ability: 'The more complex jobs are, the more complex and vast the knowledge to be acquired' (p. 261).

2.3 Tests of Cognitive Ability as Predictors of Job Performance and Training Success

Ghiselli (1973) appears to have been the first person to conduct large scale research on the predictive validity of selection measures with job performance and training success. He used correlational data from hundreds of studies conducted between 1920 and 1971, both published and unpublished, but he did not use a validity generalization technique; rather, he combined the studies by averaging the weighted coefficients using the median measure rather than the mean. The range of tests covered twenty areas such as intellectual abilities, spatial and mechanical abilities, perceptual accuracy, motor abilities and personality inventories. Ghiselli found that the average correlation was 0.22 for job performance criteria and was 0.39 for training success across the wide range of types of assessment. He also noted that for every type of job, there was a test which was moderately predictive. He found that if you took the highest average validity coefficient for the 20 types of test for the 21 jobs he examined, the values ranged from 0.24 to 0.46 for job performance and from 0.28 to 0.66 for training. The averages of these validity coefficients were 0.35 and 0.45, respectively. A very important point made by Ghiselli is that, because of the differences in job requirements between jobs, a validity study would be required in the particular job setting before the predictor could be recommended for selection. Ghiselli concluded his paper by noting that the results presented were for single tests 'and that judiciously selected combinations of tests would have been [*sic*] higher validity' (p. 477).

In a meta-analysis using Ghiselli's data plus studies which had been conducted since his review, Hunter and Hunter (1984) looked at the predictive validity of selection methods and job performance with a particular interest in the alternatives to cognitive ability tests. They found that cognitive ability tests had a mean validity of about 0.55 across a wide range of jobs and that 'There is no job for which cognitive ability does not predict training success' (p. 80). Hunter and Hunter also found that for entry-level jobs cognitive ability tests had a higher predictive validity than any alternative method.

The two articles presented above reviewed validity studies carried out in the United States. Herriot and Anderson (1997) queried why no similar meta-analyses of European-based studies had been carried out. Salgado and Anderson (2002) noted that this lack of meta-analytic studies in Europe would surprise their United States colleagues, particularly since these authors found that cognitive testing was used more often in Europe than in the United States. Using forty-five validity studies from the UK and nine from Spain, with a total sample size of over eight thousand cases, Salgado and

Anderson found that cognitive ability tests correlated 0.41 with performance measures for the UK and 0.61 for the Spanish research studies. The uncorrected figures were 0.18 and 0.36, respectively. The combined figure was 0.42 (0.21 uncorrected) which is similar to the validity coefficients found in the United States studies. A summary of these results is shown in Table 2.1.

Table 2.1: Meta-analyses of cognitive ability validity studies from the UK and Spain (adapted from a table in Salgado and Anderson, 2002, p. 87). [Adapted and reproduced with permission from Taylor and Francis]

Criterion	Country	Number of Studies	Total Sample Size	Correlation	Corrected Correlation
Job Performance	UK	45	7,283	0.18	0.41
	Spain	9	1,239	0.36	0.61
	Combined	54	8,522	0.21	0.42
Training Success	UK	61	20,305	0.34	0.56
	Spain	25	2,405	0.35	0.47
	Combined	86	22,710	0.34	0.53

Salgado and Anderson also investigated the validity of these measures for predicting training success (also shown in Table 2.1) using sixty-one UK studies and twenty-five Spanish studies (a total sample of nearly 23,000 people). The results show corrected correlations of 0.56 for the UK research and 0.47 for the Spanish studies, respectively, and a correlation of 0.53 for the combined studies. The uncorrected correlations are 0.34, 0.35, and 0.34, respectively.

Table 2.2: Meta-analyses of cognitive ability validity studies from European studies (adapted from two tables in Salgado et al., 2003, pp. 586 and 590). [Adapted and reproduced with permission from John Wiley & Sons, Inc.]

Criterion	Type of Test	Number of Studies	Total Sample Size	Correlation	Corrected Correlation
Job Performance (120 samples)	General	93	9,554	0.29	0.62
	Verbal	44	4,781	0.16	0.35
	Numerical	48	5,241	0.24	0.52
	Spatial-Mechanical	40	3,750	0.23	0.51
	Perceptual	38	3,789	0.24	0.52
	Memory	14	946	0.26	0.56
Training Success (142 samples)	General	97	16,065	0.28	0.54
	Verbal	58	11,123	0.23	0.44
	Numerical	58	10,860	0.25	0.48
	Spatial-Mechanical	84	15,834	0.20	0.40
	Perceptual	17	3,935	0.13	0.25
	Memory	15	3,323	0.17	0.34

To add to this European research Salgado *et al.* (2003) took the work a stage further by combining and analysing 234 independent samples from studies published across the European Community. The resulting analysis showed overall corrected validity coefficient figures of 0.52 for job performance ratings and 0.56 for training success

criteria. Table 2.2 shows the individual results for the different types of cognitive test. The underlying, uncorrected correlations for the various types of cognitive measure and job performance were between 0.16 and 0.29 (the corrected values were between 0.35 and 0.62). For training success the uncorrected values were between 0.13 and 0.28 (between 0.25 and 0.54, corrected). In both cases the tests of general ability have the highest correlations.

A more recent meta-analysis has been presented by three of the authors of the previous study cited. Bertua *et al.* (2005) reported on the predictive validity of UK cognitive ability tests and job performance and training success. Their results are displayed in Table 2.3 and 2.4.

Table 2.3: Meta-analyses of UK cognitive ability validity studies (adapted from two tables in Bertua et al., 2005, pp. 395-396). [Adapted and reproduced with permission from the British Psychological Society]

Criterion	Type of Test	Number of Studies	Total Sample Size	Correlation	Corrected Correlation
Job Performance (283 samples)	General	12	2,469	0.22	0.48
	Verbal	14	3,464	0.17	0.39
	Numerical	20	3,410	0.19	0.42
	Spatial	7	1,951	0.15	0.35
	Perceptual*	7	1,968	0.23	0.50
Training Success (223 samples)	General	53	17,982	0.29	0.50
	Verbal	33	12,679	0.29	0.49
	Numerical	46	15,925	0.32	0.54
	Spatial	50	15,591	0.24	0.42
	Perceptual*	41	13,134	0.30	0.50
*Labelled perceptual-clerical by Bertua *et al.*					

Table 2.3 shows the validity coefficients for different sorts of cognitive tests. For job performance the coefficients range from 0.15 to 0.23 for job performance (0.35 to 0.50, corrected) and from 0.24 to 0.32 for training success (0.42 to 0.50, corrected). The perceptual-clerical and the general ability tests show the highest correlations for job performance whilst the numerical and perceptual-clerical have the highest correlations for training success. In Table 2.4, the validity coefficients are given for different types of occupations. Here, for job performance the coefficients range from 0.14 to 0.36 (0.32 to 0.74, corrected) and from 0.28 to 0.39 for training success (0.47 to 0.64, corrected). The highest correlation found for job performance was for professional, engineering and management occupations groups, while for training success all but one of the eight professional groups had validity coefficients over 0.30, uncorrected.

Table 2.4: Meta-analyses of UK cognitive ability validity studies for various occupations (adapted from two tables in Bertua et al., 2005, pp. 398-399). [Adapted and reproduced with permission from the British Psychological Society]

Criterion	Occupation	Number of Studies	Total Sample Size	Correlation	Corrected Correlation
Job Performance	Clerical	5	628	.14	.32
	Engineer	5	542	.33	.70
	Professional	4	348	.36	.74
	Driver	2	293	.16	.37
	Operator	9	3,105	.24	.53
	Manager	5	302	.33	.69
	Sales	6	483	.25	.55
	Miscellaneous	7	943	.18	.40
Training Success	Clerical	8	1989	.33	.55
	Engineer	5	1381	.39	.64
	Professional	3	295	.35	.59
	Driver	3	1674	.28	.47
	Operator	17	4322	.32	.54
	Skilled	12	3086	.33	.55
	Miscellaneous	14	7258	.33	.55

In their original meta-analytical review of this type of research in 1981 Schmidt and Hunter concluded that 'The substantive message is … professionally developed cognitive ability tests are valid predictors of performance on the job and in training for all jobs' (p. 1128). It appears that their assertion is still applicable.

2.4 How Relevant is the Historical Data used in Meta-Analysis?

The meta-analytic studies contain data from readily available published research. However, although the appendix of the Bertua *et al.* article lists the tests used in their meta-analysis, it does not reveal which of the tests provided which set of data. Some of the research findings were carried out early in the Second World War by researchers such as Vernon and Parry (1949); but there is no way of knowing what proportion of the meta-analysis samples came from this mid-twentieth century research. One issue is that during the Second World War and during the two decades after this psychologists were often trying to select academically unqualified but intelligent men and women for jobs which needed a range of abilities and aptitudes. The 1944 Education Act raised the school leaving age in the UK from fourteen to fifteen. The extreme case of this selection process during the Second World War occurred when totally unqualified, unskilled workers were screened for officer selection. This is almost analogous to selecting people from a random sample. With the second raising of the leaving age to sixteen in 1972 more young people became educationally qualified, and these academic qualifications were increasingly used as the first stage of selection. So it is likely that validity studies carried out from this time to the present day do not contain anything like

the wide range of abilities and aptitudes found in many of those early samples. Present day samples may suffer from a restriction of range not found in those earlier studies. The figures presented in these meta-analysis tables may be the theoretical maximum validity coefficients of a random sample technique, and these correlations should not be expected in any practical selection situation. However, these estimates do give the practitioner an idea of the relative merits of the various types of test available and the sorts of jobs for which particular tests can be recommended.

2.5 Personality Theories and Inventories

Hampson (1999) notes that defining personality is tricky:

> Personality is impossible to define succinctly because it means different things to different personality psychologists. Whereas most would accept that the field of personality is the study of how individuals differ from one another, they would disagree on the best way to conceptualise these individual differences (p. 284).

Hough and Dilchert (2010) note that 'Personality variables have had a roller-coaster ride in employee selection during the 20th and 21st centuries. They have been denounced and rescued several times over' (p. 299). The development of personality theories is outlined in Appendix B (p. 124). This Appendix summarises the work of the two most influential personality theorists of the mid Twentieth Century: Catell and Eysenck. In the 1980s when more use was being made of personality tests for real life applications, like assessment and selection, many users felt that the number of dimensions used by both Eysenck (not enough) and Cattell (too many) was unhelpful. Tett *et al.* (1991) explained that whilst Cattell's 16 factors are too complex, Eysenck's three-factor model is too broad to have serious predictive validity.

How many personality factors are necessary? Zuckerman *et al.* (1993) note that the different tests are equally valid alternatives that depend on the level of description necessary. Three factors might be limited, whilst sixteen factors are probably too many for untrained assessors to appreciate. Even when interpreted by a properly trained practitioner, this number of factors is probably too many for assessment centre selectors to handle. Five factors, on the other hand, can be comprehended more easily. In the late 1980s and early 1990s several theorists (Costa and McCrae, 1988; Digman, 1990; Goldberg, 1993) proposed a five-factor model of personality[2] based on factor-analytic studies. This soon became known as 'the Big Five' or the 'Five Factor Model'[3] and has

---

[2] Goldberg (1993) credits Thurstone (1934) as being the first to produce five factors.
[3] Strictly speaking, the Big Five refers to inventories developed from Allport and Odbert's (1936) trait list, whereas the Five Factor Model was developed from questionnaire items, but the terms will be used as equivalent in this thesis.

become increasingly popular in both selection practice and research. The Big Five dimensions are Extraversion (A), Neuroticism (N), Conscientiousness (C), Agreeableness (A) and Openness to Experience (O). These five dimensions can be remembered by the acronyms OCEAN or CANOE.

Hampson (1999, p. 285) noted that 'the winning number in this lottery is undoubtedly five', but also notes that 'one major advantage of the Big Five framework is that it can assimilate other structures'. Barrick and Mount (1991) note that the two dimensions on which there is most agreement, Extraversion and Neuroticism (or Emotional Stability) 'represent the 'Big Two'[4] described by Eysenck over forty years ago' (p. 4). Eysenck himself (1992) noted that Agreeableness and Conscientiousness could be linked negatively to Psychoticism and that Openness was a factor of Extraversion. Commercially, the Cattell sixteen factors personality inventory is now available as a Big Five version, the 16PF5, where the five factors are calculated from weighted combinations of the original 16 factor scores (Russell and Karol, 1994).

Hampson (1999) summarises the Big Five trait descriptions and these are shown in Table 2.5. Some of the five factor dimensions of different inventories are not equivalent even though they use the same dimension label.

Table 2.5: The Big Five personality domains and representative traits (adapted from Hampson, 1999, p. 285). [Adapted and reproduced with permission from the British Psychological Society]

Domain	Desirable traits	Undesirable traits
Intellect or **O**penness	imaginative, intelligent, creative	shallow, unsophisticated, imperceptive
Conscientiousness	organized, thorough, tidy	careless, unreliable, sloppy
Extraversion	outgoing, sociable, assertive	introverted, reserved, passive
Agreeableness	kind, trusting, warm	hostile, selfish, cold
Emotional stability [**N**euroticism]	calm, even-tempered, imperturbable	moody, temperamental, nervous

Hough and Dilchert (2010) note that 'today the Five Factor Model (FFM) is the most widely accepted structure of personality variables' (p. 299). The names of these five factors are very similar to those found by Norman (1969) in the mid-1960s but with the exception of 'Culture' rather than 'Openness'. Norman had reworked data from Tupes

---

[4] Seven years later, Salgado (1998) calls Emotional Stability (Neuroticism) and Conscientiousness the 'Big Two'.

and Christal (1961, republished in 1992) who had devised a five factor model of personality, but this work was not published outside the US military[5].

2.6 Personality Inventories as Predictors of Job Performance and Training Success

The ability of personality to shed light on job performance and training success is the most controversial of all the topics covered in this thesis. Guion and Gottier (1965) conducted a review of personality research carried out between 1952 and 1962 and summarised their findings by writing that 'It is difficult in the face of this summary to advocate, with a clear conscience, the use of personality measures in most situations as a basis for making employment decisions about people' (pp. 703-704). Much of the published meta-analytic research on this topic since 1965 has endeavoured to counter Guion and Gottier's view and to support the assertion that personality dimensions can be useful predictors of workplace performance.

The results of Ghiselli's (1973) meta-analysis on the predictive validity of various types of ability and aptitude test were reported earlier in this chapter. Ghiselli also investigated and reported on the contribution of personality and interest inventories in predicting job performance and training success. In his introduction to personality measures he reports one of the problems with personality tests: construct validity.

> A number of different trait names are used to distinguish the various aspects of personality. In some cases different names are used to denote the same, or very nearly the same, quality and in others the same name is used to denote quite different qualities (p. 464).

It was noted earlier that this problem still exists with the more specific five factor models.

Ghiselli looked at the predictive validity of personality measures for different occupational groups. Most of the higher correlations found were between 0.2 and 0.3, and these are uncorrected. So, unlike the results for cognitive tests where Ghiselli found a test with moderate validity for every type of job, these results were relatively disappointing.

Barrick and Mount (1991) reported meta-analysis results for the Big Five personality factors and their relationship to job performance, training proficiency and other personnel data. Like Ghiselli, they looked at the predictive ability of personality tests for various occupational groups (professionals, police, managers, sales, and skilled and semi-skilled workers). Barrick and Mount hypothesized that Conscientiousness and Emotional Stability would be valid predictors of job performance for all jobs. They

---

[5] The OCEAN personality inventory derived from this work is used for the research in this thesis.

were particularly sure that Conscientiousness would be predictive since it is linked to aspects such as hard work, planning, responsibility, and so on, 'which are important attributes for accomplishing work tasks in all jobs' (p.5) as well as for enabling the person to succeed in training. Conversely, those employees who report that they are on the negative side of emotional stability, Neuroticism, with traits such as worrying and nervous 'will tend to be less successful than more emotionally stable individuals in all occupations studied because these traits tend to inhibit rather than facilitate the accomplishment of work tasks' (p. 5). They also proposed that Extraversion and Agreeableness would be a useful predictor in occupations such as management and sales, but less useful in more technical areas such as engineering. Finally, they noted that Openness to Experience should be a good predictor of training success because individuals with higher scores on this dimension would be keener to learn.

Barrick and Mount gathered together data from 162 samples reported in 117 acceptable studies with a total sample size of nearly 24,000 people. The main result they found was that Conscientiousness produced consistent results in predicting performance for all occupational groups, but for the other dimensions the predictive validity depended on the type of criteria and the particular occupational group. Extraversion, for example, predicted performance across all criteria in sales and management jobs. Barrick and Mount explain that this is probably because sales and management jobs require social interaction. They also found that Openness and Extraversion gave some prediction of training success across the different groups. Agreeableness was not found to be an important predictive dimension even in the field of sales and marketing. Neuroticism/Stability had little predictive success though, interestingly, for professionals there was a low negative correlation. This result could be explained by hypothesizing that professionals, who were slightly worrying and neurotic in a homogeneous group, might pay more attention to details, double-check things, be better prepared and also turn up on time. These dimensions showed prediction for some other occupations but the correlations were often less than 0.1 even when corrected.

Table 2.6 shows Barrick and Mount's results for different occupational groups across all three criteria of performance, training and personnel data. The separate tables in their article which present results for the separate criteria are fairly similar in the coefficients found. However, only Conscientiousness had non-zero correlations with job performance across all occupational groups. The corrected correlations for Conscientiousness were between 0.20 and 0.23 for the three criteria (based on observed correlations of 0.11 and 0.13) and the correlations between Extraversion and

performance in management and a sales occupations were 0.18 and 0.15, respectively (based on observed correlations of 0.11 and 0.09).

Table 2.6: Validity coefficients of the Big Five for different occupational groups (adapted from Barrick and Mount, 1991, p. 13). [Adapted and reproduced with permission from John Wiley & Sons, Inc.]

Occupational group	Validity coefficients [corrected shown in square brackets]				
	O	C	E	A	N
Professional	-.05 [-.08]	.11 [.20]*	-.05 [-.09]	.01 [.02]*	-.07 [-.13]
Police	.00 [.00]*	.13 [.22]	.05 [.09]*	.06 [.10]*	.06 [.10]*
Managers	.05 [.08]	.13 [.22]	.11 [.18]	.05 [.10]	.05 [.08]
Sales	-.01 [-.02]	.09 [.23]*	.09 [.15]	.00 [.00]	.04 [.07]
Skilled/Semi-skilled	.01 [.01]	.12 [.21]	.01 [.01]	.04 [.06]	.05 [.12]
Mean	.01 [.01]	.13 [.22]	.08 [.13]	.04 [.07]	.05 [.08]
O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism * More than 100% of the variance accounted for.					

It should be noted that the validity coefficients for the Big Five personality inventories presented here are very different from those in the earlier section on cognitive tests and the uncorrected figures are much lower than those found in Ghiselli's earlier report. Barrack and Mount comment on the low values of the coefficients by explaining:

> We would like to re-emphasize that our purpose was not to determine the overall validity of personality; in fact, we question whether such an analysis is meaningful. Rather, the purpose was to increase our understanding of the way the Big Five personality dimensions relate to selected occupational groups and criterion types (p. 17).

This Barrick and Mount article appears to confirm Guion and Gottier's pessimistic view on the efficacy of personality tests. Yet the Barrick and Mount paper is often quoted in books and journals to support the predictive validity of personality measures in comparison to Ghiselli's earlier work which, it is reported, does not support the idea. Cheryshenko *et al.* (2011) note that meta-analysis tries to estimate performance under optimal conditions where job performance is measured without error and where there is no restriction of range on the individual difference variables. Barrick and Mount's results, summarized in Table 2.6, shows that, in some cases, over 100 percent of the variance is explained. It would appear from these instances that the authors have over-corrected for restriction of range and the unreliability of the personality measures. The disagreement between the findings of these two studies led Goldberg (1993) to write 'This inconsistency in the findings between two large-scale quantitative reviews of a similar body of literature is *befuddling*' (p. 31, emphasis added).

Using meta-analysis Tett *et al.* (1991) looked at the predictive validity of personality tests as predictors of job performance to try to investigate conflicting findings in previous meta-analytical studies. They divided the types of study into two groups. In

one group were studies which were theoretically driven, that is, the study had a rationale for investigating the personality measure. In the other group were studies which were empirically driven, that is, they were exploring personality relationships to develop personality theory. Tett *et al* used 97 studies from a wide variety of jobs with a total sample of over thirteen thousand people. They found that the overall correlation between personality measures was 0.24 (0.16 uncorrected) but, in the studies which were theoretically driven, a correlation of 0.29 was found compared to 0.12 for the studies which were exploratory. In addition, Tett *et al.* examined the theoretical studies which were based on a job analysis to define which personality traits would be useful in which kinds of jobs. Here they found an even higher correlation of 0.38. This demonstrates that, when there is a clear hypothesis about the relationship between personality and job performance, a higher and more practically useful coefficient can be found. Tett *et al.* note: 'Contrary to conclusions of certain past reviews, the present findings provide some grounds for optimism concerning the use of personality measures in employee selection' (p. 703). It should be noted that, in their results, only emotional stability had non-zero correlations with performance.

Tett *et al.* explain that their findings support those of Barrick and Mount (1991) 'but ours are notably more positive' (p. 727). Tett *et al.* suggest that the reasons for the discrepancies in the findings are that Barrick and Mount used different methods and procedures. One of these was that Barrick and Mount used an averaging of coefficients, so that positive and negative values cancelled each other out, rather than using an absolute value when averaging which Tett *et al.* adopted. Another difference was that Tett *et al.* did not include non-significant studies. Barrick *et al.* (2001) have criticized some of the small sample sizes used by Tett *et al*. For example, their agreeableness sample size was only 280 people. The discrepancies in the meta-analysis findings are a problem for practitioners. In addition, practitioners may be wary of using the findings from United States research samples to base their decision about which personality measures to use for selection. Unlike cognitive ability tests personality tests may not be appropriate because of cultural differences.

In addition to his meta-analytic studies on cognitive tests from European research samples, Salgado has researched the validity of personality tests and job performance from the same perspective. Salgado (1997, 1998) found that the corrected validity coefficients for Conscientiousness and job performance were 0.23. as was the equivalent coefficient for Emotional Stability. Salgado also investigated whether these two personality measures could add anything to tests of general cognitive

ability. His results showed that both Conscientiousness and Emotional Stability added incremental validity of about ten percent beyond that obtained by general cognitive testing.

Hermelin and Robertson (2001) investigated some of the major meta-analytic studies of personality measures described above. They discuss the different methods and assumptions used by earlier reviewers. They note that a variety of procedures and methods were used, and that the research was produced whilst meta-analytical methods were still being developed. They report that:

> … as things stand, a meaningful comparison of meta-analytic validity coefficients estimated by different meta-analyses, is often impossible.
> Different meta-analytic studies have corrected the means and variance of their observed validity coefficients for different experimental artefacts (p. 253).

To take this work forward, Hermelin and Robertson applied standard procedures to deal with restriction of range and unreliability of criteria to the data in the original samples. They then sorted their findings into three categories: high validity selection methods, medium validity selection methods, and low validity selection methods.

The high validity selection methods, which have corrected validity coefficients exceeding 0.45, are cognitive ability tests (used in medium to high complexity jobs) and structured interviews. The next group of medium validity, with corrected validity coefficients form 0.25 to 0.45 are for biographical data, unstructured interviews, some personality measures used before 1982[6] and integrity tests. The low validity selection methods with coefficients from 0.0 to 0.25 are the Big Five personality scales. Hermelin and Robertson list the Big Five dimensions in order with Conscientiousness having the highest mean validity, followed by Emotional Stability and Extraversion and then, after these, Agreeableness and Openness.

Dudley *et al.* (2006) point out that the majority of meta-analyses on the prediction of personality to job performance have been on the broader traits of the Big Five rather than investigating the narrower traits or subscales. Dudley *et al.* found that the narrower traits of Conscientiousness predicted job performance over and above the broad global personality dimension, but this depended on the particular performance criteria and the particular job. They concluded that there could be benefits in using narrower traits of Conscientiousness to predict job performance.

---

[6] Hermelin and Robertson (2001) only mention 'the personality tests covered by Schmitt *et al.* (1984)'. These are studies which were published in two journals between 1964 and 1982.

2.7 Issues with Personality Inventories and Prediction

The enthusiasm for using personality tests to select people for jobs based on meta-analytical studies has waxed and waned over the previous six decades. Morgeson *et al.* (2007) reviewed the pros and cons and report that 'the observed validities of personality predicting job performance criteria are low and have not changed much over time; …when evaluating the usefulness of using personality tests to select applicants, one must not ignore the observed, uncorrected validity' (p. 1029). Morgeson *et al.* (2007) conclude their review of personality tests by stating that 'They are poor predictors of criteria such as job performance and are difficult to justify as a basis for making high-stakes decisions about individuals' (p. 1032).

Some experts have even stronger views. The respected pair of consultants, Johnson and Blinkhorn (1994) wrote an article for *The Psychologist* which began:

> Proponents of the use of personality tests for occupational selection continue to play fast and loose with statistical methods, and to make claims which do not stand up to close inspection. They are not the only offenders in the psychological community. They may not be the worst offenders. But they are amongst the most conspicuous offenders in so far as the impact of psychology on the everyday lives of the population at large is concerned (p. 167).

Johnson and Blinkhorn cite several reasons why these meta-analytic studies are at fault, including the use of over correction for the inventory reliability. They describe Tett *et al.*'s study as '… a shining example of the style of heroic labour, incompetent statistics and wishful thinking that make up the supporting literature, and it neatly supports our original claim that evidence for the validity of the tests is thin' (p. 168).

So why does such a lot of literature focus on personality inventories? They have, to some extent, a kind of face validity in that they appear to measure important traits which should help people succeed in what they are doing or prove an obstacle to their success. Wood (2003) notes that personality research can be useful for practitioners: 'The Big Five offer a simple and quantitative predictive framework … a conveniently and theoretically sound way of filling the 'personality' side of such studies' (p. 12). Wood also explains that managers, when dealing with personnel practitioners, strongly express the view that personality matters. For example, managers see conscientiousness to be as important as general mental ability. In addition, they see conscientiousness together with agreeableness and emotional stability (opposite of neuroticism) as factors which contribute to a manager's success. Managers view people without these traits to have tendencies which can be counter-productive for the organization. Oswald and Hough (2011) note that 'Although laypeople take the influence of personality on individual and

group outcomes for granted, the research evidence has not been so obvious' (p. 153). They recommend the use of personality tests as incremental on top of cognitive tests.

In a recent review of personality as a predictor of work-related behaviour and performance, Burch and Anderson (2008) started their article by noting that 'Personality has an important role in helping both psychologists and managers understand the nature of work-related behavior and performance' (p. 261). However, when reviewing the literature, the authors reported that critics had noted that the predictive validities are so low that practitioners are still no better off than in the 1970s or 1980s: '... the variance explained by personality in workplace behavior and performance is apparently not high' (p. 263). Burch and Anderson recommend that research should be more focused and they list seven ideas about the future directions. The following four may be of use in this study. Their first proposal is to investigate more of what Hogan and Hogan (2004) have called 'the dark side', the extreme behaviours as a reaction to certain stressful situations. Second, to look at the role of low level traits in the Big Five model[7]. Third, to look for non-linear and curvilinear relationships and, lastly, to look at the interaction of personality with other variables. These ideas will be built into this research.

One of the problems with personality tests as a selection device, compared to cognitive ability tests, is that personality tests can be faked (Searle, 2003). Candidates can direct their responses in ways they believe the organization is selecting for. Matthews (1997) also notes the problems of response bias, acquiescence and social desirability. The Big Five inventories do not have a 'lie scale' like the Eysenck Personality Inventory nor do they have an ipsative format to control some of these issues like the Gordon Personal Profile Inventory[8] (Dyer, 1984). The ipsative format presents the test-taker with a series of four equally agreeable or equally disagreeable statements from which the test-taker has to choose which he or she is 'most like' and 'least like'. However, ipsative tests are not without problems particularly if attempting to compare scores between candidates, since each person has their own baseline (Johnson *et al.* 1988). This means that comparisons of norms or techniques, like correlation, are not applicable to ipsative tests.

Even though the Big Five became established quickly, researchers like Hough (1992) questioned the use of the Big Five because there was 'not an adequate number of dimensions for predicting job performance and other important life criteria' (p. 139). She continues: 'Not only are the Big Five too broad and heterogeneous, additional

---

[7] In the previous section it was mentioned that Dudley *et al.* (2006) had noted the increased prediction by using the narrower traits of Conscientiousness.
[8] Gordon's scales measure eight dimensions: ascendancy, responsibility, emotional stability, sociability, cautiousness, original thinking, personal relations and vigour.

constructs above and beyond the Big Five appear important if the goal is prediction rather than description' (p. 139). Using meta-analysis Hough produced further useful dimensions based on Tupes and Christal's original work[9] and found nine factors[10] including 'Locus of Control'. Locus of Control, with a validity coefficient (uncorrected) of 0.19 was, together with 'Achievement', the highest predictor of overall job performance. Noteworthy, too in this research were the low correlations she found between the nine dimensions and technical proficiency in the job; Hough's dimension 'Intellectance' correlated 0.16 with technical proficiency, but the coefficients for the other eight dimensions were 0.06 or less. Hough and Furnham (2003) report that the Five Factor Model is not as comprehensive as is often imagined. Several important traits like rugged individualism, aggression and hostility are missing. These authors also note that other personality factors like Locus of Control and Type A Behaviour are important in work settings. Other critics, Oswald and Hough (2011), note that the basic problem lies with the actual development of some personality theories. They comment that 'FFM's roots are in the English lexicon, not in psychological theory' (p. 155).

Two other pieces of research cast doubt on the usefulness of the Big Five. Robertson *et al.* (2000) argue that, whilst many see Conscientiousness as the 'g' of personality, this may not be the case. Given the nature of managerial work it may not extend to all managerial jobs. In their research they found that the validity coefficient for Conscientiousness and current job performance was close to zero, and for promotion potential was -0.2. They concluded 'that conscientiousness is not influential in determining managerial performance' (p. 171) and argue 'for a more multi-faceted view of both performance and personality' (p. 179). The second piece of research, reported by Cheung *et al.* (2001), attempted to replicate the five factor structure of personality using a Chinese sample. They were not able to do this which may suggests that the Big Five dimensions are a social construct of personality in Western culture.

The final point on the role of personality as a poor predictor of job performance and training success is the influential book by Mischel (1968). He concluded that behaviour is much more determined by situation than by personality; personality has only a small role to play. In a recent review of individual differences in predicting organizational outcomes Cheryshenko *et al.* (2011) remind the reader about the importance of Mischel's ideas that individual behaviour is not consistent enough across time and situation to allow prediction by personality measures. The predictive validity of

---

[9] Hogan also developed his own personality inventory from the same Tupes and Christal material.
[10] Hough (1992) lists nine dimensions: Affiliation, Potency, Achievement, Dependability, Adjustment, Agreeableness, Intellectance and Rugged Individualism.

personality tests rarely exceeds 0.3. This only explains about ten percent of the variance. Other aspects of assessment and the overall prediction of assessment centres are often higher (eg, Hunter and Hunter, 1984). The background to some of these alternative assessment methods will be described below.

2.8 Situational Tests

A major hurdle with selecting young people directly from schools or universities is that the more practical assessment exercises, which are geared towards some general future job requirement, cannot be direct simulations of actual jobs. The candidates have little experience of work let alone specific job roles, so some generic exercise to assess their skills has to be developed. In selecting military officers there is little point in asking candidates to role-play being a junior officer in a tactical situation. No one could do this without proper training. This means that the tasks given to the candidates must not discriminate against those with little experience and be generic in nature. The two types of task used at the Admiralty Interview Board, the practical Gym task and the group Discussion Planning exercise (described in Appendix D, p. 129) were developed to examine skills such as communications skills, decision making, cooperation and leadership potential.

The Gym task is an example of a 'command task'. Ansbacher (1951) describes the early use of Command Tasks (Führeprobe – leadership test) and Leaderless Group Discussions (Rundgespräch – round table discussions) in the German Army from 1925. These were adopted by the other German services, particularly the German Navy. Ansbacher also describes the Hungarian Army using a Command task involving planks, ropes, etc., in 1936. Jones (1991) notes that these exercises had good predictive validity (0.28 to 0.51) five years after assessment.

In a review of 85 years of research in this area Schmidt and Hunter (1998) looked at combining tests of cognitive ability with various other types of selection measures such as work samples and personality tests. They found that cognitive ability tests plus measures of work samples had high validity.

2.9 Scored Biodata

Mael (1994), reviewing the literature on biodata, notes that a person's previous behaviours and experience matter because in this case the past predicts the future. Scored biodata are used by some organizations to help predict a person's success in the job. Biodata, or life-history information, is gathered on a biographical questionnaire which is often an application form specifically designed to ensure that those attributes

necessary for the job are covered. Interviewers and assessors should collect facts from application forms which are relevant to the job description criteria but these are often badly designed. This not only hinders the candidate when filling in the form but also the interviewer or assessor who has to extract the pertinent information. Selectors have to make judgements about the relevance of a candidate's qualifications and experience when selecting people even if they have a person specification developed from a job analysis. Scored biodata can help ease some of these problems.

Only organizations which have a large number of candidates can use scored biodata successfully. The biodata is modelled against a predictor of success, e.g., success in training or annual appraisals. This has to be computed using a sample of several thousand people who have already entered the organization. After the biodata scoring model has been completed these predictors have to be cross-validated with about another thousand people who were not in the original sample. In total you need the records of four thousand people! The Admiralty Interview Board was the first assessment centre in the UK to use biodata developed by Drakeley (1988) and Drakeley *et al.* (1989). Reilly and Chao (1982) note that for a military sample the validity coefficient for biodata was 0.39 for predicting training success.

2.10 References

Many books, both theoretical and practical, which look at selection and assessment, do not even mention references or just give the topic scant treatment; but, after the interview, references are the most widely used selection tool. Robertson and Makin (1986) found that about 96 percent of the organizations they surveyed used references to aid selection; but Kingston (1971) pointed out that often references were not used in the selection of candidates and were only taken up after the best candidate had been chosen. Employment Tribunal cases have altered people's views about the use of references since both individuals and organizations have been found negligent in not giving accurate references.

2.11 Meta-Analysis Research on Assessment Centre Methods

Gaugler *et al.* (1987) conducted a meta-analysis using results from twelve thousand candidates in 50 Assessment Centres producing 107 validities. Their results showed a correlation of 0.37 for combined dimensions and 0.5 for intelligence tests. Since then reviewers have gathered together the meta-analytic research and produced tables in book chapters to aid practitioners in selecting appropriate selection measures. Robertson and Smith (1989) present combined figures for the results of separate meta-

analytic studies published by Hunter and Hunter (1984), Schmitt *et al.* (1984), Reilly and Chao (1982) and Ghiselli (1973). Over ten years later, Anderson and Cunningham-Snell (2000) produced a diagram which showed the predictive validity of popular methods. Their work is also derived from Reilly & Chao and Hunter & Hunter but, in addition, they add Gaugler *et al.* (1987), Ones *et al.* (1993), McDaniel *et al.* (1994) and Schmitt *et al.* '(1994)'[11]. These sources are given in Table 2.7 together with figures quoted by Hough (1992) which she took from a paper presentation in 1985. All the coefficients shown are corrected. It can be seen that there is disagreement on the coefficients for some measures, noticeably for personality assessment depending on which sources have been used.

Table 2.7: Validity coefficients from meta-analysis for various selection measures (adapted from Robertson and Smith, 1989, p. 93, Anderson and Cunningham-Snell, 2000, p. 83 and Hough, 1992, p. 140). [Adapted and reproduced with permission from Wiley-Blackwell and from Taylor and Francis]

Selection Method	*Robertson and Smith*	*Anderson and Cunningham-Snell*	*Hough*
	Range of mean validity coefficient	*Validity coefficient*	*Validity coefficient*
Work sample	0.33 – 0.54	0.54	0.54
Ability Composite	0.53		
Ability Tests		0.54	0.53
Assessment Centre	0.41 – 0.43	0.68	0.50
Supervisor/Peer Evaluation	0.43		0.49 (peer)
General Mental Ability	0.25 –0.45		
Biodata	0.24 – 0.38	0.37	0.40
References	0.17 – 0.26	0.13	0.14
Interviews	0.14 – 0.23		
Structured Interviews		0.44	0.25
Unstructured Interviews		0.33	0.20
Personality Assessment	0.15	0.38	0.10
Integrity Tests		0.41	
Interest Assessment	0.10		
Self-assessment	0.15	0.13	

Work sample tests, general mental ability and psychomotor procedures have the best validity coefficients whilst references, interviews and personality assessments have low but positive correlations which mean they can be used to add information to the

---

[11] Do the authors mean Schmitt *et al.* (1984) already used by Robertson and Smith? Neal Schmitt, Michigan University, does not list a relevant publication for 1994 only 1984.

selection process, but should be used with caution. Reporting on a survey of test use, Robertson and Makin (1986) note that 'the frequency of use of various mainstream techniques for managerial selection in the UK is inversely proportional to their known validity' (pp. 171-2). For example, the interview can be a very poor predictor of performance if it is unstructured, yet both selector and candidate would not wish to cut out this part of the selection process out.

More recently Hermelin *et al.* (2007) conducted a meta-analysis of previous assessment centre research using standardised procedures. The observed, uncorrected correlation found for overall assessment centre ratings was 0.17 with supervisors' job performance ratings. When corrected for restriction of range this figure is 0.20 and, when further corrected for criterion unreliability, it becomes 0.28. The 95% confidence interval is between 0.24 and 0.32. This figure is much lower, in fact statistically significantly lower, than the figure of 0.36 found by Gaugler *et al*. (1987) which is often reported. Hermelin *et al.* (2007) also point out that the validity coefficients for cognitive tests used as part of the assessment centre procedures were 0.10 rather than the much higher figures for cognitive tests reported elsewhere where these tests are used as the main predictors of success. Hermelin *et al.* also point out that validity coefficients have decreased over time. They suggest that this is due to more pre-selection to screen out some candidates before the expensive assessment centre process. This could also be due to social trends where people have more qualifications which are used for screening and so produce a more homogeneous candidate population for validity research. This same point was made about cognitive tests earlier in this chapter (Section 2.4).

In a recent review of assessment centre research Arthur and Day (2011) note that a smaller rather than a larger number of dimensions gave a more accurate assessment of candidates. Jansen (1997) lists what he considers to be the most important dimensions in assessment centres: power of intelligence, social skills, power of determination, will-power (by this Jansen means personal strength and tenacity). He notes that potential is usually easier to predict than actual performance. Arthur *et al.* (2003) used meta-analysis in their research into the most predictive assessment centre dimensions rather than the overall assessment rating. They found that these have a corrected validity coefficient for job performance of between 0.25 and 0.39 which is similar to the results found by Gaugler *et al*. (1987). Arthur *et al*. (2003) found that the four best predictors were dimensions which assessed problem solving, influencing others, organizing and planning, and communication.

2.12 Chapter Summary

This chapter has reviewed some of the selection measures used in assessment centres. It appears that measures of cognitive ability have reasonable predictive validity, but personality inventories have a more chequered outlook, even with the adoption of the Big Five model of personality structure for both selection and research. There has also been criticism of the underlying methods used to aggregate the data in meta-analytic reviews. Despite this, managers and those who advise them on selection have a belief in the usefulness of personality differences to predict managers' outcomes. Other methods such as situational test, scored biodata and references have been shown to have useful predictive properties. The next chapter examines selection in the military.

3 Selection in the Military

3.1 Introduction

This chapter reports some results from selection research in the armed services. Other issues related to the job performance and training success of armed service officers, such as leadership and command, will also be discussed. One of the aims of military training is to ensure that personnel can perform their roles under stressful conditions, so workplace stress and selecting personnel for hazardous working environments are examined. In addition, Locus of Control and Type A Behaviour measures are included, because they have consistently shown promise in predicting success in stressful situations.

3.2 Testing in the Military

The predictive validity studies reported in the previous chapter usually included military samples, sometimes very large samples, in their meta-analytic research. For example, Hunter and Hirsh (1987) used a sample which included 500 military studies involving nearly half a million service personnel. They found that the tests used by the military were equally as good as those in civilian situations in predicting job performance and training success. The two main differences were that cognitive tests appeared to be slightly more predictive in military studies, particularly for training success rather than job performance. Personality tests were more predictive in the civilian samples.

One of the main reasons for cognitive tests to be better predictors of training in the military is that the tests used were usually designed for a specific purpose and validated against the relevant training criteria. Often military testing takes place across all prospective recruits with a very wide ability range to select infantry soldiers through to electronic technicians. Civilian testing may be designed to be more focussed on testing groups of people within a narrower range of abilities where educational qualifications are used as an initial filter.

Milgram (1991) reported that the emphasis of military selection research was on cognitive areas rather than on personality because large organizations are much more interested with structure and function than individual differences. This is evident in the military. Service personnel of all ranks are continually involved in training their junior staff to be able to take over from them. This can happen very quickly in conflict situations but also in peacetime as well, with people being posted frequently and at short notice. Milgram recommended the use of self-description instruments with closed-

ended questionnaires to aid selection. However, some individual studies have shown the potential of personality testing, for example, Bartram (1995) found an uncorrected correlation of around 0.2 for extraversion and UK military flying training success.

Ree and Earles (1991) looked at the role of general ability (*g*) and various specific abilities in predicting the training scores of nearly 80,000 US Air Force personnel. They found that the specific abilities did not add any prediction above *g*. A follow up investigating job performance criteria (Ree *et al.,* 1994) found similar results with *g* as the best predictor and the specific factors adding only a small (though statistically significant) prediction.

According to Schmidt *et al.* (1992) Project A was the largest and most expensive selection research programme ever carried out. McHenry *et al.* (1990) reported the results from this US army study of over 40,000 personnel (but not officers) which used a predictor battery of tests. This battery included cognitive and perceptual ability measures and also the ABLE scale (Assessment of Background and Life Experiences). The ABLE scale quantifies personal factors like temperament, personality, interest and job outcome preferences, and effort and achievement orientation to help predict performance aspects like leadership. McHenry *et al.* found that cognitive and perceptual-psychomotor ability tests were the best predictors of job specific and general task efficiency. The personality and temperament composite was the best predictor of the serviceman giving extra effort, supporting peers and displaying personal discipline. However, the best predictor was when cognitive and temperament aspects were used jointly. McHenry *et al.* concluded that the US Army could improve the predictive ability of their current test battery by adding non-cognitive predictors.

Borman *et al.* (1991) who used the same data from Project A noted that achievement orientation and dependability explained the variance in supervisory ratings. Despite these findings, a review of testing in the military by Hardinge (1997) reported that 'the lower reliability and validity of personality tests has meant that they have not been able to add anything useful to the predictive power of the selection process so far' (p. 173).

3.3 Leadership

It is likely that the topic of leadership will feature in any research study of military officers, but the reviewers of the research in this field are not always convinced of the efficacy of the research. Certainly, there is a lot of research on the topic. Recently, Young and Dulewicz (2008) noted that 'The proliferation of literature on leadership … has led to the suggestion that the construct has been studied more extensively than

almost any other aspect of human behaviour' (p. 17). Many years earlier Bass's (1960) 500-plus page book on leadership had expressed a similar opinion, with the caveat that leadership was the most researched and least productive area of psychology. Hogan and Hogan (2004) who also commented on the tremendous effort which has gone into researching the correlates of personality and leadership, wrote that 'There is a huge speculative literature on this topic that is useful primarily for its entertainment value' (p. 3). Some writers, like Den Hartog and Koopman (2001), report that there is much research on leadership which has 'always been an important topic in work and organizational psychology' (p. 166); and a recent summary of the research on the evidence on leadership and performance by Kaiser et al. (2008) states that leaders do affect the performance of organizations.

Judge *et al.* (2004) reported a meta-analysis of 150 studies on intelligence and leadership. Their results gave an unrestricted correlation of 0.21 (0.27 corrected) between intelligence and leadership. However, they report that perceptual measures of intelligence were stronger predictors than paper and pencil measures and that stress levels moderated this relationship. They conclude that 'Overall, results suggest that the relationship between intelligence and leadership is considerably lower than previously thought' (p. 542), although they do note that problem of restriction of range given that leaders are often selected on their intelligence. Ilies *et al.* (2004) have even looked at how genetics can effect leadership emergence and found, using meta-analysis of previous studies, that the genetic component accounts for about 17% of the variance in the leadership emergence data. They also found that the genetic effect is mediated by intelligence and the Big Five personality traits, though the personality factors do not influence some personal factors which explain emergent leadership such as height and attractiveness.

Judge *et al.* (2002) reviewed both the quantitative and qualitative research using the five-factor models of personality and leadership and conducted a meta-analysis using 226 correlations from 75 samples, including military research. Judge *et al.* found that Extraversion was the most consistent dimension of the Big Five in predicting leadership ratings (0.31, corrected), but three of the four other dimensions (with the exception of Agreeableness) were also correlated with leadership assessment. Table 3.1 gives the uncorrected and corrected validity coefficients for Judge *et al.*'s total research sample. Also presented in this table are the corrected coefficients for the Big Five and leadership for the military samples in the research. It is noticeable that, except for Neuroticism, the coefficients are very much lower than in civilian samples. Perhaps military personnel

are more homogeneous group on leadership skills, because it forms such an essential part of their training and work.

Table 3.1: Validity coefficients for the Big Five Personality dimensions and leadership ratings (adapted from two tables in Judge et al., 2002, p. 771 and 773). [Copyright: The American Psychological Association]

	Total Sample		Military Sample
	Uncorrected Coefficient	*Corrected Coefficient*	*Corrected Coefficient*
Openness	0.16	0.24	0.06
Conscientiousness	0.20	0.28	0.17
Extraversion	0.22	0.31	0.16
Agreeableness	0.06	0.08	-0.04
Neuroticism	-0.17	-0.24	-0.23

The second table, Table 3.2, adapted from Judge *et al.,* shows the validity coefficients for some of the lower order personality traits which had significant coefficients where the confidence limits of the coefficient did not cross zero. Achievement and Dependability, (which are traits which make up Conscientiousness) and also Sociability and Dominance (which are traits of Extraversion) produced higher validity coefficients than the higher order dimension of Conscientiousness and Extraversion. Judge *et al.* also included a Locus of Control and a Self-esteem measure as part of the analysis, and found that both of these measures gave some prediction (see Table 3.2). They noted that there was a growing body of research suggesting that Locus of Control and Self-esteem are measuring the same thing as Neuroticism. Judge *et al.* also computed a multiple correlation for all five personality dimensions and found the overall relationship to be 0.48, corrected.

Table 3.2: Validity coefficients for lower order personality traits and leadership ratings (adapted from a table in Judge et al., 2002, p. 772). [Copyright: The American Psychological Association]

	Uncorrected Coefficient	Corrected Coefficient
Locus of Control	0.08	0.13
Self-esteem	0.14	0.19
Achievement (C)	0.23	0.35
Dependability (C)	0.18	0.30
Sociability (E)	0.24	0.37
Dominance (E)	0.24	0.37

In another review of personality and leadership Bass (1998) concluded, 'When it comes to predicting transformational leadership and its components, there is no shortage of

personality expectations. However, the empirical support has been spotty' (p. 122). Bono and Judge (2004) carried out the first meta-analysis to investigate the relationship between the Big Five personality dimensions and transformational and transactional leadership. Using data from 26 studies they found that Extraversion was 'the strongest and most consistent correlate of transformational leadership' (p. 901), but these were 'generally modest validities' (p. 910) Bono and Judge's other findings were in the same order as the results found by Judge *et al.* (2002) presented in Table 3.1. Bono and Judge conclude that narrower, specific, relevant personality traits and non-dispositional factors should be the focus of future research. Judge *et al'*s (2009) recent review of research of the Big Five dimensions and the 'bright side' (core self-evaluations, intelligence, and charisma) and 'dark side' (narcissism, hubris, dominance, and Machiavellianism) of leadership traits reported that the research findings were not altogether conclusive in their predictive directions.

Young and Dulewicz (2008) used the more organizationally-oriented Occupational Personality Questionnaire (OPQ) (SHL, 1999) together with their own Leadership Dimensions Questionnaire (LDQ) (Dulewicz and Higgs, 2005) and found statistically significant differences between the personality ratings of Royal Navy officers and ratings (non-officer personnel). Young and Dulewicz noted that conceptual skills were more important at the senior level while technical and supervisory skills were more important at the junior level. A major contrast was that officers were more likely to want to be 'in charge' and the OPQ dimension 'controlling and ambitious' was a distinguishing indicator. However, the LDQ dimension 'motivation' (energy and drive) explained most of the variance found in 'overall performance' in the context of annual reports and predicting marks on 'leadership and management'. In an earlier report, Young and Dulewicz (2003) had found that 34% of the OPQ scales were significantly correlated with 'overall performance' and for the LDQ dimensions this was 73%. This again suggests that specifically tailored measures may be more useful than off-the shelf inventories.

3.4 Command and Hazardous Environments

Command in the military context refers to the leadership, management and administration of a self-contained unit or formation like a regiment, battalion, surface ship or submarine by one person: the commanding officer. The commanding officer may be in charge of sub-units within the formation, with different sub-groups of specialists such as warfare, engineering, supply, administration and training, but he or she will have overall responsibility for these formations. The submarine commander, in

particular, can be working on his own and out of communication with other units and formations for many months. The role of a submarine commanding officer can be a very lonely position, since he will be the only person aboard with senior rank and has nobody else to confide in, whereas junior officers can consult with colleagues of the same rank.

One strand of research on personality which relates both to command and stress (discussed in the next section) is concerned with the personality of managers on North Sea oil rigs. Flin carried out research looking at job performance in stressful situations like oil rig emergencies over several years; and Flin and Slaven (1996) started their paper on this topic by noting that:

> A relationship between personality and incident/emergency command ability is often assumed to exist, yet little research has explicitly examined such a relationship. Good leaders are expected to be calm, decisive under pressure and confident in action (p. 40).

Then quoting Downes' (1991, p. 145) they list the qualities of a naval leader as: 'Intelligence, commonsense, integrity, judgement, enthusiasm, loyalty, cheerfulness, sense of humour, energy, fortitude, moral courage, the will to dominate and decisiveness'. Slaven and Flin (1997) proposed that leadership and command ability, stable personality and decision-making skills were all required in the emergency management role. Flin and Slaven used a personality inventory, the Occupational Personality Questionnaire (SHL, 1999), with Offshore Installation Managers who took part in six simulated emergency exercises on a four-day emergency management course. The managers on the training course were rated by experienced senior managers and trainers on several aspects of performance during six simulated emergency scenarios. Flin and Slaven collected data on 93 managers and found some significant correlations of between 0.22 and 0.31 between the personality dimensions and the ratings, but noted that 'there were few significant correlations and they were modest in size' (p. 45). This emphasizes the difference between research conducted to develop theory and that carried out by practitioner for selection purposes. A statistically significant relationship was found here but could not be used for selecting out unsuitable managers.

In a review of the selection of personnel for hazardous performance, Hogan and Lesser (1996) looked at the results of personality measures to distinguish effective performance in areas like bomb disposal and naval diver training. They found that Conscientiousness was extremely important for performance in hazardous jobs. People need to follow rules but, as Hogan and Lesser noted, the job holders must also be willing to put themselves into the excitement of hazardous environments. They found that the most effective performers were not spontaneous, thrill-seeking or impulsive and did not take risks. They were 'non-conforming, but not reckless' (p. 214) and they were 'cautious, careful,

attentive to detail and not impulsive' (p. 219). Hogan and Lesser also found some links to other personality traits like maturity and emotional stability which were linked to well-being. Openness was also considered to be an important link, since successful personnel were found to be 'open, curious, analytical, and interested in new ideas and experiences (pp. 218-9). These findings led Hogan and Lesser to postulate a link between Openness and general mental ability.

This real world research suggests that personality tests have only low predictive validity in stressful environments reinforcing Mischel's (1968) argument, presented towards the end of the previous chapter, that the situation is far more important than personality in predicting a person's behaviour. Hannah *et al.* (2009) reported that, before they conducted their literature review, leadership in extreme contexts was assumed to be very similar. However, after reviewing the literature and discussing the topic with professionals involved with hazardous situations they concluded that leadership was uniquely context specific: 'We believe extreme contexts create particularly unique contingencies, constraints and causations; requiring researchers to view such leadership as inherently contextualized' (p. 898). So the research undertaken in this thesis may help by looking at another stressful context; and recognising that important situational factors affect the amount of stress in a particular job.

3.5 Stress

Sonnentag and Frese (2003) review of the research on stress in organizations reported that 'empirical research … shows that organizational stress has detrimental effects on individual health and well-being' (p. 479). More recently, Tetrich *et al.*'s (2010) review of the research on work-related health, stress and safety concluded that psychological risk factors are prevalent in organizations and account for a significant amount of absence and ill health.

Cox (1978) used three basic models to explain stress. The first is taken from a lay or dictionary definition model, very similar to the mechanical concept of stress used in physics or engineering. Pressure is applied to the person by stress forces and, in attempting to cope with these forces, the person exerts or strains themselves. Here stress is seen as the dependent variable describing the person's response to the unpleasant environment. The second approach looks at stress as an unpleasant stimulus; an independent variable in the environment. The third model, which Cox considers to be the 'most adequate approach', views stress as a reflection of the lack of fit between the person and their environment. In this case, stress is seen as an intervening variable

between the stimulus and response. This is often called the transactional model (Cox and Mackay, 1981) which recognises the dynamic, reciprocal nature of the relationships between the factors which contribute to stress.

In the transactional model the environment is understood in the widest possible terms to reflect both the person's internal and external environment and their physical and psychological environment. The transactional approach has led to the development of theories such as French *et al.* (1982), McGrath (1976), Cox (1978) and Lazarus and Folkman (1984). These theories incorporate both (1) the notion of person-environmental fit, where stress is seen as a mismatch between the individual's characteristics, like abilities and goals, and (2) aspects of the organization such as work demands and organizational climate. In the transactional model stress is seen as a perceptual phenomenon where the person makes a comparison between the demands placed upon them and the resources they believe they have to cope with the stressful situation. The theories note that stress is the imbalance between the person's perceived environmental demands and their perceived ability to cope with these demands. This interplay between appraisal and coping strategies results in both psychological and physical outcomes. Cox describes coping as psychological, including both cognitive and behavioural strategies, as well as physiological.

Cooper (1998) has reported the usefulness of process theories of stress which have a strong occupational orientation and earlier he outlined a model of stress at work (Cooper, 1986) which shows the interlinking processes. The basic structure of Cooper's model is shown as Figure 3.1. This model forms the basis of the Occupational Stress Indicator used in this research study and will be described in the Method chapter. The additional features within each box of the figure are given at Appendix C (p. 128). On the left-hand side of the diagram the sources of stress at work are shown. Features like 'role in the organization' contain aspects like role ambiguity and role conflict which can be extremely stressful for the 'individual'. Above and below the individual in the diagram are factors which may enhance or inhibit the individual's perception of stress. Individual characteristics like neuroticism or the person's tolerance of ambiguity can alter the person's response to the stress associated with the situation at work. People also spend a lot of time outside work so 'home-work interface' factors and relationships can increase or reduce the perceived level of stress. To the right of the individual in the diagram are shown the individual and organizational symptoms of stress which range from individual outcomes like increased blood pressure and escapist drinking to poor outcomes for the organization such as high absenteeism and poor quality products.

Figure 3.1: A basic model of organizational stress (adapted from Cooper, 1986, p. 326). [Adapted and reproduced with permission from the British Psychological Society]
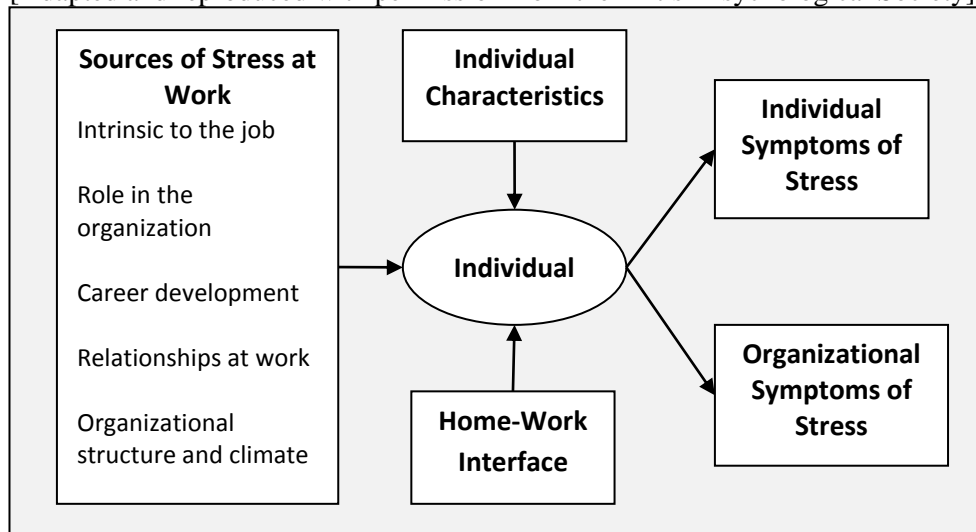
Dewe and Cooper (2007) reviewed the importance of coping strategies in work related stress including personality dispositions. They noted that Neuroticism, Extraversion and Type A Behaviour stand out as the best predictors. In a review of the relationship between personality and stress, Matthews *et al.* (2009) noted (1) that many reported studies are cherry-picked and so the reported results could be due to chance and (2) that the correlations are of a small magnitude so that small changes in the error can affect whether the reported correlation is predictive or not predictive.

Orasanu and Backer's (1996) article on stress and military performance started by explaining that 'The debilitating effects of stress on military performance have long been recognized' (p. 89). They list several stress factors which particularly confront the military: danger and threat; fatigue which includes sleep deprivation, sustained and continuous operations; workload and informational load; and environmental factors such as noise, heat or cold and altitude. They also noted the effects of modern technology which gives the individual little time to think about their response and course of action. On the issue of 'danger and threat' factor the authors note that research has shown that people who believe that they can control the situation are more comfortable in what appear to be stressful situations.

3.6 The Moderating Effects of Type A Behaviour, Locus of Control and Hardiness

This section will look at behavioural factors which have been found to affect performance at work. The sources of stress in the organization, such as interpersonal relationships, are moderated by the individual's characteristics such as Locus of Control, Type A Behaviour and hardiness. The stress effects are negated or modified by

the coping strategies adopted by the individual such as time management and social support. The outcomes for the individual can vary from poor mental or physical health to various aspects of job satisfaction, which are very similar to the hygiene factors and motivators described by Herzberg (1987). As well as outcomes for the individual, the effects of stress often have repercussions for the organization, such as low morale or high absenteeism.

Earlier in this section the notion that personality could predict job performance was examined. Cooper (1986) proposed that stress and personality variables could predict disease. H. S. Friedman and Booth-Kewley (1987a) conducted a meta-analysis of research studies which correlated personality to five diseases: asthma, arthritis, ulcers, headaches and coronary heart disease (CHD). They found that a 'disease-prone' personality related to depression, anger and hostility, and anxiety may exist, but that evidence is weak except for coronary heart disease. Typical findings were correlations of 0.1, but Friedman and Booth-Kewley (1987a) point out that the link between cholesterol and CHD and between smoking and CHD is only about 0.15.

Type A Behaviour is a term coined by the cardiologists M. Friedman and Rosenman (Friedman, 1996) following their work on the association of a driven personality type and coronary heart disease. Their studies indicated that people who had symptoms of time-urgency, impatience and 'free-floating hostility' were two to three times more likely to suffer from coronary heart disease. The original research has been criticized on methodological grounds and some later studies do not support the findings, except for the prediction by free-floating hostility symptoms. Landsbergis *et al.* (2003) reviewed the role of Type A Behaviour in the workplace and occupational health and noted that, although there is limited evidence it suggests that Type A Behaviour linked to recent trends in work organization such as increased safety climate, shift work, working hours and new technology may be increasing the risk of occupational illnesses.

Other researchers have suggested that, rather than being an inbuilt way of operating, Type A Behaviour could be a person's sensible coping reaction to their stressful work environment. For example, in the environment of a submarine, it would be very difficult to laze about and not take time pressures and crises seriously. Friedman and Booth-Keeley (1987b) note that the 'traditional emphasis on hurry sickness in coronary proneness are deemed totally inadequate' (p. 783) and that a more refined model is needed. These authors also found that social support was important, with a correlation of -0.19 between social support and coronary heart disease.

Cooper *et al.* (2001) noted that the Type A personality 'is one of the most interesting dispositional characteristics in stress research, in that it may lead to both positive and negative outcomes for individuals' (p. 121). For example, strong achievement motivation may help vitality and enthusiasm, but the aggressiveness and need for control can lead to frustration and poor relationships both at work and at home. Cooper *et al.* note that those with a Type A personality 'may actively position themselves in situations that require ambition, driveness [*sic*], and competitive behavior; these situations, in turn, may induce greater psychological strain' (p. 119).

In a review of the implications of stress research for management practice Jex (1998) reported that personality has been studied 'most extensively' as a moderating variable. In particular, Type A Behaviour, self-esteem and Locus of Control have been widely studied; and the internal Locus of Control has been shown to predict more desirable outcomes when linked to cognitive and role overload.

In an earlier section of this chapter, Locus of Control was found to be an important factor in predicting job performance (Hough, 1992) and leadership (Judge et al., 2002) but this concept has not yet been defined. Locus of Control (Rotter, 1975) is a variable or construct which is believed to account for a variety of behaviours. External Locus of Control refers to a person's belief that outcomes in a person's life, or the world at large, are not determined by the individual's action but by external factors. On the other hand, internal locus of control refers to the individual's belief that they can control events by their own actions. Research has indicated that people who believe that they can control events may interpret these events as challenging rather than stressful.

Cohen and Edwards (1988) review of the literature on Locus of Control reported that it does have a powerful moderating effect on stress. Theorell (2003) also noted the importance of Locus of Control and argues that this would appear to be a necessary condition to be able to exert control over one's situation for coping with stressors. Sonnentag and Frese (2003) noted that control at work is an important factor in workplace stress and that, if an individual is able to influence their activities this can reduce their stress. These authors note that stress may not only influence the individual's health but also their job performance. However, they pointed to a lack of evidence in actual field studies. They suggest that, when compared to laboratory studies, there may not be such a large stress effect in the real situation because people can alter their situation by switching tasks and using different sorts of coping strategies

Hart and Cooper (2001) note that research has shown that personality variables have a role in the appraisal and coping processes. 'They are an informative and important part of the process that enables people to interpret and respond to their environment' (p. 98). In particular, they point to the roles of Neuroticism and Extraversion that are stable over time, and can determine to some extent the psychological meanings people attribute to events. Frome (2003) points out that extraversion and hardiness are important resources for coping and that such characteristics 'may be conceived of as individual resources' (p. 151). They help people actively cope with problems at work and home, and reduce the likely conflict between the two. On the other hand, these authors note that people with a high Neuroticism score may avoid problems at work and home, but increase the likelihood of work-family conflict. Quillian-Wolever and Wolever (2003) report studies suggesting that social support is the most important psychological buffer.

Kobasa (1979) originally developed the idea of hardiness to explain why some highly-stressed executives coped with the situation while others became ill. She hypothesized that there were three components which helped people cope with stress: that they feel that they can control events (internal Locus of Control); have a sense of commitment, purpose and involvement; and see problems as a challenge and an opportunity for growth. Kobasa *et al.* (1983) also examined cognitive appraisal in high stress conditions and the link between Type A Behaviour and hardiness. They found that these factors were independent and that high hardiness scores protected high Type A executives from illness. Those with high Type A behaviour and low hardiness scores reported most illness. They also found that Type A behaviour was linked to extrinsic motivation and hardiness was linked to intrinsic motivation.

Bartone *et al.* (2009) researched both the Big Five personality factors and hardiness as predictors of leadership performance of 296 army officer cadets at West Point The performance was measured by leadership grades given by supervisor ratings aggregated over four years in two different contexts: summer field training camps and during academic periods. The results showed that leadership performance in the summer field training environment is predicted by the Big Five Extraversion dimensions and also by hardiness after controlling for general intellectual abilities. On the other hand, during the academic period context, leader performance is most strongly predicted by mental abilities than by the Big Five Conscientiousness followed by Hardiness. This research demonstrates that personality factors which may predict leadership are different in different contexts and that hardiness may be a useful predictor across contexts. Johnsen

*et al.'*s (2009) study of 71 Royal Norwegian Navy officer cadets has also shown that aspects of hardiness can predict leadership styles.

3.7 Chapter Summary

This chapter has reported some of the relevant assessment and selection research in military and hazardous contexts and also reviewed the important issue of stress on job performance. The research presented reinforces the findings in the second chapter that cognitive tests can predict job performance and training success in military samples. However, there are more inferences in this chapter about specific personality and other behavioural factors, which can influence the way a person reacts to stress in hazardous and extreme situations. Factors like internal Locus of Control and the person's coping strategies can mitigate the way stressful contexts are perceived.

4 Context and Methods

4.1 Introduction

This chapter presents the background and context of the research and outlines the methods used to collect the data. The chapter will describe some previous research carried out by the author of this present study to set the scene for the investigation presented in the rest of the chapter. The previous study had the same aims and used similar methods as this study, but the previous research was undertaken as part of professional practice rather than for academic research. The specific assessment centre measures, personality inventory and stress indicator used in the research will be presented as well as the course outcomes and performance measures.

4.2 Research Aims and Hypotheses

The two aims of the study were to investigate (1) the relative value of psychometric versus other assessment centre selection measures and (2) the value of these and other approaches for selecting naval commanding officers for practitioners. In pursuing these two aims the following investigations will be made:

1. To investigate the relationship between assessment centre measures and command performance measures to evaluate the predictive validity of the assessment and selection measures over an extended time period, with particular focus on the psychometric tests.

2. To evaluate the predictive validity of two concurrent measures that might be used in professional practice to predict command performance, one for personality and one for occupational stress.

3. To investigate the combined prediction of these various aspects of selection.

4. To investigate the non-psychometric parts of the assessment centre such as the leaderless exercises, both practical and discussion, in order to evaluate the relative merits of these compared with the psychometric procedures.

5. To investigate if the assessment centre dimensions or the overall assessment centre Final Board Mark is a better predictor of success than the individual or combined psychometric tests and other individual predictors.

6. To investigate links between the cognitive ability tests and the personality variables in this sample of participants.

7. To discuss the use of these measures and the use of predictive validity research both in the wider professional practice context of occupational psychology and in the specific contexts of military, naval and similar professions.

It is hypothesised that the AIB measures and the Occupational Stress Indicator scales will correlate positively with performance on the course and the performance grades; so a one-tailed test will be normally used. A two-tailed test will only be used for OCEAN personality measures, where it is not clear which way to predict performance.

4.3 Sequence of Research Events

Table 4.1 below outlines the sequence of events for data collection for this research.

Table 4.1: Table of the sequence of the research.

Approximate dates	Activity
1978 to 1990	Candidates attend the Admiralty Interview Board for three days and are assessed using various psychometric and other assessment measures. Those who are successful join the Royal Navy as officers. The selection data is kept by the Royal Navy for research purposes.
1993 to 1995	The first phase of the research study takes place over three years at the Submarine Command Course with 36 students. This is, on average, 13 years after the students had attended the Admiralty Interview Board. The measures used were a spatial ability test, the 16PF Personality Inventory and a Locus of Control measure.
1996 to 2002	The second phase of the research takes place at the Submarine Command Course with an additional 57 students over a six year period. The original concurrent measures are replaced by the OCEAN Big Five personality inventory and the Occupational Stress Indicator.

4.4 The Admiralty Interview Board Procedures

The selection procedure for AIB candidates takes place over a three day period. The procedures are summarised in Table 4.2 and a full description of the procedures[12] is given in Appendix D (p. 129). The candidates arrive on the afternoon of the first day and are briefed about the following two days. They fill in a biographical questionnaire to give the Board an up-to-date record of their qualifications and achievements. The next morning is taken up with testing the whole group of candidates (twelve people) in

---

[12] The information contained within this thesis that relates to the Admiralty Interview Board (AIB) represents the Royal Navy's Officer Selection Process prior to the year 2005. Since that date, the Royal Navy has introduced competency based assessment for all Officer Candidates and undertakes regular external validation to ensure that the AIB's processes replicate UK Assessment Centre best practice.

one room. The four tests which make up the psychometric test battery (a verbal test, a non-verbal test, a numeracy test and a clerical instructions test) are completed. After lunch the candidates rehearse the two types of tasks they will perform in front of the assessors on the following morning: the Gym tasks and the Discussion Planning exercise. This rehearsal completes the second day.

Table 4.2: Timetable for AIB showing candidates' activities and outputs from this.

Day	Candidates' Activity	Output
Afternoon First Day	Arrive at AIB and are briefed Fill in biographical questionnaire	Scored biodata* Interview material
Morning Second Day	Briefed by a Board President Take psychometric and other tests	Test scores Composite test score
Afternoon Second Day	Practice Gym tasks and group planning Discussion	none

Morning Third Day	Observed by Board Members whilst completing:
	    Gym task
	    Group Discussion Planning Exercise
	
	Two Interviews (Senior Board Members and PSO)
	Candidate is discussed by full Board
	Informed of decision by Board President
	Leave AIB
		Gym mark
		Discussion mark
		
		Board dimensions**
		Final Board Mark

* From 1989 onwards ** Effective Intellect, Leadership Potential, Character and Personality, Service Motivation from 1985		

The actual assessment of the candidates by Board Members (assessors) takes place on the third morning. Candidates are assessed in groups of four with four assessors assigned to each group. The four assessors watch the candidates tackle four Gym tasks; each candidate takes it in turn to lead the exercise. After this the candidates and assessors return to a Board room for the Discussion Planning exercise. This is leaderless, and the assessors rate the four candidates on various dimensions for each exercise. After observing each exercise the Board Members give a set of marks for each candidate and then share their thoughts about candidates with the other members of the Board. In the final part of the assessment process the four candidates are interviewed individually.

When the interviews are finished the Board Members look at the various pieces of information collected on each candidate and discuss this. An overall composite test score is calculated from the four tests which make up the test battery. A biodata score is also calculated based on the biographical details which candidates gave on their first afternoon. These two scores are not disclosed to the Board Members until the discussion of the candidates' practical performance has finished. The composite test score and the

biodata score have been shown to be good predictors of success in future training, by comparing the candidates' AIB scores with results during initial officer training at Dartmouth (DERA, 1999). These two scores feed into the Effective Intellect dimension of the final assessment. Other characteristics of Effective Intellect which have been observed during the assessment process, such as the candidates' performance on the Gym and Planning exercise can moderate the Effective Intellect score.

The other three Board dimensions are Leadership Potential, Character and Personality (moderate predictors) and Service Motivation (a weak predictor). These are also built up from marks on the exercises and interviews. The Final Board Mark is agreed, using the three strongest predictor dimensions. There are again, as in the rest of the process, strict procedures and protocols in place for arriving at the Final Board Mark (see Appendix E, p. 132) The Final Board Mark for each candidate is agreed just before lunchtime, and each candidate is individually informed of their result by the Board President. About half the candidates pass the Board, but they then have to wait to see if they have a high enough score to be selected when all the candidates' scores are compared at a meeting about a month before the next training intake is due. There is a moderate correlation between the Final Board Mark and the training results (DERA, 1999).

An early study on the validity of the Admiralty Interview Board, carried out by Gardner and Williams (1973), found that the correlations between the Final Board Mark and performance of seaman (warfare) and engineering officers were 0.39 and 0.34, respectively. The psychometric tests predicted training marks with correlations of 0.37 and 0.34 for the respective groups. Interestingly, Gardner and Williams noted a low correlation between the Final Board Mark and the psychometric tests, suggesting that the Board did not take much notice of the tests; but these tests do correlate with training marks later on.

Gardner and Williams (1973) also looked at the longer term prediction by investigating the time it took for an officer to be promoted to Lieutenant Commander. The correlation here was 0.40; but this research could be dated, because modern officer entrants are a more heterogeneous group (for example, educated to graduate level) than previously. Edwards and Morrison (1994) found this to be the case in United States naval research. Using time to get promoted as a performance indicator was no longer possible since most Lieutenant Commanders were promoted after 14 years of service.

Dobson and Williams (1989) carried out an analysis of the assessment data gathered in 1982 at the Regular Commissions Board and compared the assessment data of

candidates with their training data at the Royal Military Academy, Sandhurst. They found an overall correlation of 0.33 (0.43 if resignations are included). However, the correlations varied according to the particular group of trainees, which ranged from 0.16 for non-graduates, 0.45 for graduates and 0.65 for those promoted from the ranks.

Internal research (DERA, 1999) was carried out every year to validate the Admiralty Interview Board procedures against the professional training mark (examination scores) at Britannia Royal Naval College (BRNC). Table 4.3 gives the validity coefficients of these measures against BRNC performance.

Table 4.3: Prediction of AIB Tests against Professional Marks at BRNC.

AIB Test or other measures	Correlation (raw)	Correlation (corrected)
Verbal test	0.34	0.41
Non-verbal test	0.31	
Numerical test	0.36	
Instructions test	0.29	0.39
Verbal and Instructions		0.52
Composite Test Score	0.40	0.48
Professional Aptitude Predictor		0.46^
Service Knowledge Test	0.38	
Gym task	0.19	
Discussion Planning task	0.10	
Final Board Mark	0.32	
^ 0.40 for male and 0.54 for female		

However, this is only one set of marks that the selection measures were validated against. Other studies were conducted with (1) navigation (both theoretical and practical), (2) leadership, (3) validation studies linking the predictors to the junior officers' reports in their first jobs after they had completed training and (4) grades on further training course.

It can be seen from Table 4.3 that test scores correlate about 0.30 with the trainee officers' performance and that where there are corrected figures, this brings the coefficients up to around 0.40 for individual tests. The Composite Test Score with coefficients of 0.40 (0.48 corrected) gives a higher prediction. The table also shows that a combination of the verbal and instructional test gives the highest prediction (0.52). Since the professional marks are related to examinations following conventional study this is not surprising. In other areas, e.g., navigation, the non-verbal and numerical tests were found to be better predictors.

4.5 The Submarine Command Course

The Submarine Command Course (SMCC) started in 1917 and is one of the most prestigious command courses in the armed services. It is also known as 'Perisher' because of the high failure rate of (about 30 percent)and a pun on periscope. SMCC is a qualifying course for warfare officers in the submarine service if they wish to serve as Executive Officers on a Royal Navy submarine. This may later lead to the officer being given command of a submarine. The course lasts 24 weeks and normally takes place twice each year from March to July and August to November. There are usually four to six students on each course, supervised by a course Commanding Officer who is the course instructor and known (and addressed) as 'Teacher' by the students.

The first few weeks of the course are spent in the classroom, where the students study command and leadership, and tactical aspects of submarine command. These skills are practised and developed in a simulator before being tried out in a real submarine. At the most intense part of the later exercises two or three frigates are used to assist in the training. Later training involves exercises in taking the submarine into shallow waters to conduct such activities as surveillance and to drop off marines. During the simulator training and real exercises, the students take it in turn to act as the commanding officer of the submarine and are responsible for all aspects of the submarine including navigation, safety and attack.

The SMCC is an extremely expensive course to run but, in addition to reducing the cost of the course, any reduction in the failure rate would also be useful in preventing officers from going through the process of failing the course. This failure can be quite traumatic for students after many years in the submarine service. Failure on this course means that the officer has to leave the submarine service and join the surface fleet.

4.6 A Previous Study

The research presented in this report follows on from a previous research project carried out and reported by the author (Beadle, 1997)[13] when he was employed as a Principal Psychologist working for the Ministry of Defence (Navy). This study examined whether psychometric assessment could be used to reduce the high failure rate on the Submarine Command Course (SMCC). The background and some of the results of this previous study, from now on called 'the first phase' in this thesis, are given in Appendix F (p. 133).

---

[13] This material is reproduced with permission from the Defence Science and Technology Laboratory.

In the first phase three measures were trialled to see if they could predict failure on the SMCC. These were:

1. A spatial awareness test 'Directions and Distances' developed internally,

2. The 16 Personality Factor Questionnaire (Cattell, 1989),

3. A Locus of Control Inventory (Rotter, 1975).

In 1996, after three years and six courses, 36 sets of data were available, 28 of which included scores on the spatial test and the two inventories. The results showed that the spatial ability test did not predict success on the course and only one of Cattell's sixteen personality factors predicted the Pass-Fail result. An analysis of the difference in means between the pass and fail groups showed that students who reported themselves as 'Group-oriented' rather than 'Self-sufficient' were more likely to pass the course [U = 44.5, p = 0.04; two-tailed]. Another 16PF factor, *Suspicious* as opposed to *Trusting* did not have a significant difference in means for the pass and fail groups, but their scores were associated with the overall A to F Course Grade with a correlation of 0.42 [p = 0.03; two-tailed]. The *Suspicious* end of this factor includes traits such as hard to fool, distrustful and sceptical, deliberate in their actions, unconcerned about other people and being poor team members. However, whilst these results were statistically significant the contingency tables drawn for these factors showed a lot of misclassification, making these results theoretically interesting, but not useful in the practical situation[14].

The theory behind the Locus of Control measure suggests that those who attribute their success and failures to internal rather than external factors would be more likely to pass the course. These findings were confirmed with a statistically significant difference between those who passed or failed the course [U = 51, p = 0.04; one-tailed]. The mean score for those passing the course was 11.8 and for those failing the course 14.0. The correlation of 0.25, though significant, meant that the prediction was far from perfect. Again, an analysis of the misclassified students[15] illustrated that using this measure as a selection device would have denied good students the chance of taking the course.

The historical selection data from the Admiralty Interview Board was also examined. The results of the Non-verbal test, taken on average 13 years before the course, and the recorded GCE Mathematics grade predicted the overall grading on the course with statistically significant correlations of 0.39 and 0.51, respectively. Again, whilst this is a useful theoretical result, the table shows that a lot of good candidates would be

---

[14] See the relevant contingency table in Appendix F, Table F.1, p. 135.
[15] See Appendix F, Table F.2, p. 135.

rejected[16], even those with a correlation of nearly 0.4 [N= 28]. The even higher correlation of the Maths GCE grade as a predictor meant that a Grade A, B or C gave an approximate probability of passing the course of 83%, 71% and 57%, respectively [N= 20]. An attempt was made to analyse the cognitive and personality aspects together using linear regression, but no significant findings were found.

4.7 The Second Phase of the Research for this Thesis

Beadle's (1997) report on the above research concluded that psychometric testing might not be the answer to the problems of failure on the course, and that training, development and appraisal procedures should be examined as ways of reducing failure. However, a further investigation with more up-to-date psychological measures might be worthwhile. It was recommended that a different measure of personality should be trialled, with fewer dimensions, such as the Five Factor Model of personality. At the time the first phase report was being written, a United States Services Five Factor Model had become available to UK Defence psychologists who re-named the personality inventory 'OCEAN' (Collis, 1997). A second measure, the Occupational Stress Indicator (OSI) developed by Cooper *et al.* (1988), was also included to examine locus of control and some of the wider issue of stress, like sources of stress, Type A Behaviour and coping strategies.

The second phase of the data collection started in March 1996 with the OCEAN and OSI inventories replacing the ones used in the first phase of the study. Unfortunately, the funding for this research was withdrawn after about two years after the second phase started, but data collection was continued by the author until July 2002. This second phase of research forms the major part of the analysis presented here. The type of information generated on each student on the Submarine Command Course (SMCC) when they attended the Admiralty Interview Board, in particular the psychometric test data, is common to both phases of the study and is the major focus of the research presented in this thesis.

Table 4.4 presents an outline of the data collection in the two phases of the research project, and gives details of the research focus, the timings and student numbers. The table then lists the inventories used at the beginning of the SMCC course and the data collected on the students' performance on the SMCC are shown. Finally, the table outlines the data which was derived from the AIB candidates' database.

---

[16] See Appendix F, Table F.3, p. 136.

Table 4.4: Details of the research carried out.

	Research Phase	
	Phase One	*Phase Two*
Focus of the research	Cognitive skills Personality	Personality Coping with stress
Data collected	March 1993 to November 1995	March 1996 to July 2002
Number of students (93 in total)	36	57
Specific measures administered at the start of the SMCC	Spatial ability test Locus of Control scale 16 PF Personality Inventory	OCEAN personality inventory Occupational Stress Indicator
SMCC Performance measures (Common to both phases of the research)	Pass-Fail Result Course Grade (A to F) Twenty Quality Grades and derived from this: a Total score and subtotals for Tactical, Administrative and Personality qualities	
AIB measures (Common to both phases of the research)	Verbal ability test score Non-verbal agility test score Numerical test score Instructions test score Composite Test Score Gym task mark Discussion mark Board Effective Intellect mark Board Leadership Potential mark Board Character and Personality mark Board Service Motivation mark Final Board Mark Professional Aptitude Predictor biodata score Headteacher's Reference score Maths and Physics GCE grades	

4.8 Research Design

This research is a correlational design. Unlike experiments and quasi-experiments, which test the effect of one variable on another, a correlation study looks at how two variables are related. The biggest problem with this type of approach is that when two variables are shown to be related, this does not mean that one causes the other. There may be other factors which explain the link between these variables. In social research factors like income, social class and education levels can often explain why variables are apparently related. There is also the problem that when two measures, which have similar underlying properties, are used in a research study, a high correlation may be found because of an individual's way of completing the inventories.

One complication with this study design is that some material was collected concurrently at the beginning of the SMCC course, rather than with the original AIB data collected at the Admiralty Interview Board. Another related complication is that the AIB data was collected for selection purposes, whereas the OCEAN and OSI data

were collected purely for research. Schmitt *et al.* (2010) note that meta-analytical research suggests that there is not too much difference in validity coefficients for the two types of designs.

4.9 Participants

The participants in phase two of the study were Royal Navy officers taking part as students on the Submarine Command Course. Fifty-seven students on the twelve courses which took place between March 1996 and July 2002 completed the OCEAN personality inventory and the Occupational Stress Inventory for this new phase of the research. The average age of the students on starting the course was 32.5 years old. The average age of this group when they attended the AIB was 19.8 years old and this was on average 12.6 years before the start of the course.

For the analysis of the Admiralty Interview Board selection data, the selection information from the 57 candidates from the phase two study was combined with the information on the 36 students from phase one of the research, already described in Section 4.6 (pp. 49 -51) of this chapter. [The 36 students in the phase one study attended one of the six courses held between March 1993 and November 1995.] This gives selection information on a total of 93 students and this will be referred to as the combined study. The average age of the students in the combined study was 32.4 years old at the start of the course. The combined group attended the AIB between May 1978 and July 1990 and their average age when they attended was 19.8 years old. Their attendance at AIB was, on average, 12.7 years before the start of the command course. Table 4.5 gives summary details of these ages and the time elapsed relevant to the participants, as well as giving the range of ages and years elapsed. The course is 24 weeks long, or 0.46 of a year, so this makes the length of time between initial assessment at the AIB and assessment on the command course about 13 years on average.

Table 4.5: Ages and time elapsed relevant to participants.

	Phase Two Study (N=57) (1996 to 2002)			The Combined Study (N=93) (1993 to 2002)		
	Lower	Mean	*Upper*	*Lower*	Mean	*Upper*
Age at SMCC	*29.5*	32.5	*36.7*	*29.5*	32.4	*36.7*
Age at AIB	*16.8*	19.8	*28.3*	*16.3*	19.8	*30.8*
Time between AIB and SMCC	*7.8*	12.6	*15.8*	*5.8*	12.7	*15.8*

Table 4.6 shows the overall Pass-Fail course results for various time periods. The table also shows that there are incomplete records and missing data in the research. The results and grades for four students in phase two of the research were, inadvertently, not collected and the results for one student in phase one was omitted; he had left the course for other reasons than failure.

Table 4.6: Pass-Fail Results for the SMCC for different time periods.

Time Period	Number of SMCC Students	Pass	Fail	Pass Rate
1979 to 1992	188	132	46	70.2%
1993 to 1995 (first phase)	36	26	10	72.2%
1996 to 2002 (second phase)	57 (but only have records for 52)	39	13	75.0%
Total for both phases of the research	93 (but only have records for 88)	65	23	73.9%

4.10 Materials

This section explains the way information was collected from the participants and starts with a description of the Admiralty Interview Board procedures and measures. This is followed by a description of the two inventories which were completed by students at the start of the SMCC course for phase two of the study; the OCEAN personality inventory and the Occupational Stress Indicator. Finally, the performance grades collected at the end of the SMCC course will be described.

4.10.1 The Admiralty Interview Board Measures

The information about the students' performance at the Admiralty Interview Board was derived from the Royal Navy Candidates' database. The information was collected as part of the normal selection process between May 1978 and July 1990 for the combined group. The reliabilities, intercorrelations and other properties of some of the AIB measures are given in Appendix G (p. 137). Validity coefficients have already been shown in Table 4.3 above (p.48).

Verbal ability test. This test is the GT90A which was published by Nelson/ase and introduced at the AIB in 1972 to replace an earlier version (GT35). This test was originally developed by the National Institute of Industrial Psychology, and comprised four parts; synonyms and antonyms, analogies, jumbled sentences, and completing sentences. It lasted for 20 minutes.

Non-verbal ability test. This test (GT70/23) was introduced in 1978 and is in two parts: testing matrix completion and sequencing. It is similar to the familiar standard and advanced progressive Matrices non-verbal test devised by Raven (1963). This lasts for 13 minutes and is a measure of intelligence only minimally influenced by verbal aptitude and educational background. This too was obtained from Nelson/ase.

Numeracy test. This 25 minute test (SP225) was introduced in 1978 and was devised in-house. The Numeracy test has three sub-tests which look at: fluency (basic arithmetical approximations), problems (reasoning with algebra), interpretation (statistical information from graphs and tables).

Instructions test. This is a clerical checking and instruction test (SP21) lasting 15 minutes. Candidates have to carry out clerical operations by checking, filing, classifying and coding a list of 30 stores items. This test dates from the original AIB in 1947. The instructions for this test are quite complex, so a tape recording of the instructions is used to ensure standardisation.

A Composite Test Score is produced using the four tests just described. Research (DERA, 1999) showed that this was the best overall predictor of training performance at Britannia Royal Naval College, Dartmouth, and that an equal weighting of the four individual test scores produced the best prediction.

Biodata. This item (the Professional Aptitude Predictor or PAP) was introduced towards the end of 1989 following a review and update of the AIB procedures by Birkbeck College, University of London. This was reported by Jones *et al.* (1991). Only two students out of the 36 in phase one of the study had a biodata score, but 27 students out of the phase two study group of 57 had this score, having attended the AIB after its introduction. The predictors used in scoring the biodata are given in Appendix H (p. 138). The strong predictors are all academic qualifications. It is notable that GCE results, particularly GCE Maths and Physics, are more predictive than A Level results.

Gym task and the group Discussion Planning exercises. The other major assessment features of the AIB are the marks given by the four Board Members for the candidates'

performance on (1) four practical Gym tasks and (2) a group Discussion Planning exercise. The marks for these two exercises provide an average of the four Board Members individual marks.

Headteachers' References. In a research study for the Admiralty Interview Board, Jones and Harrison (1982) tried to make the references from head teachers more valid. They designed a report form which defined the competences the referees were to use and provided a behaviourally-anchored rating scale for each attribute. The research showed that the resulting references were a great improvement on the previous scoring system, where the AIB Board Members gave a score for a purely narrative reference. For obvious reasons, references from head teachers were only asked for up to two years after a candidate finished school.

The scored sections of the reference reported on seven aspects of school life:

- Application to studies (Application),
- Involvement in clubs, societies, sports, etc. (Involvement),
- Discharge of responsibility (Responsibility),
- Strength of character (Character),
- How well respected by contemporaries (Relationships),
- Influence and leadership of contemporaries (Influence),
- Overall contribution to school/college (Contribution).

These items are scored on a 1 to 9 scale.

The four board dimensions and the Final Board Mark. In a recent review of assessment centre research Arthur and Day (2011) recommend that researchers should investigate data at this dimension level rather than just the overall board mark. The four dimensions of the Admiralty Interview Board are: Effective Intellect, Leadership Potential, Character and Personality, and Service Motivation. Anderson *et al.* (1994) showed that assessment centre selectors are overloaded with information so tend to assess on the overarching dimensions, but weigh the observational sources too heavily. So it will be interesting to see if these dimensions are better predictors than the Final Board Mark.

Other miscellaneous information. GCE Maths and Physics results collected at the AIB were also included in the analysis.

4.10.2 The personality inventory: OCEAN

The OCEAN personality inventory was originally devised by Tupes and Christal in 1961 (see chapter 2). Lord and Rust (2003) note that Tupes and Christal factor-analysed results from self-ratings, peer-ratings and supervisor ratings 'and found that a five factor solution generated strong and recurrent factors across all these types of assessment' (p.

16). The original inventory was revived in the early 1990s with the emergence of the five factor model of personality. An anglicized version of the Tupes and Christal inventory was developed and trialled and found to have a similar structure to the original US version (Collis, 1997)[17]. This inventory was named OCEAN by UK defence psychologists and it measures the five factors of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism.

The inventory is split into two sections; one which measures traits and a second which measures self-descriptions. The trait part of the questionnaire consists of a list of 64 traits, for example: 'inventive', 'responsible', 'bashful', 'affectionate', 'touchy', etc. The person completing the trait inventory has to use a nine-point Likert scale representing how characteristic the trait is for them. The scale ranges from 'extremely uncharacteristic' (1) through 'neutral' (5) to 'extremely characteristic' (9).

The self-description part has 110 statements such as 'I am a very shy person', 'I tend to be a loner' (to quote the two shortest statements). Here the test taker has to write down a number next to the statement which represents how much they agree with the description. For the self-descriptions there is a seven-point Likert scale which goes from 'very strongly disagree' (1) through 'neutral' (4) to 'very strongly agree' (7). To investigate if traits or self-descriptions give a different prediction, the two OCEAN measures will be investigated in the analysis in both separate and combined forms.

The combined OCEAN inventory can also be separated into lower order factors or sub-composites which were extracted by Christal (Collis, 1997). Table 4.7 shows a breakdown of the five OCEAN personality factors into the subcomposites. Factor analysis showed that the subcomposites load only on their main factors except for 'Socially Active' which loads slightly less on Agreeableness. The literature review presented earlier noted that these subcomposites may be better predictors of performance than the five wider personality dimensions. An analysis will be undertaken to see if these subcomposites are better at predicting outcomes than the five main dimensions in specific areas like the grades given for individual aspects of performance.

---

[17] This material is reproduced with permission from the Defence Science and Technology Laboratory.

Table 4.7: The subcomposites of the OCEAN Personality Inventor (adapted from Collis, 1997, p. 12). [Adapted and reproduced with permission from the Defence Science and Technology Laboratory]

The subcomposites of the OCEAN Personality Inventory		
Openness	**Conscientiousness**	**Extraversion**
Philosophical	Efficient and Dependable	Shy and Bashful
Scientific Interest	Hard Working	Talkative
Creative	Organized	Socially Active
Reflective		Assertive
Cultured		Unsociable
Agreeableness	**Neuroticism**	
Warm and Sympathetic	Nervous and Stresses Out	
Friendly	Worrying	
Considerate	Irritable	
Cold and Insensitive	Envious and Jealous	
Helpful		

Collis (1997) described the development of OCEAN and the technical background to the reliability and validity of both the US and UK versions. One study which was carried out in the UK included over 600 Royal Naval officer applicants at the Admiralty Interview Board. Collis reports that the split-half reliabilities for the five factors of the UK version were between 0.83 and 0.88 (0.91 and 0.94 when corrected). The test-retest reliabilities over a nine-month period for 88 successful AIB candidates joining the Royal Navy are shown in Table 4.8. The test-retest reliabilities for four of the five dimensions were above 0.7; but the test-retest figure for Agreeableness was only 0.58.

Table 4.8: Test-retest reliability for Royal Navy officers (adapted from Collis, 1997, p. 22). [Adapted and reproduced with permission from the Defence Science and Technology Laboratory]

Personality Factor	*Test-retest Reliability* *(all significant $p < 0.05$)*
Openness	0.79
Conscientiousness	0.71
Extraversion	0.75
Agreeableness	0.58
Neuroticism	0.75
N = 88. Time between testing was nine months.	

Collis gives test-retest data for longer periods and describes various research studies which examined the construct, concurrent and predictive validity of the OCEAN inventory. She concludes that the reliability 'is very acceptable and comparable with

published retest data for commercially available personality instruments' (p. 23). The OCEAN personality inventory takes about 40 minutes to complete.

4.10.3 The Occupational Stress Indicator

Hart and Cooper (2001) note that the definition of stress in terms of the interaction between person, the environment or both, means that stress cannot be measured by a single variable. They also note that stress is a 'relatively abstract construct' (p. 98) and is not measured directly but by other variables such as coping and personality. For these reasons Cooper *et al.* (1988) devised the Occupational Stress Indicator (OSI) which has been used in the diagnosis and management of stress in the workplace and for research purposes. It has been used with a wide range of occupational groups. The OSI follows Cooper's organizational model of stress (Cooper, 1986) which is based on the transactional model of stress outlined in Chapter 3 (Section 3.5).

The Occupational Stress Indicator (OSI) was devised specifically for management use rather than as a research instrument. The eight pages of questions are divided into six separate questionnaires. Table 4.9 shows how these six questionnaires relate to the four elements of the model of stress: sources of stress, individual reactions, the effects of stress on the individual and organization and coping with stress. The questionnaire titles appear at the top of the pages to the person completing the OSI. The health questionnaire is divided into two parts investigating physical health and mental health.

Table 4.9: The OSI stress model elements and the six questionnaires.

The four elements of the model of stress	OSI Questionnaire title
Sources of stress	How you feel about your job
The individual	How you assess your current state of health: • Part A. How you feel or behave • Part B. Your physical health
	The way you behave generally
	How you interpret events around you
Effects of stress	Sources of pressure in your job [two pages]
Coping with stress	How you cope with stress you experience

With the exception of the heath questionnaires the questionnaire items in the other five questionnaires outlined in Table 4.9 are broken down further into the 28 scales listed in Table 4.10. The list also includes three *'broad view'* scales, questionnaires, based on selected items from the other scales within a particular questionnaire theme. For

example, the *'J6 Broad view of job satisfaction'* scale is scored using five items, one from each of the other five 'How you feel about your job' scales.

Evers *et al.* (2000) noted that the OSI is a popular research instrument, which published at least 38 articles in journals between 1990 and 1997. However, although questionnaires like those in the OSI have been designed to help organizational interventions, Brinner and Reynolds' (1999) review of the evidence could not find any support for the assertions made that these models do actually help.

Table 4.10: The OSI Questionnaires and subscales.

Questionnaire title	OSI scale *[Note the "Broad View" for three questionnaires]*
How you feel about your job	J1 Satisfaction with achievement, value and growth
	J2 Satisfaction with the job itself
	J3 Satisfaction with organisational design and structure
	J4 Satisfaction with organizational processes
	J5 Satisfaction with personal relationships
	J6 Broad view of job satisfaction
How you assess your current state of health	H1 Mental ill health
	H2 Physical ill health
The way you behave generally	B1 Attitude to living
	B2 Style of behaviour
	B3 Ambition
	B4 Broad view of Type A
How you interpret events around you	I1 Organisational forces
	I2 Management processes
	I3 Individual influence
	I4 Broad view of control
Sources of pressure in your job	S1 Factors intrinsic to the job
	S2 The managerial role
	S3 Relationships with other people
	S4 Career and achievement
	S5 Organisational structure and climate
	S6 Home/work interface
How you cope with stress you experience	C1 Social support
	C2 Task strategies
	C3 Logic
	C4 Home and work relationship
	C5 Time
	C6 Involvement

4.10.4 Submarine Command Course Performance Measures

In addition to the Pass-Fail Result for the course, an overall performance grading scheme was devised specifically for this research to try to show more gradations of performance. This overall performance grade or Course Grade, which runs from A (top

performers) to F (poorest performers), is shown in Table 4.11. Students who pass the course are graded A to D; those who fail are assigned grades E or F.

Table 4.11: Overall performance grade (Course Grade).

Overall Performance Grade (Course Grade)	
A - Exceptional	(Very promising career prospects)
B - Good	(Promising career prospects)
C - Average	(Likely to be reliable)
D – Just Pass	(Some shortcomings)
E - Fail	(Unable to fulfil SMCC requirements)
F – Severe Fail	(Unlikely to succeed in General Service)

Table 4.12 shows a breakdown of the Course Grades given to students on Phase 2 of the research and for the combined research study. The most used grades were the pass grades B and C, followed by the fail grade E. This performance measure is not normally distributed, but skewed and bimodal, as shown in Figure 4.1.

Table 4.12: Overall Course Grade for students on the SMCC.

Course Grade and Description	Phase 2 of the Research		Phase 1 and Phase 2	
	Number	*Percent*	*Number*	*Percent*
A - Exceptional	3	5.7	6	6.7
B - Good	16	30.2	31	34.8
C - Average	16	30.2	22	24.7
D - Just Pass	5	9.4	7	8.67
E - Fail	12	22.6	18	20.2
F - Severe Fail	1	1.9	5	5.6
Total	53	100	89	100
(Missing Course Grade)	(4)		(4)	

Figure 4.1: Histogram of the Course Grades (N = 89).
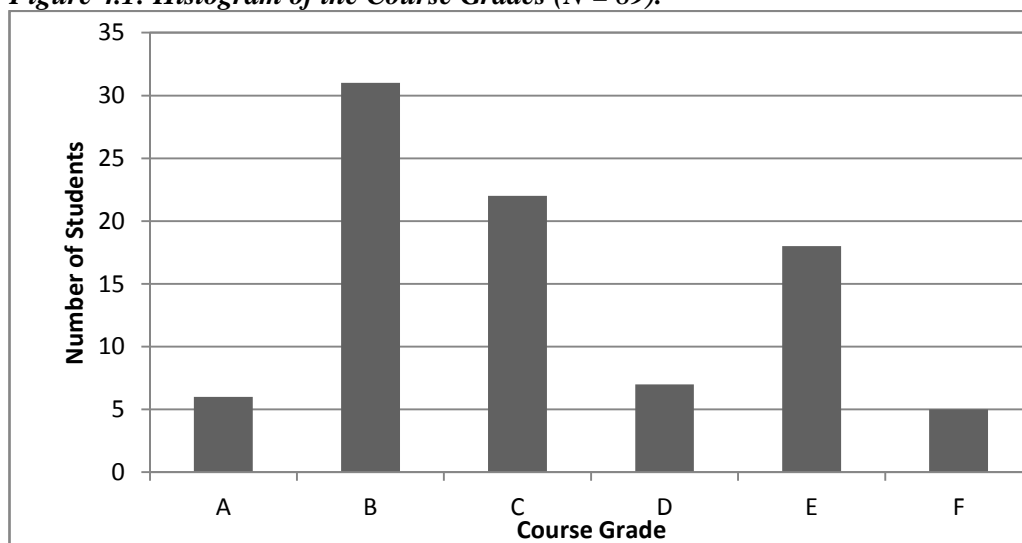
To help develop the performance grades list of three determinants and eight performance components proposed by Campbell *et al* (1993) were used for bringing together a complete model of job performance. The three important determinants are:

1. Declarative knowledge: factual knowledge and understanding of things that must be done and knowing what to do.
2. Procedural knowledge: able to perform the tasks and the skills in knowing how to do things and,
3. Motivation: the direction, degree and persistence of effort in doing them.

The first two of these determinants were also used by Kyllonen and Christal's (1990) theory of links between the psychometric and informational processing models, but they did not use "motivation" and added two forms related to memory: processing speed and working memory (Kyllonen, 1966).

  Campbell *et al.* also listed eight components of job performance:
1. Job-specific task proficiency.
2. Non-job-specific task proficiency.
3. Written and oral communications.
4. Demonstrating effort.
5. Maintaining personal discipline.
6. Facilitating peer and team performance.
7. Supervision/leadership.
8. Management/administration.

The last two components in this list usually occur in face to face communication, but the others may not. Campbell *et al.* (1990) noted that five of the eight dimensions were found in a sample of military jobs.

Campbell *et al.* (1993) argue that the three most important components of every job are task-proficiency, demonstrating effort and maintaining personal discipline, though not every component may be relevant to all jobs; for example, not all jobs require management or communication. Campbell *et al.* (1996) argue that performance is not under the total control of the individual. Organizational factors like work group, management and external factors such as demand for a product or geography can also contribute to a person's job performance.

To get a clearer understanding of the underlying performance factors, the author of this thesis used a critical incident technique (Flanagan, 1954) and a repertory grid technique (Kelly, 1955) to develop the important performance criteria. These two techniques are often used in job analysis. The author interviewed the two 'Teachers' who were in charge of the first two courses during the first phase of the study and developed a grading sheet of attributes that emerged as important performance criteria on the course.

Campbell *et al.*'s scheme was used to try to cover the important aspects of performance on SMCC. The resulting Twenty Quality Grades, shown in Table 4.13 cover twenty aspects of performance: eight grades on aspects of Tactical performance, three grades

on Administrative performance and nine grades on Personality qualities. It may appear strange to see aspects like maturity, sense of humour, and bearing and dress appearing in the list, but these were all aspects of Rodger's (1952) 'seven-point plan' to aid selection. The students were graded from 0 to 3 on each aspect.

Table 4.13: Twenty Quality Grades.

Twenty Quality Grades		
Tactical	*Administration*	*Personality*
Professional Knowledge	Management Ability	Leadership
Tactical Awareness	Staff Work	Stamina
Practical Ability	Use of English	Command Presence
TWS Technical Knowledge		Bearing and Dress
Navigation and Pilotage		Intelligence
Instinctiveness (Safety)		Common Sense
Courage		Maturity
Caution		Charm/Sense of Humour
		Honesty

The SMCC Course Assessment sheet is given in Appendix I (p. 139). Table 4.14 shows the mapping between the Twenty Quality Grades and Campbell *et al.*'s components and dimensions to show how these are covered by the assessment.

Schmitt *et al.* (2010) argue that underlying performance constructs should be specified as carefully as possible, particularly where there are contextual dimensions and best practice suggests that it would have been useful to produce an assessment scheme with behaviourally anchored rating scales. Viswesvaran (2002) credits Smith and Kendall (1963) with proposing the use of BARS designed to tie the level of performance to a particular grade by having a common framework of reference to which each scale point refer; but the assessment recording method for this research had to be simple and easy to use, so this line was not pursued. In a review of different types of ratings and their quality Viswesvaran and Ones (2002) noted that research suggests that the type of rating does not make a large difference to the quality of the ratings.

The end of course assessment was completed for each student, but not for the first two courses, and data was missed for one later course of four students. Four commanding officers were responsible for completing the assessment during the nine year period when this data was collected.

Table 4.14: The links between the Twenty Quality Grades and Campbell et al.'s components and dimensions.

Campbell et al.'s eight components of job performance	Twenty Quality Grades	Important Dimensions
1. Job-specific task proficiency*	Professional Knowledge	Declarative Knowledge
	Tactical Awareness	Procedural Knowledge
	Practical Ability	Procedural Knowledge
	TWS Technical Knowledge	Declarative Knowledge
	Navigation and Pilotage	Procedural Knowledge
	Instinctiveness (Safety)	Procedural Knowledge
	Caution	Procedural Knowledge
2. Non-job-specific task proficiency*	Staff Work	Procedural Knowledge
	Intelligence	Declarative Knowledge
	Common Sense	Procedural Knowledge
3. Written and oral communications	Use of English	Procedural Knowledge
4. Demonstrating effort*	Stamina	Motivation
	Courage	Motivation
5. Maintaining personal discipline*	Command Presence	Motivation
	Bearing and Dress (Maturity)	Motivation Motivation
6. Facilitating peer and team performance	Charm/Sense of Humour (Honesty)	Motivation Motivation
7. Supervision/leadership	Leadership	Motivation
8. Management/administration	Management Ability	Procedural Knowledge/ Motivation
* Important components in every job.		

Figure 4.2 presents an elementary linkage analysis (McQuitty, 1957) of the Twenty Quality Grades produced from the inter-correlations. This shows four clusters and one independent grading. The main cluster of ten gradings, with Tactical Awareness, Professional Knowledge, Practical Ability and Instinctiveness (Safety) at the centre have links to other clusters. Thus Leadership links to another cluster with Management Ability and this cluster links to another cluster around Use of English. Command Presence links the main cluster to another headed by Charm/Sense of Humour. Bearing and Dress is by itself, linked to Professional Knowledge.

Figure 4.2: Major links between the Twenty Quality Grades.

Correlations are shown without the decimal points.
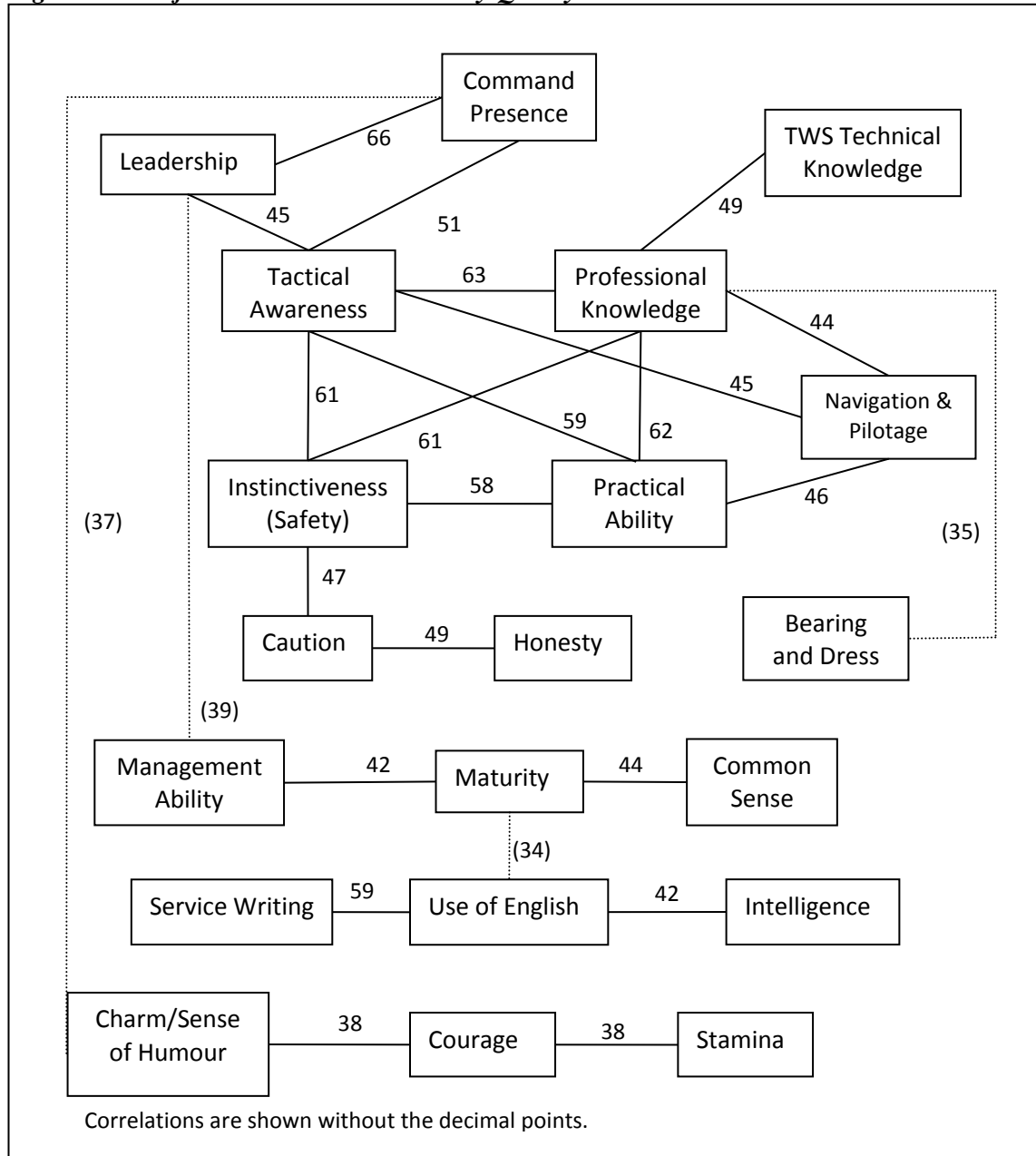
4.11 Procedures

The OCEAN personality inventory and OSI were sent to the course instructor a week before the course started. He briefed the students on the research and gave each student an envelope containing the inventories and a personal letter from the author about the research. The students were reassured that the information would not be used by anyone else in the Royal Navy and that their names would be removed and a code used to assure anonymity. The students were asked to complete the two inventories during the first evening of the course. The envelopes were collected by the author at a session the next afternoon when the author presented two workshops; one on the psychology of

management and one on stress and stress management. One student on the first course of phase one did not wish to complete the spatial tests; otherwise, the rest of the material was completed fully except for one student who only completed the first page of the OSI (he probably didn't turn over from the first page!).

4.12 Analyses Undertaken

The data was analysed using the *Statistical Package for the Social Sciences* (SPSS) version 14. The results are presented in the next chapter of the thesis in eight sections as outlined below:

- Analysis A: The AIB measures and the Pass-Fail Result and the Course Grade (Section 5.2),
- Analysis B: The AIB measures and the Twenty Quality Grades (Section 5.3),
- Analysis C: The OCEAN personality inventory and the Pass-Fail Result and the Course Grade (Section 5.4),
- Analysis D: The OCEAN personality inventory and Twenty Quality Grades (Section 5.5),
- Analysis E: The Occupational Stress Indicator and the Pass-Fail Result and the Course Grade (Section 5.6),
- Analysis F: The Occupational Stress Indicator and the Twenty Quality Grades (Section 5.7),
- Analysis G: Miscellaneous AIB items and Pass-Fail Result, the Course Grade and the Twenty Quality Grades (Section 5.8).
- Analysis H: Links between the AIB assessment measures and the personality dimensions (Section 5.9).

Since many of the measures have a small range of scores, their distributions may not be normally distributed like the Course Grades. So the non-parametric correlation statistic, Spearman's rho (ρ), was used. Siegel (1956) notes that the efficiency of the Spearman ρ method is about 91 percent when compared to the most powerful parametric correlation, the Pearson product-moment correlation method. This means that, if the Spearman method can find an association in a set of normal data containing a hundred cases, the Pearson method would find 91 cases.

4.13 The Limitations of the Research

One of the major problems with the research is that the group of students investigated are a highly selected and well-trained group of people, so any validity coefficients produced may be low. However, the literature review shows that. whilst there may be some clear results both for cognitive tests and personality inventories for research in selecting junior staff, the results for more senior managers are less clear and sometimes

in the opposite direction (for example, for Neuroticism and job performance). The low number of items which make up some of the scales of the OSI probably means that these scales are not very reliable; and this comment also applies to the sub-composites of the OCEAN personality inventory. Other major limitations are the performance measures, which have only a four-point scale, and the possible influence of the four different people, who were in charge of the course at different times, and also assessed the students. There is no way of assessing the inter-rater reliability of these assessors.

4.14 Chapter Summary

This chapter has set out the context in which the original selection assessment took place, the Admiralty Interview Board (AIB). The Submarine Command Course, which the students attended on an average thirteen years after attending AIB, was also described. The various measures used in the research, including performance measures, were explained. The next chapter presents the results of the analysis.

5 Results

5.1 Introduction

This part of the thesis presents the results of the analyses of the various measures which might predict success on the Submarine Command Course (SMCC). The correlations and other analyses of how the **AIB measures**, the **OCEAN personality inventory** and the **Occupational Stress Indicator (OSI)** correlate with the course results and other grades are reported. This analysis plan was outlined in Section 4.12 (p. 66) of the previous chapter and is also given here in Table 5.1 in extended form.

Table 5.1: A breakdown of the analysis reported.

Possible Predictor Measure	Results Section	SMCC Course Assessment
Admiralty Interview Board (AIB) Psychometric and other measures	**Analysis A 5.2**	*Pass-Fail Result Course Grade (A to E)*
	Analysis B 5.3	*Twenty Quality Grades: Total Score Total Tactical Total Admin 20 Individual Grades*
The OCEAN Personality Inventory	**Analysis C 5.4**	*Pass-Fail Result Course Grade (A to E)*
	Analysis D 5.5	*Twenty Quality Grades: Total Score Total Tactical Total Admin Total Personality 20 Individual Grades*
The Occupational Stress Indicator (OSI)	**Analysis E 5.6**	*Pass-Fail Result Course Grade (A to E)*
	Analysis F 5.7	*Twenty Quality Grades: Total Score Total Tactical Total Admin Total Personality 20 Individual Grades*
Miscellaneous AIB measures *Professional Aptitude Predictor (PAP) The four assessment centre dimensions The Headteacher's Reference score*	**Analysis G 5.8**	*Pass-Fail Result Course Grade (A to E) Twenty Quality Grades: Total Score Total Tactical Total Admin Total Personality 20 Individual Grades*
OCEAN personality inventory and the AIB Measures	**Analysis H 5.9**	Not used in this analysis

The analysis is presented in eight sections and the first six of these sections are in pairs. The first section of each pair (Analyses A, C and E) examines if the selected measure can predict the overall course results (the *Pass-Fail Result* and the *Course Grade*). The following section of each pair (Analyses B, D and F) reports the correlations between

the measures and the ***Twenty Quality Grades***. These latter analyses are broken down into two parts. The first part examines the relationship between the measure and the *Total Score* of the ***Twenty Quality Grades*** and the three *subtotals* which score the tactical, administrative and personality aspects of performance. Following this, the links to the *individual Twenty Quality Grades* are explored. The next section (Analysis G) examines some miscellaneous data which were collected at AIB as part of the assessment centre process: *Professional Aptitude Predictor (PAP)*, the *four assessment centre dimensions* and the *Headteacher's Reference* score. The final analysis, Analysis H, investigates any links between the **OCEAN personality inventory** and the **AIB measures.** The correlation coefficients given in the tables are Spearman rho (ρ).

To help the reader through this Results chapter and the following Discussion chapter the different levels of the measures will be given in different styles of type. At the top level, the **AIB predictor measures**, the **OCEAN personality inventory** and the **Occupational Stress Indicator** will be given in **bold** typeface. For the next level below these, items such as the ***Pass-Fail Result***, the ***Course Grade*** and the ***OSI subscales*** will be given in ***bold and italics***. At the next level, items like the *Verbal test,* the *Headteacher's Referecne, Total Score and Extraversion* will be given in *italics*. The last level will be in plain text, for example, Command Presence, 'Ambition'.

5.2 Analysis A: The AIB Measures and the Pass-Fail Result and the Course Grade

Before looking at the actual results of the **AIB measures** as predictors of success on the Submarine Command Course the intercorrelations between the various AIB measures will be examined. The **AIB psychometric measures** are a *Verbal test*, a *Non-verbal test*, a *Numerical test* and an *Instructional test*. These are combined to give a *Composite Test Score* made up from the four psychometric tests. In addition to this, the candidates' scores on the practical *Gym Tasks* and on the *Discussion Planning Exercise*, as well as the overall *Final Board Mark* are included and the candidates' *Maths* and *Physics* GCE grades are also included. The intercorrelations between these measures are shown in Table 5.2. Since the **AIB measures** would be expected to correlate with each other and predict success on the course, a one-tailed test was used.

It can be seen from the table that many of the predictors are correlated significantly with each other. The ***psychometric tests*** are correlated with each other above the 0.3 level. The *Final Board Mark* correlates with the *Verbal, Non-verbal* and *Numerical test* at around the 0.2 level, but the *Instructions test* is not correlated. The *Final Board Mark* correlates with the other selection procedures which are observed by the Board

Members at a much higher level than the psychometric tests; the *Gym Task mark* at around 0.4 and the *Discussion mark* approaching 0.5.

It is interesting that both the *Maths* and *Physics grades* correlate with the *Verbal test* and *Numerical test*, but not with the *Non-verbal test* or *Instructional test*. Another finding is that, with one exception, the *Gym* and *Discussion marks* did not correlate with the **psychometric tests**. The exception is the *Gym mark* which is correlated with the *Non-verbal test*. There are no significant negative correlations in this table.

Table 5.2: Intercorrelations between the AIB measures.

AIB Predictor	1	2	3	4	5	6	7	8	9	10
1 *Verbal ability test*		0.31 ** (93)	0.49 ** (91)	0.38 ** (92)	0.71 ** (90)	-0.16 (93)	0.03 (93)	0.24 * (93)	0.26 * (76)	0.41 ** (72)
2 *Non-verbal ability test*	0.31 ** (93)		0.47 ** (91)	0.32 ** (92)	0.70 ** (90)	0.18 * (93)	0.14 (93)	0.18 * (93)	0.18 (76)	0.19 (72)
3 *Numerical test*	0.49 ** (91)	0.47 ** (91)		0.36 ** (90)	0.85 ** (90)	-0.11 (91)	0.06 (91)	23* (91)	0.40 ** (76)	42** (72)
4 *Instructions test*	0.38 ** (92)	0.32 ** (92)	0.36 ** (90)		0.64 ** (90)	-0.08 (92)	-0.07 (92)	-0.05 (92)	0.15 (75)	12 (71)
5 *Composite test score*	0.71 ** (90)	0.70 ** (90)	0.85 ** (90)	0.64 ** (90)		-0.09 (90)	0.03 (90)	0.20 * (90)	0.36 ** (75)	0.43 ** (71)
6 *Gym tasks mark*	-0.16 (93)	0.18 * (93)	-0.11 (91)	-0.08 (92)	-0.09 (90)		0.23 * (93)	0.39 ** (93)	-0.01 (76)	-0.04 (72)
7 *Discussion mark*	0.03 (93)	0.14 (93)	0.06 (91)	-0.07 (92)	0.03 (90)	0.23 * (93)		0.47 ** (93)	-0.15 (76)	-0.04 (72)
8 *Final Board Mark*	0.24 * (93)	0.18 * (93)	0.23 * (91)	-0.05 (92)	0.20 * (90)	0.39 ** (93)	0.47 ** (93)		0.06 (76)	0.23 (72)
9 *Math grade*	0.26 * (76)	0.18 (76)	0.40 ** (76)	0.15 (75)	0.36 ** (75)	-0.01 (76)	-0.15 (76)	0.06 (76)		0.48 ** (71)
10 *Physics grade*	0.41 ** (72)	0.19 (72)	0.42 ** (72)	0.12 (71)	0.43 ** (71)	-0.04 (72)	-0.04 (72)	0.23 (72)	0.48 ** (71)	
* $p < 0.05$; ** $p < 0.01$ (one-tailed). The figures in brackets indicate the number of candidates, e.g. N = 93 is shown as (93).										

A non-parametric test of difference between means of independent samples, the Mann-Whitney U test, was carried out to look at the difference between the means of the various **AIB predictors** and the ***Pass-Fail Result*** status of students who completed the SMCC course. Of the ten predictors examined in this section only one, the *Non-verbal test*, had a statistically significant result [U = 590, N = 88, p = 0.047 (one-tailed), *d* =

0.4]. The mean score of the pass group (64 students) was 39.8 (SD = 6.1) and the mean of those who failed (24 students) was 37.5 (SD = 5.8). Although Cohen's *d* was developed to give an estimate of the effect size for the parametric t-test (Cohen, 1992) the figure of 0.4 for *d* shows that this difference in means is a small to moderate effect size[29].

Table 5.3 displays the correlations found between the **AIB predictors** and the *six-point Course Grade* (see Table 4.11, p. 61). The only statistically significant correlation found was for the *Non-verbal test* with a correlation 0.20. The other correlations were below 0.13.

Table 5.3: Correlations between the AIB measures and the Course Grade.

AIB Predictor	Correlation with Course Grade
Verbal ability test	-0.05 (88)
Non-verbal ability test	0.20* (88)
Numerical test	0.02 (86)
Instructions test	0.13 (87)
Composite test score	0.09 (83)
Gym task mark	-0.02 (88)
Discussion mark	-0.11 (88)
Final Board Mark	-0.07 (88)
Maths grade	0.06 (72)
Physics grade	0.05 (67)
* p < 0.05. The figures in brackets indicate the number of students.	

5.3 Analysis B: The AIB Measures and Twenty Quality Grade

At the end of the SMCC course, the course instructor rated each student on *Twenty Quality Grades* (these are given in Table 4.13, p. 63). The intercorrelations of the *Twenty Quality Grades* are shown in two tables, Table 5.4 (first part) and Table 5.5 (second part) together with the *Total Score* obtained by adding up the marks for all the *Grades* for each student. The Tables show that most items are correlated with each other and with the *Total Score*. In particular, the eight *Tactical Grades* are mostly highly correlated with each other. The exceptions to this pattern are the three *Administration Grades* which mainly correlate with each other. The Staff Work and Use of English grades are the only ones which do not correlate with the *Total Score*. None of the negative correlations are statistically significant.

The Course Instructor completed the *Twenty Quality Grades* at the same time as giving each student an overall A to F *Course Grade*. It is likely that the *Twenty Quality Grades* were influenced by the overall *Course Grade* awarded rather than the other way round. These two measures have a correlation of 0.85 (p < 0.001). In addition to the

---

[29] Cohen's *d* was calculated using a pooled variance method

Total Score, the **Twenty Quality Grades** were subdivided into three areas of performance: tactical, administration and personality and these have a *subtotal* score designated in this report as *Total Tactical, Total Admin and Total Personality*.

Table 5.4: Intercorrelations between the Twenty Quality Grades (first part).

	Total	PK	TA	PA	TWS	N&P	I(S)	Co	Ca	MA
Tactical										
Professional Knowledge	67**									
Tactical Awareness	68**	63**								
Practical Ability	61**	62**	59**							
TWS Technical Knowledge	50**	49**	38**	35**						
Navigation and Pilotage	52**	44**	45**	46**	16					
Instinctiveness (Safety)	75**	61**	61**	58**	27*	36**				
Courage	41**	11	22*	08	11	09	30**			
Caution	43**	29**	26**	21*	14	14	47**	04		
Administration										
Management Ability	26*	-10	-01	03	-08	03	06	09	02	
Staff Work	19	09	-10	05	10	01	12	-15	20*	21*
Use of English	09	-17	20*	-04	-18	-01	-02	-05	06	34**
Personality										
Leadership	57**	34**	43**	36**	25*	23*	34**	30**	14	39**
Stamina	32**	24*	22*	22*	17	27*	06	38**	-03	-16
Command Presence	67**	33**	51*	31**	38**	32**	37**	37**	03	31**
Bearing and Dress	41**	35**	31**	08	26*	14	09	14	12	18
Intelligence	42**	23*	14	30**	22*	24*	38**	-01	17	15
Common Sense	48**	23*	13	22*	35*	30**	28**	14	12	15
Maturity	51**	05	14	06	20*	19	20*	28**	25*	44**
Charm/Sense of Humour	34**	-02	18	-02	06	05	12	38**	19	06
Honesty	43**	26*	21*	13	19	18	35**	28**	49**	-09
Decimal points are omitted. * p < 0.05; ** p < 0.01. N = 75.										

Table 5.5: Intercorrelations between the Twenty Quality Grades (second part).

	SW	UoE	L	S	CP	B&D	I	CS	M	C/SoH
Use of English	59**									
Personality										
Leadership	-12	-13								
Stamina	22*	20*	09							
Command Presence	-11	7	66**	16						
Bearing and Dress	13	09	17	17	27**					
Intelligence	37**	42**	-09	-02	12	02				
Common Sense	08	10	17	10	35**	20*	34**			
Maturity	16	12	21*	10	30**	19	27**	42**		
Charm/Sense of Humour	-02	07	35**	15	31**	15	03	14	33**	
Honesty	15	-10	00	34**	05	03	13	15	26*	20*
Decimal points are omitted. * p < 0.05; ** p < 0.01 (one-tailed). N = 75.										

Table 5.6 shows the intercorrelations between the three *subtotals* and the *Total Score*. The *Total Tactical* score and *Total Personality* score are highly correlated with the *Total Score*, but the *Total Admin* score has only a weak correlation with the *Total Score*. The *Tactical* and *Personality subtotals* are moderately correlated with each other, the *Total Admin* score does not correlate with the other two *subtotals*.

Table 5.6: Intercorrelations between the Twenty Quality Grades Total Score and subtotals.

	Total Score	Total Tactical	Total Admin
Total Tactical	0.88**		
Total Admin	0.25*	-0.17	
Total Personality	0.90**	0.66**	0.16
* p < 0.05; ** p < 0.01 (one-tailed). N = 75.			

Table 5.7 examines the correlations between the *Total Score* and the three *subtotals* of the **Twenty Quality Grades** and the **Course Grade**. The table shows that the various *subtotals* correlate with the **Course Grade** as might be expected, but the *Total Admin* score is only weakly correlated.

Table 5.7: Correlations between the Course Grade and the Twenty Quality Grades Total score and subtotals.

	Total Score	Total Tactical	Total Admin	Total Personality
Course Grade	0.85**	0.81**	0.16*	0.74**
* p < 0.05; ** p < 0.01 (one-tailed). N = 75				

If the three subtotals are examined using the Mann-Witney U test to look at the difference in means between the pass and fail groups, then the same pattern of results is found as in Table 5.7. However, using this statistical method the means of the *Total Admin* score for the students who pass or fail is not statistically different. The difference in means for the *subtotals* of the *Tactical* and *Personality Quality Grades* are highly significant. These results are shown in Table 5.8.

Table 5.8: Difference between means of the Pass-Fail groups for the three Quality Grades sub-totals.

Significant difference in means	Means and (SDs)		Statistics
	Fail	Pass	
Total Tactical	12.5 (2.7)	17.9 (2.6)	U = 71; p < 0.001
Total Admin	6.0 (1.5)	6.0 (1.6)	U = 497; p = 0.37
Total Personality	16.1 (2.2)	20.1 (2.6)	U = 129; p < 0.001
Mann-Witney U test (one-tailed). N = 74 (19 Fail and 55 Pass).			
Figures in brackets following the means are standard deviations (SDs).			

Table 5.9 reports the correlations between the *Total Score* and the three *subtotal* scores of the *Twenty Quality Grades* (summarising eight *Tactical* items, three *Administration* items and nine *Personality* items) and the **AIB measures**. The main finding is that the

Non-verbal test correlates at 0.30 with the *Total Tactical* score. The *Verbal test* has a significant correlation (0.23) with the *Total Administrative* score which would be in line with prediction. However, the *Verbal test* shows negative correlations with the *Total Score* (-0.19) and with the *Total Personality* score (-0.22). A one-tailed test was used here since the direction of prediction was expected. If a data exploration technique had been used and a two-tailed test was employed then these negative correlations would not be significant. The other negative correlation found was between the *Final Board Mark* and the *Total Personality* score (-0.24) but this would still be significant if a two-tailed test had been used.

Table 5.9: Correlations between the AIB measures and the Twenty Quality Grades subtotals.

	Total Score	Total Tactical	Total Admin	Total Personality
Verbal ability test	-0.19* [0.05]	-0.16	0.23*	-0.22*[0.03]
Non-verbal ability test	0.17	0.30**	0.01	0.07
Numerical test	-0.07	-0.02	0.07	-0.12
Instructions test (N = 74)	0.06	0.11	-0.09	0.02
Composite test score (N = 74)	-0.03	0.06	0.08	-0.09
Gym task mark	-0.10	-0.06	0.03	-0.15
Discussion mark	-0.09	-0.06	-0.02	-0.10
Final Board Mark	-0.14	-0.11	0.16	-0.24* [0.02]
Maths grade (N = 71)	-0.04	-0.02	-0.01	-0.01
Physics grade (N = 66)	0.03	0.12	0.05	-0.08

* $p < 0.05$; ** $p < 0.01$ (one-tailed). N = 75 except where indicated. Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.

Table 5.10: Correlation between the AIB measures and the individual Twenty Quality Grades.

Verbal test		Non-verbal test	
Courage	-0.27* [0.00]	Professional Knowledge	0.35**
Staff Work	0.25*	Tactical Awareness	0.29**
Leadership	-0.22* [0.03]	Practical Ability	0.31**
Stamina	-0.22* [0.03]	Instinctiveness (Safety)	0.23*
Command Presence	-0.30**		
Numerical test		**Instructions test (N = 74)**	
Courage	-0.29**	Tactical Awareness	0.23*
		Caution	-0.20* [0.04]
Composite Test Score (N = 74)		**Discussion**	
Courage	-0.21* [0.03]	Caution	0.26*
Gym task		**Final Board Mark**	
Stamina	-0.20* [0.04]	Courage	-0.28**
Common Sense	0.20* [0.04]	Staff Work	0.21*
Maturity	-0.20* [0.04]	Leadership	-0.30**
		Command Presence	-0.26* [0.01]
Maths grade (N = 71)		**Physics grade (N = 66)**	
Management Ability	-0.23* [0.03]	Leadership	-0.32**
Leadership	-0.21* [0.04]	Command Presence	-0.27* [0.02]

* $p < 0.05$; ** $p < 0.01$ (one-tailed). N = 75 except where indicated. Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.

The final results given in this section are for the correlations between the **AIB measures** and the *individual Twenty Quality Grades*. As there is so much data from this analysis only the statistically significant correlations are reported in Table 5.10.

5.4 Analysis C: The Personality Inventory (OCEAN) and the Pass-Fail Result and the Course Grade

The personality inventory **OCEAN** measures the Big Five dimensions of *Openness, Conscientiousness*, *Extraversion, Agreeableness* and *Neuroticism*. This particular inventory has two separate inventories; one measuring **traits** and the other **self-descriptions** (see subsection 4.10.2, pp. 56-59 for a complete description). These two personality measures are combined to give a **composite score** for each personality dimension. Before looking at an analysis of the three inventories and the course results, the correlations within and between these three inventories will be presented.

Since the direction of intercorrelations between the five *OCEAN dimensions* was not as certain as for the other predictor measures used in this research, a two-tailed test of significance was used. Tables 5.11, 5.12 and 5.13 give the intercorrelations between the various **OCEAN versions**[30]. There are some discrepancies between the intercorrelations in the **composite** and the two other versions. There are only two correlations which can be seen across all three tables: *Extraversion* is highly correlated with *Agreeableness* and *Conscientiousness* has a high negative correlation with *Neuroticism*. *Openness* seems to be independent of the other dimensions except for a link to *Conscientiousness* in the composite version. In the first two tables, looking at the **composite** and **trait** versions, *Agreeableness* is linked with *Conscientiousness* and negatively linked with *Neuroticism*. *Extraversion* and *Conscientiousness* are only linked in the **self-description** version.

Table 5.11: Intercorrelations between the composite OCEAN dimensions.

	Openness	*Conscientiousness*	*Extraversion*	*Agreeableness*
Conscientiousness	0.30*			
Extraversion	0.04	0.23		
Agreeableness	0.09	0.34**	0.50**	
Neuroticism	0.07	-0.44**	-0.30*	-0.34*
* p < 0.05; ** p < 0.01 (two-tailed). N = 57.				

---

[30] To help distinguish more easily between the three versions of **OCEAN** the *composite* version has the straightforward dimension names, e.g. *Openness*, *Conscientiousness*, etc. The **trait** version dimension labels are suffixed with *T* and the **self-description** dimension labels with *SD*, e.g. *OpennessT*, *ExtraversionSD*, etc.

Table 5.12: Intercorrelations between OCEAN (trait version) dimensions.

	OpennessT	ConscientiousnessT	ExtraversionT	AgreeablenessT
ConscientiousnessT	0.20			
ExtraversionT	-0.05	0.13		
AgreeablenessT	-0.10	0.42**	0.45**	
NeuroticismT	-0.01	-0.33*	-0.29*	-0.51*
* p < 0.05; ** p < 0.01 (two-tailed). N = 57.				

Table 5.13: Intercorrelations between OCEAN (self-description version) dimensions.

	Openness SD	Conscientiousness SD	Extraversion SD	Agreeableness SD
ConscientiousnessSD	0.26			
ExtraversionSD	0.10	0.32*		
AgreeablenessSD	0.09	0.18	0.50**	
NeuroticismSD	0.17	-0.40**	-0.21	-0.14
* p < 0.05; ** p < 0.01 (two-tailed). N = 57.				

An analysis of the difference between the means of the *OCEAN dimensions* for the **Pass-Fail Result** showed no significant differences for any of the fifteen comparisons with the three different versions of **OCEAN**. (To save space the tables are not presented here but can be found in Appendix J, Section 1, p. 140.)

The next table, Table 5.14, gives the results of the correlations between the various *OCEAN personality dimensions* and the **Course Grade**. The table shows that no significant correlations were found with any of the fifteen dimensions.

Table 5.14: Correlations between the various versions of OCEAN and the Course Grade.

Correlations with the Course Grade	Version of OCEAN inventory		
	Composite	Trait	Self-description
Openness	-0.08	-0.10	-0.04
Conscientiousness	0.12	0.09	0.18
Extraversion	0.14	0.14	0.15
Agreeableness	0.05	0.08	0.03
Neuroticism	0.10	0.03	0.15
No significant results were found (two-tailed). N = 53.			

The next analysis investigated the difference between the means of the scores on the various **OCEAN subcomposites** (see Table 4.7, p. 58) on predicting the **Pass-Fail Result**. No significant differences were found in any of the twenty-two **subcomposites**. (To save space the tables are not presented here but can be found in Appendix J, Section 2, p. 141.)

The last analysis in this section investigated the correlation between the twenty-two **OCEAN subcomposites** and the overall **Course Grade**. None of the **subcomposites** correlated significantly with the overall **Course Grade**. (To save space the tables are not presented here but can be found in Appendix J, Section 3, p. 142.)

5.5 Analysis D: The Personality Inventory and Twenty Quality Grades

The first results to be presented in this section are the correlations between the three versions of the **OCEAN** inventory and the *Total Score* of the *Twenty Quality Grades*. These are given in Table 5.15. None of the correlations are statistically significant, though some are heading towards the 0.2 level.

Table 5.15: Correlations between the various versions of OCEAN and the Total Score for the Twenty Quality Grades.

Correlations with Total Score	Version of OCEAN inventory used		
	Composite	*Trait*	*Self-description*
Openness	-0.06	-0.17	0.03
Conscientiousness	0.17	0.18	0.17
Extraversion	0.13	0.12	0.16
Agreeableness	0.09	0.11	0.07
Neuroticism	0.13	0.07	0.17
No significant results were found (two-tailed). N = 53.			

The second analysis looked at the correlations between the scores on the various versions of the **OCEAN** inventory and the *subtotal* scores of the *Twenty Quality Grades*. The first table, Table 5.16, presents the results for the **composite version of OCEAN** and shows that there are no statistically significant findings.

Table 5.16: Correlations between the composite OCEAN dimensions and the Twenty Quality subtotals.

Composite version of OCEAN	Twenty Quality Subtotals		
	Total Tactical	*Total Admin*	*Total Personality*
Openness	-0.23	0.09	0.01
Conscientiousness	0.05	0.05	0.24
Extraversion	0.15	0.03	0.12
Agreeableness	-0.01	0.23	0.16
Neuroticism	0.08	0.06	0.15
No significant results were found (two-tailed). N = 53.			

Table 5.17 shows the only statistically significant correlation found in this set of three analyses of the various versions of **OCEAN** and the *Twenty Quality Grades subtotals* scores; *OpennessT* is negatively correlated with the *Total Tactical* section score with a figure of -0.34 on the *trait* version of **OCEAN**. In the previous table (Table 5.16) a correlation of -0.23 was observed, but this was not significant in this two-tailed condition.

Table 5.17: Correlations between the OCEAN (trait version) dimensions and the Twenty Quality subtotal.

Trait version of OCEAN	Twenty Qualities subtotals		
	Total Tactical	*Total Admin*	*Total Personality*
OpennessT	-0.34*	-0.13	-0.07
ConscientiousnessT	0.04	0.11	0.25
ExtraversionT	0.18	-0.02	0.09
AgreeablenessT	0.02	0.23	0.16
NeuroticismT	0.06	-0.01	0.08
* significant at the 0.05 level (two-tailed). N = 53.			

The last table in this series, Table 5.18, shows that there were no significant associations found between the *self-description* version of **OCEAN** and the *Twenty Quality Grades subtotal* scores.

Table 5.18: **Correlations between the OCEAN (Self-description) dimensions and the Twenty Quality subtotals.**

Self-description version of OCEAN	Twenty Qualities subtotals		
	Total Tactical	*Total Admin*	*Total Personality*
OpennessSD	-0.08	0.03	0.05
ConscientiousnessSD	0.07	0.01	0.22
ExtraversionSD	0.14	0.10	0.16
AgreeablenessSD	-0.03	0.21	0.15
NeuroticismSD	0.11	0.12	0.20
No significant results were found (two-tailed). N = 53.			

It should be noted that in each of the last three tables (5.16 to 5.18) that *Conscientiousness* is correlated with the *Total Personality* score above the 0.2 level but these figures are not significant using the nonparametric Spearman correlation method which was used throughout the analysis. However, the two highest correlations between *Conscientiousness* and *Total Personality* in these three tables are statistically significant if the parametric Pearson method is used. These two would also have been significant if a one-tailed test had been used.

The next analysis looks at the relationship between the various versions of **OCEAN** and the individual *Twenty Quality Grades*. Table 5.19 shows that there are some statistically significant correlations between these two measures for a few of the *Twenty Quality Grades* and some of these are about the highest seen in these analyses. All three versions of *Extraversion* correlate with Command Presence and all three versions of *Agreeableness* correlate with Use of English. Two measures of *Extraversion* correlate with Tactical Awareness, but this *Grade* also correlates negatively with the *trait version* of *Openness*. The *composite version* of *Openness* correlates negatively with Caution and Honesty. *Conscientiousness* has two links; with Command Presence and TWS Technical Knowledge. There are no links to any of the three versions of *Neuroticism*.

Table 5.19: Correlations between the OCEAN subcomposites and the individual Twenty Quality Grades.

OCEAN Dimension	Twenty Quality Grades	Correlation with OCEAN		
		Composite	Trait	Self-description
Openness	Tactical Awareness		-.29*	
	Caution	-.34*	-.40**	
	Honesty	-.32*	-.33*	
Conscientiousness	TWS Technical Knowledge			.29*
	Command Presence		.29*	
Extraversion	Tactical Awareness	.30*	.33*	
	TWS Technical Knowledge			.28*
	Command Presence	.39**	.36**	.40**
Agreeableness	Service Writing	.27*		
	Use of English	.36**	.36**	.34*
Neuroticism	none			
* p < 0.05; ** p < 0.01 (two-tailed). N = 53.				

The final analysis in this section investigates the relationships between the twenty-two *OCEAN subcomposites* and the *individual Twenty Quality Grades*. Firstly, Table 5.20 gives the statistically significant correlations between the *OCEAN subcomposites* and the **Twenty Quality Grades** *Total Score* and *subtotals*. This shows just four significant correlations with three for *Neuroticism* (Envious and Jealous) linked to the *Total Score* as well as *Total Tactical* (0.37) and *Total Personality*. *Openness* (Philosophical) is linked negatively to the *Total Tactical* score.

Table 5.20: Correlations between the OCEAN subcomposites and the Twenty Quality Grades Total Score and subtotals.

Openness (Philosophical)	
Total Tactical -.28*	
Neuroticism (Envious and Jealous)	
Total Score .31*	
Total Tactical .37**	
Total Personality .28*	
* p < 0.05; ** p < 0.01 (two-tailed). N = 53.	

The next table, Table 5.21, shows the statistically significant correlations between the twenty-two *OCEAN subcomposites* and the *individual Twenty Quality Grades*. Care needs to be taken when interpreting this table since the labels given to some of the *subcomposites* are negatively phrased, e.g., *Extraversion* (Shy and Bashful) and *Extraversion* (Unsociable). However, these *subcomposites* all contribute to the overall *Extraversion* score so the results need to be interpreted with the extravert component having the higher score. For example, the Unsociable end of the scale has a low score for *Extraversion*. It is interesting to note that there were no significant correlations found between any of the five *Conscientious subcomposites* which might have been predicted from the literature review.

The links to the various *Openness subcomposites* are in two directions with the two measuring 'Philosophical' and 'Creative' aspects having negative correlations. Of note is the high correlation between both *Openness* (Philosophical) and *Openness* (Creative) with the grading for Caution (-0.34 and -0.38, respectively) which might be expected. The link of *Openness* (Reflective) to the Common Sense grading and *Openness* (Cultured) to Use of English are again logical, but the link of this *subcomposite* to Navigation and Pilotage is puzzling. Could this be that the student is better able to express themselves and be understood when giving navigational commands?

Table 5.21: Correlations between the OCEAN subcomposites and the individual Twenty Quality Grades.

Openness (Philosophical)		*Openness* (Creative)	
Courage	-0.28*	Tactical Awareness	-0.30*
Caution	-0.34*	Caution	-0.38**
Honesty	-0.35**		
Openness (Reflective)		*Openness* (Cultured)	
Common Sense	0.32*	Navigation and Pilotage	0.29*
		Use of English	0.35*
Extraversion (Shy and Bashful)~		*Extraversion* (Talkative)	
Tactical Awareness	0.29*	Tactical Awareness	0.30*
Command Presence	0.40**	Command Presence	0.33*
		Honesty	-0.32*
Extraversion (Assertive)		*Extraversion* (Unsociable)~	
TWS Technical Knowledge	0.44**	Tactical Awareness	0.27*
Command Presence	0.30*	Command Presence	0.34*
Agreeableness (Warm and Sympathetic)		*Agreeableness* (Considerate)	
Use of English	0.36**	Staff Work	0.30*
Charm/Sense of Humour	0.28*	Use of English	0.36**
		Intelligence	0.28*
Agreeableness (Cold and Insensitive)~		*Agreeableness* (Helpful)	
Practical Ability	0.29*	Bearing and Dress	-0.29*
Use of English	0.31*		
Neuroticism (Worrying)		*Neuroticism* (Envious and Jealous)	
Maturity	0.28*	Professional Knowledge	0.32*
		Tactical Awareness	0.30*
		TWS Technical Knowledge	0.40**
		Instinctiveness (Safety)	0.29*
		Courage	0.34*
		Use of English	-0.28*
* p < 0.05; ** p < 0.01 (two-tailed). N = 53. ~ *Subcomposite* label is negatively phrased.			

Table 5.21 also shows that four *Extraversion subcomposites* have links to Command Presence and three to Tactical Awareness and the four *Agreeableness subcomposites* link to some qualities which might be expected like Charm/Sense of Humour. However, the *subcomposite* with the most links is *Neuroticism* (Envious and Jealous) with six links to the *Quality Grades*. There are no links between five of the OCEAN *subcomposites* and the *Quality Grades*. These are: *Openness* (Scientific Thinking), *Extraversion* (Socially Active), *Agreeableness* (Friendly), *Neuroticism* (Nervous and

Stressed Out) and *Neuroticism* (Irritable). One of those without any links is *Openness* (Scientific Interest) which, it might be thought, would link to the some of the *Grades* on the more technical aspects of performance.

To illustrate the coverage of the correlations the next table, Table 5.22, gives the same results which were displayed in Table 5.21, but the opposite way round, so that the number of significant correlation for each *Quality Grade* can be seen. Tactical awareness and Use of English have the highest number of links to various *OCEAN subcomposites* with five matches. Command Presence is next with four correlations which are all linked to various *Extraversion subcomposites*.

Table 5.22: Correlations between the individual Twenty Quality Grades and the OCEAN subcomposite scores.

Twenty Quality subscale	*Quality Grades*	*OCEAN subcomposite*	
Tactical	Professional Knowledge	*Neuroticism* (Envious and Jealous)	0.32*
	Tactical Awareness	*Openness* (Creative)	0.30*
		Extraversion (Shy and Bashful)	0.29
		Extraversion (Talkative)	0.30*
		Extraversion (Unsociable) ~	0.27*
		Neuroticism (Envious and Jealous)	0.30*
	Practical Ability	*Agreeableness* (Cold and Insensitive)~	0.29*
	TWS Technical Knowledge	*Extraversion* (Assertive)	0.44**
		Neuroticism (Envious and Jealous)	0.40**
	Navigation and Pilotage	*Openness* (Cultural)	0.29*
	Instinctiveness (Safety)	*Neuroticism* (Envious and Jealous)	0.29*
	Courage	*Openness* (Philosophical)	-0.28*
		Neuroticism (Envious and Jealous)	0.34*
	Caution	*Openness* (Philosophical)	-0.28*
		Openness (Creative)	0.38**
Administration	Management Ability	none	
	Staff Work	*Agreeableness* (Considerate)	0.30*
	Use of English	*Openness* (Cultured)	0.35*
		Agreeableness (Warm and Sympathetic)	0.36**
		Agreeableness (Considerate)	0.36**
		Agreeableness (Cold and Insensitive)~	0.31*
		Neuroticism (Envious and Jealous)	-0.28*
Personality	Leadership	none	
	Stamina	none	
	Command Presence	*Extraversion* (Shy and Bashful) ~	0.40**
		Extraversion (Talkative)	0.33*
		Extraversion (Assertive)	0.30*
		Extraversion (Unsociable) ~	0.34*
	Bearing and Dress	*Agreeableness* (Helpful)	-0.29*
	Intelligence	*Agreeableness* (Considerate)	0.28*
	Common Sense	*Openness* (Reflective)	0.32*
	Maturity	*Neuroticism* (Worrying)	0.28*
	Charm/Sense of Humour	*Agreeableness* (Warm and Sympathetic)	0.28*
	Honesty	*Openness* (Philosophical)	-0.35**
		Extraversion (Talkative)	-0.32*

* $p < 0.05$; ** $p < 0.01$ (two-tailed).N = 53. ~ Subcomposite label is negatively phrased.

5.6 Analysis E: The Occupational Stress Indicator, the Pass-Fail Result and the Course Grade

This analysis starts with the intercorrelations between the various *Occupational Stress Indicator (OSI) scales* (see Table 4.10, p. 60). Each table in this series will also include the correlation found between the *OSI scales* and the overall *Course Grade*. These tables are Tables 5.23 to 5.29. Since it was expected that the **OSI measures** would predict the course outcome a one-tailed test of significance was used. The positive and negative signs of the correlations reported have been modified to take into account the way the various **OSI measures** are scored so that the positive correlations reflect the expected scenario, which is that satisfaction with the job, good mental and physical health, being able to influence things, lack of job pressure and good coping strategies will predict a successful outcome on the course.

Table 5.23: Correlations between the 'How you feel about your job' questions and the Course Grade.

How you feel about your job	J1	J2	J3	J4	J5	J6
J1 Satisfaction with achievement, value and growth						
J2 Satisfaction with the job itself	0.69**					
J3 Satisfaction with organisational design and structure	0.59**	0.58**				
J4 Satisfaction with organizational processes	0.71**	0.57**	0.53**			
J5 Satisfaction with personal relationships	0.66**	0.61**	0.63**	0.55**		
J6 Broad view of job satisfaction	0.81**	0.76**	0.73**	0.78**	0.75**	
Course Grade (N = 53)	0.17	-0.02	0.05	0.15	0.07	0.13
* $p < 0.05$; ** $p < 0.01$ (one-tailed test). N = 57 except for the Course Grade.						

Table 5.24: Correlations between the 'How you assess your current state of health' questions and the Course Grade.

How you assess your current state of health	H1	H2
H1 Mental Ill Health		
H2 Physical Ill Health	0.55**	
Course Grade (N = 52)	-0.14	-0.19
* $p < 0.05$; ** $p < 0.01$ (one-tailed). N = 56 except for the Course Grade.		

As might be expected there are some high correlations between the various scales of the questionnaires. However, the only statistically significant correlations with the *Course Grade* are with two of the '*The way you behave generally'* scales of the OSI: B3 'Ambition' (0.43) and B4 'Broad view of Type A' (0.23) which are reported in Table 5.25.

Table 5.25: Correlations between the 'The way you behave generally' questions and the Course Grade.

The way you behave generally	B1	B2	B3	B4
B1 Attitude to living				
B2 Style of behaviour	0.29*			
B3 Ambition	0.50**	0.29*		
B4 Broad view of Type A	0.69**	0.49**	0.63**	
Course Grade (N = 52)	0.16	0.04	0.43**	0.23*
* p < 0.05; ** p < 0.01 (one-tailed). N = 56 except for the Course Grade.				

Table 5.26: Correlations between the 'How you interpret events around you' questions and the Course Grade.

How you interpret events around you	I1	I2	I3	I4
I1 Organisational forces				
I2 Management processes	0.11			
I3 Individual influence	0.32**	0.27*		
I4 Broad view of control	0.67**	0.44**	0.27*	
Course Grade (N = 52)	0.15	0.00	0.11	0.01
* p < 0.05; ** p < 0.01 (one-tailed test). N = 56 except for the Course Grade.				

Table 5.27: Correlations between the 'Sources of pressure in your job' questions and the Course Grade

Sources of pressure in your job	S1	S2	S3	S4	S5	S6
S1 Factors intrinsic to the job						
S2 The managerial role	0.72**					
S3 Relationships with other people	0.46**	0.74**				
S4 Career and achievement	0.62**	0.80**	0.74**			
S5 Organisational structure and climate	0.57**	0.68**	0.80**	0.75**		
S6 Home/work interface	0.50**	0.43**	0.43**	0.52**	0.53**	
Course Grade (N =52)	0.21	0.13	-0.06	0.18	0.06	-0.14
* p < 0.05; ** p < 0.01 (one-tailed test). N = 56 except for the Course Grade.						

Table 5.28: Correlations between the 'How you cope with stress you experience' questions and the Course Grade.

How you cope with stress you experience	C1	C2	C3	C4	C5	C6
C1 Social support						
C2 Task strategies	0.31*					
C3 Logic	0.03	0.25*				
C4 Home and work relationship	0.16	-0.03	0.14			
C5 Time	0.27*	0.54**	0.42**	0.15		
C6 Involvement	0.30*	0.34**	0.35**	0.16	0.35**	
Course Grade (N = 52)	0.02	0.16	0.10	-0.21	-0.04	0.18
* p < 0.05; ** p < 0.01 (one-tailed test). N = 56 except for the Course Grade.						

The results of an analysis of the difference of means of those who passed or failed the course are presented in Table 5.29. The means of four questionnaire scales of the **OSI**

showed a statistically statistical difference. Two of these are the same as those in the correlation study presented above: B3 'Ambition' and B4 'Broad view of Type A', together with two from the *'Sources of pressure in your job'* questionnaires: S3 'Relationships with other people' and S6 *'*Home/work interface'. According to Cohen (1992) an effect size of 0.5 is a medium effect and a large effect size is 0.8. It can be seen from the values of *d* given in Table 5.29 that these effects are medium to large[31]. The first two results are in the expected direction, that is, students with higher 'Ambition' and higher 'Broad view of Type A' scores are (on average) more successful on the course. Students who passed the course completed two of the '*Sources of pressure in your job'* scales not in the way expected. They reported poorer 'Relations with other people' and reported a worse 'Home/work interface'.

Table 5.29: Statistically significant difference between means of the OSI scales for the Pass-Fail Result.

Significant difference in means	Means (SDs)		Statistics	Cohen's *d*
	Fail	Pass		
B3 Ambition	10.2 (1.6)	11.6 (2.5)	U = 160; p = 0.03	0.6
B4 Broad view of Type A	22.8 (3.0)	24.4 (3.2)	U = 170.5; p < 0.05	0.5
S3 Relationships with other people	27.8 (5.9)	31.1 (4.6)	U = 167.5; p = 0.04	0.7
S6 Home/work interface	31.5 (6.8)	36.1 (6.8)	U = 149.5; p = 0.02	0.7
Mann-Witney U test (one-tailed). N = 51 (13 Fail and 38 Pass). Figures in brackets following the means are standard deviations (SDs).				

5.7 Analysis F: The Occupational Stress Indicator and the Twenty Quality Grades

The results of the analysis of the *OSI scales* and the **Twenty Quality Grades** are given in this section. Table 5.30 shows the correlations between the **OSI scales** and the **Twenty Quality Grades** *Total* and *subtotals*. B3 'Ambition' again features showing high correlations with the *Total Score* and the *subtotals* for both the *Tactical* and *Personality Grades*. The I3 'Individual influence scale', which is a locus of control measure, is correlated with the *Total Tactical* scores. However, H2 'Physical Ill Health' correlates negatively with the *subtotals* for both the *Admin* and *Personality Grades*. Students who reported themselves as having poorer 'Physical Ill Health' received better ratings on the *Total Admin* and *Personality Grades*.   Likewise, those students who reported having poorer 'Home and work relationships' received higher *Grades* on the *Total Score* and *Total Personality subtotal*.

---

[31] Cohen's *d* was calculated using a pooled variance method (except for S6 where the standard deviations for both groups are very similar).

Table 5.30: Correlations between the OSI and the Total Quality Grades subtotals.

H2 Physical Ill Health	
Total Admin	-0.23* [0.05]
Total Personality	-0.24* [0.05]
B3 Ambition	
Total Score	0.37**
Total Tactical	0.31*
Total Personality	0.33**
I3 Individual influence	
Total Tactical	0.28*
C4 Home and work relationship	
Total core	-0.28* [0.02]
Total Personality	-0.30* [0.01]
* p < 0.05; ** P < 0.01 (one-tailed). N = 52.	
Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.	

Table 5.31 reports the significant correlations found between the **OSI** and the *individual Twenty Quality Grades*. Altogether forty-nine significant correlations were found with 'Ambition' having the greatest number of links with eight significant correlations. Three of these correlations are approaching 0.4 and link to Practical Ability, Leadership and Command Presence. The J4 'Job satisfaction with organizational processes' is linked to three of the *Quality Grades* including Tactical Awareness and Command Presence and another in the same set of questionnaires, J5 'Satisfaction with personal relationships' is also linked positively to three of the *Quality Grades*.

The Locus of Control measure, 'Individual influences', is linked to three *Quality Grades*, particularly Courage, but also Instinctiveness (Safety). Another **OSI scale**, S1 'Factors intrinsic to the job', links to three *Quality Grades*: Management Ability, Common Sense and Maturity. However, there are also twelve significant negative correlations in these results. The one which most stands out is the apparent link between poorer 'Mental Ill Health' and higher *Grades* for Leadership. This is not the relationship which might have been expected. Similarly, those students who reported poorer 'Home and work relationships' (C4) were assessed as more Mature and having more Courage. One of the highest correlations found in this Results section is the negative connection of -0.43 between C5 'Time' and Honesty. This suggests that those who have strategies to help time management are viewed as less honest.

There were no significant links found between any of the **Twenty Quality Grades** and the following **OSI scales**: J1 'Satisfaction with achievement, value and growth'; I4 'Broad view of control'; S3 'Relationships with other people'; S4 'Career and achievement'. Table 5.32 shows the results of this analysis the opposite way round

linking the *Twenty Quality Grades* to the *OSI scales*. There are five positive links to TWS Technical Knowledge and four positive links to Command Presence.

Table 5.31: Correlations between the OSI scales and the Twenty Quality Grades.

J2 Satisfaction with the job itself	J3 Satisfaction with organisational design and structure
Navigation and Pilotage -0.31* [0.01]	Navigation and Pilotage -0.35**

J4 Satisfaction with organizational processes	J5 Satisfaction with personal relationships
Tactical Awareness 0.29* Command Presence 0.29* Charm/Sense of Humour 0.27*	TWS Technical Knowledge 0.24* Navigation and Pilotage -0.27* [0.03] Staff Work 0.27* Stamina 0.31*

J6 Broad view of job satisfaction	H1 Mental Ill Health
Navigation and Pilotage -0.25* [0.04]	Leadership -0.28* [0.02]

B1 Attitude to living	B2 Style of behaviour
TWS Technical Knowledge 0.23*	TWS Technical Knowledge 0.24* Navigation and Pilotage 0 .24*

B3 Ambition	B4 Broad view of Type A
Professional Knoweldge 0.26* Practical Abitity 0.36** TWS Technical Knowledge 0.29* Instinctiveness (Safety) 0.27* Management Ability 0.29* Leadership 0.36** Command Presence 0.37** Bearing and Dress 0 .23*	TWS Technical Knowledge 0.25*

I1 Organisational forces	I2 Management processes
Tactical Awareness 0.29 Command Presence 0.30	Bearing and Dress -0.23* [0.05]

I3 Individual influence	S1 Factors intrinsic to the job
Instinctiveness (Safety) 0.28* Courage 0.37** Honesty 0.27*	Management Ability 0.33** Common Sense 0.25* Maturity 0.39**

S2 The managerial role	S5 Organisational structure and climate
Maturity 0.26*	Maturity 0.23*

S6 Home/work interface	C1 Social support
TWS Technical Knowledge -0.24* [0.04] Command Presence -0.36**	Charm/Sense of Humour 0 .28*

C2 Task strategies	C3 Logic
Bearing and Dress 0.25* Charm/Sense of Humour 0.27*	Use of English 0 .34** Command Presence 0 .28*

C4 Home and work relationship	C5 Time
Maturity -0.36** Courage -0.28* [0.02]	Professional Knowledge -0.25* [0.04] Use of English 0.24* Honesty -0.43**

C6 Involvement	
Tactical Awareness 0.23* Practical Ability 0.25* Maturity -0.24* [0.04]	

* p < 0.05; ** p < 0.01 (one-tailed) N = 52 except for J items where N = 53. Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.

In addition there are three positive links to Tactical Awareness, Maturity and Charm/Sense of Humour. The negative correlations include four for Navigation and

Pilotage and two to Maturity. There were no links for the Caution or Intelligence *Quality Grades* to any of the **OSI questionnaires scales**.

Table 5.32: Correlations between the individual Twenty Quality Grades and the OSI scales.

	Quality Grade	OSI Scale
Tactical	Professional Knowledge	B3 Ambition 0 .26* C5 Time -0.25* [0.04]
	Tactical Awareness	J4 Satisfaction with organisational processes 0.29* I1 Organisational forces 0.29* C6 Involvement 0.23*
	Practical Ability	B3 Ambition 0.36** C6 Involvement 0.25*
	TWS Technical Knowledge	J5 Satisfaction with personal relationships 0.24* B1 Attitude to living 0.23* B2 Style of behaviour 0.24* B3 Ambition 0.29* B4 Broad view of Type A 0.25* S6 Home/work interface -0.24* [0.04]
	Navigation and Pilotage	J2 Satisfaction with the job -0.31* [0.01] J3 Satisfaction with organisational design and structure -0.35** J5 Satisfaction with personal relationships -0.27* [0.04] J6 Broad view of job satisfaction -0.25* [0.04] B2 Style of behaviour 0.24*
	Instinctiveness (Safety)	B3 Ambition 0.27* I3 Individual differences 0.28*
	Courage	I3 Individual differences 0.37** C4 Home/work relationships -0.28* [0.02]
	Caution	none
Admin.	Management Ability	B3 Ambition 0.29* S1 Factors intrinsic to the job 0.33**
	Staff Work	J5 Satisfaction with personal relationships 0.27*
	Use of English	C3 Logic 0.34** C5 Time 0.24*
Personality	Leadership	H1 Mental Ill Heath -0.28* [0.02] B3 Ambition 0.36**
	Stamina	J5 Satisfaction with personal relationships 0.31*
	Command Presence	J4 Satisfaction with organisational processes 0.29* B3 Ambition 0.37** I1 Organisational forces 0.30* S6 Home/work interface -0.36** C3 Logic 0.28*
	Bearing and Dress	B3 Ambition 0.23* I2 Management processes -0.23* [0.05] C2 Task strategies 0.25*
	Intelligence	none
	Common Sense	S1 Factors intrinsic to the job 0 .25*
	Maturity	S1 Factors intrinsic to the job 0.39** S2 The managerial role 0.26* S5 Organisational structure and climate 0.23* C4 Home and work relationships -0.36** C6 Involvement -.24* [0.04]
	Charm/Sense of Humour	J4 Satisfaction with organisational processes 0.27* C1 Social support 0.28* C2 Task strategies 0.27*
	Honesty	I3 Individual differences 0.27* C5 Time -0.43**

* p < 0.05; ** p < 0.01 (one-tailed). N = 52 except for J items where N = 53.
Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.

5.8 Analysis G: Miscellaneous AIB items

In this section the results of an analysis of some other measures collected or generated at the time of selection at the Admiralty Interview Board will be examined. These measures are:

- The *Professional Aptitude Predictor (PAP)* which is a scored biodata measures using mainly exam results, the number and types of games, pastimes and part-time jobs.
- The *four assessment centre dimensions*: Effective Intellect, Leadership Potential Character and Personality and Service Motivation.
- The *Headteacher's Reference* score which has a total score made up from seven individual rating scores.

These were all tested at the one-tailed level of significance since it was predicted that these measures would all predict the students' performance on the course.

The *Professional Aptitude Predictor (PAP)* produced no significant correlations for the higher level course information such as such as the **Pass-Fail Result,** the **Course Grade** or the *Total and subtotals* of the **Twenty Quality Grades**. (These results are not shown but can be found in Appendix J, Section 4, p. 144.) However, five of the *individual Twenty Quality Grades* have significant correlations with the *PAP scores*. Table 5.33 shows these.

Table 5.33: Correlations between the Professional Aptitude Predictor and the Twenty Quality Grades .

Professional Aptitude Predictor (N = 28)	Tactical Awareness	-0.34* [0.04]
	TWS Technical Knowledge	-0.37* [0.03]
	Use of English	0.44*
	Bearing and Dress	-0.33* [0.04]
	Charm/Sense of Humour	0.38*
* p < 0.05 (one-tailed). Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.		

The relationship with Use of English appears to be a sensible link and has a high correlation coefficient of 0.44. However, the link to Charm/Sense of Humour isn't apparent. Similarly, the three significant negative correlations are puzzling.

Table 5.34 presents the results of an analysis of the four overall assessment centre dimensions: Effective Intellect, Leadership Potential, Character and Personality and Service Motivation. Significant correlations were found only at the *individual Twenty Quality Grades* level (See Appendix J, Section 5, p. 145, for the other results). The high correlation between AIB assessment dimension of Leadership Potential and Practical Ability is a reasonable link as is the link between Service Motivation and Bearing and Dress, but the even higher correlation found between the board assessment dimension of Character and Personality and TWS Technical Knowledge is not easy to fathom.

Table 5.34: Correlations between the four assessment centre dimensions and the Twenty Quality Grades.

Assessment Dimension	Twenty Quality Grades
Effective Intellect (N = 30)	none
Leadership Potential (N = 30)	Practical Ability　　　　　0.36*
Character and Personality (N = 30)	TWS Technical Knowledge 0.45**
Service Motivation (N = 30)	Practical Ability　　　-0.41* [0.01] Bearing and Dress　　　　0.34*
* p < 0.05; ** p < 0.01 (one-tailed). Figures in square brackets after the significant negative correlation coefficients are one-tailed significance levels.	

The last item to be examined in this analysis is the *Headteacher's Reference* score. The difference in means of the *Headteacher's Reference* score of the **Pass-Fail Result** groups was not statistically significant [U = 256.5, N = 56, p = 0.09]. Table 5.35 shows that the total score for the *Headteacher's Reference* does not correlate with the overall **Course Grade** nor with the *Total or subtotals* of the **Twenty Quality Grades.** However, three of the correlations, though not significant, are around the 0.2 level (one-tailed test).

Table 5.35: Correlations between Headteacher's Reference total score the Course Grade and the Twenty Qualities total score and subtotals.

	Headteacher's Reference Total Score
Course Grade	0.22
Total	0.23
Total Tactical	0.13
Total Admin	0.12
Total Personality	0.19
None of the correlations were significant. N = 56 for Course Grade and N = 46 for the others.	

The *Headteacher's Reference* score is calculated by adding seven individual scores which aim to quantify the candidate's Application, Involvement, Responsibility, Character, Relationships, Influence and Contribution. Table 5.36 shows the results of the analysis of the *Headteacher's Reference subscores* with the **Course Grade** and the **Twenty Quality Grades** *Total Score* and *subtotals*. Whilst only one of the seven *Headteacher's Reference* score for Contribution, is significantly related to the **Course Grade,** four of the seven *subscores* correlate with the *Total Score* of the **Twenty Quality Grades**; these are Involvement, Relationships, Influence and Contribution. An additional four significant correlations were found with the **Quality Grades** *subtotals*. The link between the *Headteacher's Influence* score and *Total Personality* is a likely link. The most obvious omission, though, is any link between the *Headteacher's* Character score and the *Total Personality* score. [Please note that in quite a few cases the headteacher did not complete a full set of ratings on a candidate and this accounts for the different number of cases shown in the analysis.]

Table 5.36: Correlations between Headteacher's Reference subscores and the Course Grade and Twenty Qualities Total score and subtotals.

Headteacher's subscores	Course Grade	Total	Total Tactical	Total Admin	Total Personality
Application					
Involvement		0.24* (55)	0.24* (55)		
Responsibility					
Character					
Relationships		0.25* (48)		0.31* (48)	
Influence		0.30* (48)			0.31* (48)
Contribution	0.28* (61)	0.27* (50)	0.31* (50)		
* p < 0.05 (one-tailed) Non-significant results are omitted. The figures in brackets indicate the number of candidates.					

The final analysis looks at the relationship between the subscores of the *Headteacher's Reference* and the *individual Quality Grades*. The results are shown in Table 5.37 This shows that at this level there are some significant correlations. The relationship found between *Headteacher's* Responsibility score and the Leadership *Quality Grade* is interesting and the links between *Headteacher's* Influence score and the four *Quality Grades* of Caution, Leadership, Command Presence and Maturity make sense, but some of the others do not; for example, *Headteacher's* Relationships score and Service Writing.

Table 5.37: Correlations between Headteacher's Reference subscores and the individual Twenty Quality Grades.

Head Application (N = 55)		Head Involvement (N = 55)	
Courage	-0.34**	Tactical Awareness	0.28*
Head Responsibility (N = 46)		Head Character (N = 48)	
Leadership	0.29*	none	
Head Relationships (N = 46)		Head Influence (N = 48)	
Service Writing	0.26*	Caution	0.25*
		Leadership	0.32*
		Command Presence	0.33*
		Maturity	0.27*
Head Contribution (N = 50)			
Professional Knowledge	0.26*		
Tactical Awareness	0.29*		
Caution	0.28*		
Leadership	0.29*		
* p < 0.05; ** p < 0.01 (one-tailed)			

5.9 Analysis H: Links between the OCEAN Personality Inventory and the AIB Measures

Table 5.38 reports that *Openness* correlates with just two of the ***AIB measures***; the *Numerical test* score and the *Physics grade.* The latter obtained an average of 16 years before the OCEAN inventory was completed.

Table 5.38: Correlations between the various Openness scores and the Numerical test and Physics grade.

	Numerical Test	Physics Grade
Openness	0.30*	0.29*
OpennessT	0.27*	0.16
OpennessSD	0.28*	0.34**
Openness (Philosophical)	0.36**	0.22
Openness (Scientific Interest)	0.11	0.33**
Openness (Creative)	0.24*	0.24*
Openness (Reflective)	0.23*	0.29*
Openness (Cultured)	-0.02	0.26*
Number of cases	57	54
* p < 0.05; ** p < 0.01 (one-tailed).		

One of the highest correlations found was the link between *Openness* (Scientific Interest) and *Physics* at 0.33 which appears logical. However, looking back at the Table 5.17 *OpennessT* correlated -0.34 with *Total Tactical,* and many of the *Openness subcomposites* had negative links to the *individual Twenty Quality Grades.* For example, *Openness* (Philosophical) correlated negatively with *Total Tactical* at -0.28 and with three individual *Grades* at around -0.3 each.

Having found this result, the other four personality dimensions were explored. Only *Extraversion* produced a number of correlations with ***AIB measures*** and these are shown in Table 5.39. *Extraversion* is linked negatively to the *Verbal test* and the *Maths Grade,* and positively to the *Gym* and the *Discussion* marks.

Table 5.39: Correlations between the various Extraversion scores and some AIB measures.

	Verbal Test	Maths Grade	Gym Mark	Discussion Mark
Extraversion	-0.26*	-0.26*	0.30*	0.23*
ExtraversionT	-0.24*	-0.25*	0.28*	0.25*
ExtraversionSD	-0.27*	-0.25*	0.27*	0.18
Extraversion (Shy and Bashful)~	-0.25*	-0.17	0.29*	0.28*
Extraversion (Talkative)	-0.24*	-0.39**	0.23*	0.25*
Extraversion (Socially Active)	-0.04	-0.17	0.32*	0.18*
Extraversion (Assertive)	-0.18	-0.17	0.03	-0.05
Extraversion (Unsociable)~	-0.27*	-0.23*	0.16	0.07
Number of cases	56	54	57	57
* p < 0.05; ** p < 0.01 (one-tailed). ~ *Subcomposite* label is negatively phrased.				

These results are discussed in the next chapter.

6 Discussion

6.1 Introduction

The results will be discussed to take account of the two aims of the study: (1) to investigate the relative value of psychometric versus other assessment centre selection measures and (2) the value of these and other approaches for selecting naval commanding officers for practitioners. This discussion will also link to some relevant research published after the data gathering for this thesis began. Some general comments on the results of the research will be made, as well as remarks about how the literature review, this research and the findings contribute to the wider issues of predicting naval command performance and also practitioner issues. A summary conclusion will complete the thesis.

6.2 The Psychometric Tests and other AIB Measures

The only statistically significant results in the analysis of the ten **AIB measures** were the links between the *Non-verbal Test* and the ***Pass-Fail Result*** and ***Course Grade***. The difference in means for the ***Pass-Fail Results*** had a small to moderate effect size and a corresponding correlation of 0.20 for the ***Course Grade***. Although the results are statistically significant only 4% of the variance is explained leaving 96% of the students' performance is not explained by this measure. However, the *Non-verbal test* was the most likely measure to be a good predictor. The AIB *Non-verbal test* is a matrices test and Carroll (1993) notes that matrices tests can be a good test of general intelligence as well as specific abilities. He notes that they are:

> … readily interpreted as measuring a general ability to deal with visual forms, particularly those that would be generally characterized as figural or geometric and particularly those whose perception or mental manipulation is complex and difficult (p. 609).

Carroll (1993) also notes that there may be links to perceptual speed, visual perception and reasoning ability.

The commander of a submarine will use the periscope to take as short a view as possible of the disposition, speeds and movements of ships on the surface and then lower the periscope. He may then move the submarine to another position underwater and take another look through the periscope. The commander with good spatial perception and reasoning ability will be able to estimate accurately where the ships will be. The time the periscope spends above water has to be minimised so that there is less chance of the submarine being spotted. In addition, the distance of a ship seen in the

periscope is judged by measuring the angle of the top of the mast above the sea, using a grating in the periscope eyepiece. The commander knows the mast heights of various types of ship and so can calculate the distance of the ship in his head. Both tasks are complex 'figural or geometric', so the *Non-verbal test* may indicate the SMCC students' ability in this area.

In the analysis of the AIB predictors with the *Total Score* and the *Quality Grades,* the *Non-verbal test* showed a significant correlation of 0.30 with the *Total Tactical* score. This is also in line with the expected prediction of non-verbal tests referred to by Carroll (1993) above. In addition, the *Verbal test* correlates 0.23 with the *Total Admin* score. However, since the *Total Admin* score did not correlate with the *Total Score* it would appear that the *Total Admin* score is measuring another aspect of performance linked to verbal ability rather than non-verbal ability. This distinction is similar to Vernon's (1961) two types of intelligence; mechanical/spatial (*Non-verbal test)* verbal/educational *(Verbal tests)* or Sternberg's (1985) practical and analytical intelligence.

The relationship of the **AIB measures** with the *individual Twenty Quality Grades* showed that the *Non-verbal test* was significantly correlated with four important aspects of tactical performance: Professional Knowledge (0.35), Tactical Awareness (0.29), Practical Ability (0.31) and Instinctiveness (Safety) (0.23). This reinforces Campbell *et al.'s* (1993) ideas about the importance of job-specific task proficiency. The other positive relationships found in this analysis were that:

- The *Verbal test* correlated with Staff Work (non-job-specific task proficiency),
- The *Instructions test* was linked to Tactical Awareness,
- The *Discussion mark* with Caution and
- The *Final Board Mark* with Staff Work.

However, several negative correlations found in this analysis were linked to the *Verbal test,* the *Numerical tests* and the *Final Board Mark*. This may suggest that the type of person who is successful at initial selection may not be so successful at this more senior stage of assessment, reflecting Drucker's (1955) comments on the important change from junior to senior management. At the initial selection stage the Board Members are looking for someone who is a team player who will cooperate with other people and blend in with the group of candidates. Leadership potential is high on the agenda of attributes which the Board is looking for, but this is very much in conjunction with listening to others, making suggestions and not interfering too much when the other

candidates lead their own *Gym task*. In contrast, on the Submarine Command Course, the students are assessed when they are acting alone as the submarine commander. One student is in charge while the other students help out with navigation and warfare tasks and carry out the roles they would normally perform as junior officers. The course Commanding Officer does not interfere either These sorts of real-world examples are not discussed in These sorts of real-world examples are not discussed in unless there are pressing safety issues. This may be one reason for the different style of leadership required in the two different settings.

Another major reason why more significant links between the AIB predictors and the overall course outcomes were not found is that, after an average of thirteen years in the Royal Navy, the individuals on the course have a good deal of common training and experience in the submarine service and, as mentioned in the limitations of the research, will probably be a very homogeneous group.

6.3 Biodata, Board Dimensions and Headteacher's Reference Report

The analysis of the various miscellaneous items showed some significant correlations, but only at the level of the *individual Twenty Quality Grades*. The *Professional Aptitude Predictor (PAP),* which is a biodata measure, correlated with Use of English (0.44), which might be expected since the PAP mainly consists of academic results. However, it is difficult to see how this correlates with Charm/Sense of Humour, unless those candidates who stayed on longer at school came from more affluent backgrounds!

The analysis of the four *assessment centre dimensions* suggests that there is a link between Leadership Potential, as assessed by the AIB, and the Practical Ability *Grade* on the SMCC course (0.36). The assessment of leadership qualities at AIB would come from the assessor's observation of candidates as they took part in the practical *Gym Task* and the leaderless *Discussion Exercise*. It is likely that the assessors were aware of, and had assessed, the practical ability of the candidates on these two tasks, and that this influenced the AIB Leadership Potential mark. The Character and Personality mark awarded to candidates at AIB correlates at 0.45 with the TWS Technical Knowledge assessment. There may also be a link here between the AIB assessors' observation of candidates using the *Gym task* equipment, e.g., tying knots and using cantilevers, and their speed, time and distance calculations in the *Discussion Exercise*. The link between the AIB Service Motivation score and Bearing and Dress *Grade* is more obvious, but the high negative correlation of *Service Motivation* score with Practical Ability is more puzzling.

The final analysis of the **AIB measures** looked at the prediction of a scored reference report given about the AIB candidate by their Headteacher. Neither the *Headteacher's Reference* score nor the *subscores* predicted the **Pass-Fail Result** or the **Course Grade**. However, four of the *Headteacher's Reference subscores* were correlated with the *Total Score* of the **Quality Grades**. These were the Headteacher's assessment of Involvement, Relationships, Influence and Contribution. One of the most likely links which was not seen was that between the *Headteacher's* Character rating and the *Total Personality* score. However, on reflection, these might be very different aspects of performance in widely different settings. Many of the *individual Twenty* **Quality Grades** such as Leadership, Command Presence, Common Sense, etc., are related to influencing people; and a statistically significant relationship with the *Headteacher's* Influence rating was found. Nevertheless, some of the other links, for example, the rating of *Headteacher's* Involvement score with the *Total Tactical* score are not so easy to explain. Perhaps this is a measure of conscientiousness?

Several links were found between the *Headteacher's Reference subscores* and the *individual Twenty Quality Grades*. Two of the *Headteacher subscores* had four links each. Those for *Headteacher's* Influence score which link to Leadership, Command Presence and Maturity seem appropriate, but three of the four *Headteacher's* Contribution scores linked to Professional Knowledge, Tactical Awareness and Caution, are not obvious, although the link to Leadership may be assessing a similar type of quality.

Although some of these correlations may not be easy to explain, it is interesting that these one-scale ratings by both the Headteacher and the Commanding Officer, on average over thirteen years apart, gave some significant correlations. On the other hand, seven correlations out of a possible 140 could have occurred by chance.

6.4 The OCEAN Personality Inventory

No significant differences in the means were found between the **Pass-Fail Results** and the various versions of the *OCEAN dimensions* and no significant correlations were found between the *OCEAN dimensions* and the **Course Grade**. In addition, no significant differences or correlations were found with the twenty-two **subcomposites** which make up the **OCEAN inventory**. The *Conscientious* dimension, which had been indicated as a likely predictor in the literature review, did not show up here, though it accounts for the highest correlation of 0.18 seen in Table 5.14 (p. 76), but here p = 0.2 so this is not significant.

None of the dimensions in the three versions of OCEAN are correlated with the *Total Score* of the **Twenty Quality Grades,** although *Conscientiousness* has a correlation of 0.17 and 0.18 with the *Total Score* in the various versions, but this had a probability of about 0.2 in each case. Only one significant correlation was found in the analyses with the *Twenty Quality subtotals*, that between the trait version of *Openness* and the *Total Tactical* score with a correlation of -0.34 (p = 0.01, two-tailed).

No significant correlations were found with the *Tactical, Administrative* or *Personality subtotals* which contribute to the *Total Score*. However, the *Conscientiousness* scores in the three different versions were correlated with the *Total Personality* score (between 0.22 and 0.25). It was also pointed out in the Results section that if the parametric correlation coefficient had been used, these two correlations would have been significant in the two-tailed condition. Most of the data in the meta-analytic literature uses both parametric testing and one-tailed tests; so the findings used in these analyses and the conclusions drawn from them may enhance the the results in statistical testing terms, but this does not alter the numbers of students who are allocated to the pass or fail groups in a practical selection situation.

The results of the analysis of the different version of OCEAN and the *individual Twenty Quality Grades* found some links between the OCEAN dimensions and a few of the *individual Grades.* All three versions of *Extraversion* were linked to Command Presence and all three versions of *Agreeableness* were linked to Use of English. *Openness* was again linked to the *Total Tactical* score items, in this case negatively to Tactical Awareness and Caution, as well as negatively to the personality *Quality Grade* of Honesty. Two of the three alternative versions of *Conscientiousness* showed positive links to *Tactical Grades,* suggesting that this aspect of personality may play a part in job performance.

The next analysis with the OCEAN **subcomposites** and the *Total Score* and three *subtotals* showed four statistically significant correlations. *Openness* (Philosophical) was negatively correlated with *Total Tactical* (-0.28). This result is in line with the one discussed in the previous paragraph where a negative correlation of -0.34 between the trait version of *Openness* and *Total Tactical* score was discussed. It looks as though the *Openness* (Philosophical) *subcomposite* might account for a lot of the association found in the earlier result. The *Openness* (Philosophical) scale consists of ratings on five trait items: 'deep', 'philosophical', 'contemplative', 'introspective' and 'complex' and five self-descriptions relating to 'deep thoughts', 'intellectual curiosity', 'intellectual

discussion', 'theoretical scientist' and 'philosophical discussions', etc. In this particular *Openness subcomposite* about half the items are traits in contrast to the other four subcomposites where three of the four subcomposites consist mainly self-description statements. The other *subcomposite* which is mainly formed of trait items is *Openness* (Creative) which will be discussed in more detail later.

This analysis also found that *Neuroticism* (Envious and Jealous) was correlated with both *Total Tactical* (0.37) and *Total Personality* (0.28), the first of these correlations at 0.37 being one of the highest found in the whole of the analyses. This suggests that those students who report themselves as more neurotic do better on the tactical aspects of the course and score higher on the personality *Quality Grades*. This result may indicate that some of the personality traits required by the submarine commander go against conventional expectations. Some of this has already been discussed under the negative correlations found with the *Final Board Mark*; namely that a person with a more selfish and driven attitude may be more successful at higher level of management.

When the results of the analysis with the OCEAN **subcomposites** scores were examined with the individual **Twenty Quality Grades,** there were some statistically significant correlations. However, whilst four links each to *Openness, Extraversion* and *Agreeableness subcomposites* were found and two for the *Neuroticism subcomposite*, not one link to was found to any *Conscientiousness subcomposite*, although this might have been expected from the main papers in the literature review.

Two of the *Openness subcomposites* have negative correlations with the *Quality Grades*. One is *Openness* (Philosophical) which was discussed above; the other is *Openness* (Creative), which has only four items to make up its scale: three traits: 'creative', 'innovative' and 'inventive', and a self-description: 'I love to find innovative solutions to difficult problems'. Another *subcomposite Openness* (Reflective) has self-descriptions about analysis, and reflective thinking links to Common Sense. The last *Openness subcomposite* relating to 'Cultured' has self-descriptions about reading, poetry, music and visiting art galleries which link to the Use of English *Grade*.

Of the four *Extraversion subcomposites* found to have statistically significant associations to some of the **Twenty Quality Grades**, all link to Command Presence with correlations between 0.3 and 0.4 and three link to Tactical Awareness. Three of the *Agreeableness subcomposites* are linked to Use of English and the other links are reasonable, including one of the few links to the Intelligence *Grade*.

The main finding in this section of results is that one *Neuroticism subcomposite* 'Envious and Jealous' is linked to five *Quality Grades*. The *Neuroticism* (Envious and Jealous) scale is made up of just three items, the trait words 'envious (jealous of what others have)' and 'jealous' and the self-description 'I am jealous of others who get what I would like to have'. There are no correlations with any *Neuroticism* dimension and the *Course Results,* nor with the *Total Score* or other subtotals. However, for *Neuroticism* (Envious and Jealous) there are statistically significant correlations with the *Total Tactical* and *Total Personality* scores and with five *individual Quality Grades*.

It does appear that at the *OCEAN subcomposite* level and the *individual Quality Grades* level, there are many more links than at the higher level. This supports Judge *et al.*'s (2002) research which found higher correlations with the lower level dimensions; and other reviewers in Chapter 2 have suggested that more research at this 'facet' level could prove to be useful. Cheryshenko *et al.* (2011) note that emerging research on narrower personality dimensions or the facets which make up these dimensions is a promising area of research. On the other hand, with twenty-two *subcomposites* and *Twenty Quality Grades* the number of possible correlations is 440. At the five percent level of statistical significance, it would be expected that just less than 22 of these correlations would be found by chance; and this analysis reported 32 significant correlations.

6.5 The Occupational Stress Indicator

The Occupational Stress Indicator also included scale measuring Locus of Control and Type A Behaviour. There were only a few statistically significant results found in the analysis of the OSI and the **Pass-Fail Results** and **Course Grade**. Four of the **OSI scales** showed significant differences in means for the pass-fail status of the students. However, two of these significant differences were not in the expected directions. It was expected that those who reported better 'Relationships with other people' and a better 'Home/work interface' would perform better on the course but the results were in the opposite direction. One explanation could be that the respondents who scored higher on these two scales were attuned to their situation and able to distinguish the pressures and problems they faced. Another possibility is that these people felt able to admit to the sources of pressure and the problems they perceived, such as the interface between work and home, and so gave higher ratings on the questionnaire.

Two of the *'How you feel or behave'* questionnaires did show a significant difference in means in the expected direction. These scales were 'Ambition' and 'Broad view of Type

A'. This confirms Cooper *et al.*'s (2001) view that Type A behaviour can have positive outcomes for people. 'Ambition' was also the only **OSI questionnaire scale** which correlated with the overall **Course Grade** at 0.43 (p = 0.001), the second highest correlation found in all the Results. Hogan (1986) developed a six-factor model of personality where Extraversion is split into two separate factors labelled Ambition and Sociability. An analysis of the correlations between the 'B3 Ambition' scale of the OSI and the various *OCEAN Extraversion* measures, including *subcomposites*, is given in Table 6.1. This shows that the *Extraversion* (Assertive) *subcomposite* is correlated at 0.44 with the **OSI scale** 'Ambition'; and the *self-description version* of *Extraversion* correlates at 0.24 with 'Ambition'.

Table 6.1: Correlations between the OSI scale 'Ambition' and the various versions of the OCEAN Extraversion scales.

Extraversion	Correlation with 'Ambition'
Composite version	0.20
Trait version	0.16
Self-description version	0.24*
Subcomposite Shy and Bashful	0.21
Subcomposite Talkative	0.11
Subcomposite Socially Active	0.20
Subcomposite Assertive	0.44**
Subcomposite Unsociable	0.04
* p < 0.05; ** p < 0.01 (one-tailed). N = 52.	

The OSI 'Ambition' scale is made up of only three items: '… achievement-oriented person …', '… action-oriented …' and '… concerned to learn about other people's opinions about me, particularly recognition others give me'. The *Extraversion* (Assertive) *subcomposit*e is made up of four items: the trait 'bold' and three self-descriptions about speaking up, taking charge and doing a lot of talking in meetings. These don't cover the same areas and may display different aspects of this dimension.

At the level of the *Total Score* and the *subtotals* for the **Twenty Quality Grades,** the 'Ambition' questionnaire was again prominent, with three statistically significant correlations linked to the *Total Score,* the *Total Tactical* score and the *Total Personality* score. This time, the Locus of Control scale 'Individual influence' did correlate with the *Total Score*. The other correlations are negative and not expected. Those students who reported that they had poorer 'Physical Ill Health' had better scores on the *Total Score* and the *Total Personality* score. Likewise, those reporting having fewer coping strategies for the 'Home and work relationship' questions were also given better

performance grades for *Total Score* and *Total Personality*. The first of these anomalous results is difficult to explain. The 'Physical Ill Health' questionnaire is a straightforward list of health symptoms which the respondent has to complete. However, the four items for the 'Home and work relationship' scale may help to explain these findings. The items are: 'Resort to hobbies and pastimes', 'Having a home that is a refuge', 'Deliberately separating "home" and "work"' and 'Expand interests and activities outside work'. The theory behind the **OSI** suggested that a high score on this scale shows someone who has coping strategies that will enable them to cope better with any stress they have at work. The person who does not do these things might be seen as someone who is totally committed to work, a company person (even a workaholic) who would probably get higher grades for performance measures at the end of the course. In addition, because of the nature of the submariners' work, it is not easy to see how any of these coping items apply, except when the officer is on leave or has a shore posting (which could be away from his family home).

In the first phase of this study, reported in Section 4.6, the Locus of Control scale was found to be a good predictor of performance. In this second phase the locus of control scale (I3 Individual influence) did not show any significant results for the *Pass-Fail Results* nor for the overall *Course Grade*. The only result was with the *Total Tactical* score. One probable reason for this is that the **OSI questionnaire** consists of only three items, whereas the Rotter (1975) measure, used in the first phase of the study, has twenty-three items plus six filler questions. The other scale of interest, the 'Broad view of Type A' scale correlated 0.23 with the *Course Grade,* but no other important links were found; it did not correlate with the *Total Score* or *subscores*.

The last part of this analysis investigated the correlations between the *OSI scales* and the *individual Twenty Quality Grades*. Once again, 'Ambition' was prominently featured with nine significant correlations. The *Grades* with correlations approaching 0.4 with 'Ambition' are: Practical Ability, Leadership and Command Presence; a good selection of useful *Quality Grades*. There were other links to positive aspects of job satisfaction related to organizational process and personal relationships.

Some negative correlations are worth exploring, the most striking of which is the apparent link between poor mental health and the Leadership grading. The 'Mental Ill Health' scale is very similar to the well-known General Health Questionnaire series, e.g., the GHQ-12 (Goldberg and Williams, 1988) which asks respondents to agree or disagree with various symptoms, for example: 'Would you describe yourself as being a

rather 'moody' sort of person …?'. Many of the questions are also similar to those in the *OCEAN Neuroticism* battery of items. An examination of this link is given in Table 6.2 which shows that there are many correlations between the OSI 'Mental Ill Health' scale and the various *OCEAN Neuroticism dimensions* and *subcomposites*. The one exception is the Envious and Jealous *subcomposite* which was seen earlier to differentiate some aspects of superior performance on the course. It may be that this *subcomposite* is measuring something other than *Neuroticism*.

Table 6.2: Correlations between the OSI scale 'Mental Ill Health' and the various versions of the OCEAN Neuroticism scales.

Neuroticism	Correlation with 'Mental Ill Health'
Composite version	0.66**
Trait version	0.48**
Self-descriptions version	0.67**
Subcomposite (Nervous and Stressed Out)	0.60**
Subcomposite (Irritable)	0.58**
Subcomposite (Worrying)	0.57**
Subcomposite (Envious and Jealous)	0.19
** $p < 0.01$ (one-tailed). N = 52.	

The 'S3 Relationships with other people' questionnaire, which showed a significant difference in means between the pass and fail groups of students, did not link to any of the **Twenty Quality Grades**. However, the sources of pressure questionnaire 'S6 Home/work interface', which also differentiated between the pass and fail groups, was negatively linked to Command Presence. The coping strategies scale 'C4 Home and work relationships' was negatively correlated with Maturity and with Courage. Possibly those students who were graded as more Mature may have had more insight, and might have been able to set themselves apart from home and work conflicts, but still saw this as a source of pressure and reported it on their rating of the *OSI scale*. In this analysis (unlike the one in the previous section) the three-item Locus of Control measure did produce some statistically significant links to Instinctiveness (Safety), Courage and Honesty. It could be argued that these aspects are the ones where students might believe that they can most influence events.

Lyne *et al.* (2000) have reviewed the factor structure of the **OSI**. They reported that the **OSI** had been widely used to conduct research on levels of stress in many occupational groups, but found that the factor structure was not supported by the original scoring keys. They recommended that an alternative scoring system should be substituted for

many of the scales. Whilst the physical and mental health scales are fine, those for the Locus of Control scale (I3 Individual influence) and the Type A Behaviour scale need substantial revision. Lyne *et al.*'s proposed scoring for the Type A scale uses twelve of the fourteen items in the questionnaire rather than just the first six items used in the original scoring. The rescoring for the Locus of Control scale uses eight items from the twelve in the original questionnaire rather than just the three items of the original scale (two of the items in the original scale are not used in the proposed scale!). The revised scoring was not carried out in this research since the OSI scoring method uses seven plastic overlay scoring keys, each with several subscales with reverse scoring on some items; and that is complicated enough. Evers *et al.* (2000) note that in the Dutch version of the OSI they completely changed the Type A and locus of control scales. It might be profitable to reanalyse the data using Lyne *et al.*'s new scoring system for the OSI Locus of Control scale and see if this gives an increases prediction.

6.6 Links between the Personality Dimensions and the AIB Measures

In the literature review on selection in the military in Chapter 3 it was reported that Hogan and Lesser (1996) had investigated performance in hazardous situations like bomb disposal and naval diving. Their findings led them to propose that there may be a link between Openness and general mental ability. It was not the original intention of this research to look at the correlations between the various predictor measures, but some other research has linked personality and intelligence together. Using meta-analysis, Ackerman and Heggestad (1997) found that Openness correlated 0.33 with general intelligence suggesting that personality may be a moderating variable. In a recent review of the topic, Oswald and Hough (2011) also noted that there might be a link between personality and intelligence.

An investigation was carried out with the data collected for this research. One of the highest correlations found was the link between *Openness* (Scientific Interest) and Physics at 0.33 which appears logical. However, looking back at the Table 5.17 (p. 78) *OpennessT* correlated -0.34 with *Total Tactical,* and many of the *Openness subcomposites* had negative links to the *individual Twenty Quality Grades.* For example, *Openness* (Philosophical) correlated negatively with *Total Tactical* at -0.28 and with three individual *Grades* at around -0.3 each. It is difficult to explain what is going on here but this could also be related to the different characteristics needed to become a successful as a senior officer. Oswald and Hough (2011) note that *Openness* is the most controversial of the Big Five Factors and may be better understood at the

facet level which is certainly the case with these results, where the subcomposites may be better predictors.

Having found this result, the other four personality dimensions were explored. Only *Extraversion* produced a number of correlations with **AIB measures**. *Extraversion* is linked negatively to the *Verbal test* and the *Maths Grade,* and positively to the *Gym* and the *Discussion* marks. There could be other explanations, but these linkages suggest that introverts may be better at the academic types of tests, and extraverts are better at the practical aspects of the assessment centre process. The *Non-verbal test* and the *Instructions test* showed no significant correlations with any of the *Extraversion* scales, so perhaps there is a balance with these two tests being located between the practical and the academic. This issue could be an interesting area for future research. Zimprich *et al.* (2009) have recently noted that the Big Five scheme may be too broad for prediction and that some important variables are missing so researchers should look at facets and explore further the link between personality and intelligence.

6.7 Predicting Naval Command Performance

It is, of course, disappointing that few results and correlations were found between the various predictor measures and the course results and the SMCC performance assessments. It is also disappointing that some of the results from the first phase of the study were not replicated in the second phase. The main problem with attempting to use any of the measures to predict success on the course was that of the increased homogeneity over time of the group being studied. The submarine service does not select its members; they are all people who volunteer at the end of initial training. The experience of submariners is very similar even in comparison to their colleagues in the surface fleet who serve in a variety of vessels which have a wide range of roles and tasks. The submarine environment is fairly standard and officers get to know each other well through different postings and training courses where they work alongside each other. On the subject of the prediction of specialist jobs Sternberg (1997) notes that:

> … we would probably find conventional measures of intelligence to be about as good or better than any other single measure. But measures that are highly predictive of success *across* jobs are not necessarily particularly predictive *within* jobs. (p. 201, original emphasis).

Other personality theorists like Gardner (1983) and Goleman (1995), and Sternberg (1985) himself, have proposed that other aspects of intelligence, which are not covered by conventional ability tests, may be more predictive of success in organizations.

Some of the investigations for this study set out in the Context and Methods chapter (Section 4.2 pp. 44-45) were not met because of the lack of results. It was not possible to compare the relative strengths of prediction of the AIB measures because few of them were predictive and the correlations obtained were low. Schmidt and Hunter (1998), for example, found that cognitive ability tests plus measures of work samples had high validity, so the *Non-verbal test* score plus the *Gym task* score might appear to be a likely predictor; but with the low or absent correlations this aspect of the research could not be explored.

The essence of this thesis is trying to predict performance using ability and personality measures. A major problem with trying to do this has been pointed out by Campbell *et al.* (1996). They state that the determinants of performance are either direct or indirect. The direct determinants such as declarative knowledge, professional knowledge and skills and motivation can influence performance directly but the indirect determinants such as ability, personality and interests (such as education, training and experience) can only influence performance by influencing the direct determinants. This means that there is no direct causal path between the indirect determinants and performance. Campbell *et al.* (1996) note that 'If, for example, there is a significant direct causal path between general mental ability and a supervisory rating of performance, it is by definition an artefact' (p. 274). This may mean that another, less direct model of performance is necessary for future research and that a different type of analysis, like structural equation modelling, may be appropriate.

Since the *Non-verbal tests* was the best predictor of success on the course a more dynamic battery of spatial and non-verbal test like those in MICROPAT (Bartram and Dale, 1991) could be trialled. This is a computerised battery of tests to measure psychomotor and information processing. MICROPAT is used to select armed service pilots in the UK and civilian pilots in many countries. This test battery was considered at the design stage of the research and computers with this system installed were readily available. However, this was not used for practical reasons like the length of time for testing, with only one individual on a computer at one time, and it would not have been possible to provide adequate supervision for the testing process.

There may be glimpses of the type of character required to be a successful submarine commander coming from this research, but the five, top-level *OCEAN dimensions* do not tap into this. The results of the first phase of the research found that the more conventional 16PF Factor Q2, linked to being a joiner and a team person, was correlated

with success on the course. However, another factor (Factor L) emerged which was linked to traits such as hard to fool, distrustful, sceptical and people involved with their own egos. This finding resonates with the higher correlations found in phase two of the research, like *Neuroticism* (Envious and Jealous) and 'Ambition', both of which predicted success. Research by Hogan (1986; Hogan and Shelton, 1998) has identified that Ambition appears to be an important aspect of personality and that it could be used to predict outcomes like leadership.

The Results chapter also presented various negative links for the ***Openness subcomposites*** with the performance *Grades* which suggests that people who are not open to experience have a better chance of passing the course. Some of these themes have already been reported by Drucker (1974) who writes (ignoring the gendered language of the time) that:

> The great leader is rarely 'warm', a good many have been icy. He is not often 'outgoing' or 'affable'; he tends to be austere and aloof. He has little 'empathy'; he makes demands. A good many have not had a trace of charisma. But a leader always inspires confidence, always commands respect (p. 303).

It is likely that these traits are not found in typical personality inventories since these are often developed by the factor analysis of scores from junior managers. The process of factor analysis itself could remove these important traits which might be more predictive of success at a senior management levels. Some of the evidence in Chapter 2 suggests that personality dimensions at senior levels may predict in the opposite direction (Robertson *et al.,* 2000).

The use of personality dimensions, particularly at the higher level, may not be the best way to proceed. Recent critics of the Big Five like Hough and Dilchert (2010), Zimprich *et al.* (2009) and Cheryshenko *et al.* (2010) have all suggested using more targeted personality dimensions to get a better prediction of performance. They also note that Locus of Control is a useful predictor of job performance. By focusing the attention of personality research solely on traits to predict performance, Cortina and Ingerick (2005) note that researchers and practitioners are not considering less behavioural aspects of personality such as motivation, needs, goals, beliefs, schemas. In addition, situational factors are ignored as are patterns of behaviour.

Earlier theorists like McClelland (1987) and Murray (1981) have noted that people have needs for achievement, power and affiliation, and needs to avoid failure and to do things better. These factors are not measured in conventional personality tests such as the Big Five, but appear to be important in this research in areas such as the 'Ambition' scale.

Holt (2006) argues that 'The drawback of the Assessment Centre is that this standardised approach means it struggles to assess a candidate's passion, adaptability, courage and collaboration – the very competencies that drive leadership in modern organisations' (p. 17). Young and Dulewicz (2003; 2008) who conducted research with naval officers noted the importance of a scale measuring 'controlling and ambitious' and they also recommended specifically tailored measures to assess these important characteristics which other studies in the literature review have also suggested.

The *Twenty Quality Grades* produced by the commanding officers at the end of the Course is another major problem with the research. The individual *Grades* were given on a 0 (zero) to 3 scale with zero being rarely used. Unlike the AIB *Gym* and *Discussion* marks, where four board members rated the candidates, these performance *Grades* were given by one person. Over the course of the nine years research the *Twenty Quality Grades* were assessed by the four different commanding officers; and this alone would probably make the *individual Twenty Quality Grades* fairly unreliable.

If the pattern of the zero scores for the *individual Grades* is examined, it is noticeable that no successful candidate was given a zero score. Eight of the twenty-three students who failed the course also had no zeros on their *Grade* assessment. Of the remaining fifteen students who failed, six students had zeros for Instinctiveness (Safety) and three of these same students also got a zero for Tactical Awareness. The remaining nine failures were given a zero score for other *individual Grades* covering Professional Knowledge, Navigation and Pilotage, Courage, Caution, Management Ability, Common Sense, Maturity and Honesty. This would suggest that quite a few individuals fail the course for one or two very specific reasons rather than for more generic, complex reasons. It might, therefore, be difficult to design a selection measure to target these very individual failings.

Another issue with this study is the longitudinal nature of the research. The investigation of the long-term prediction of the AIB measures was a primary aim of the study, but the findings may not be applicable to the present time. The majority of young people entering the navy as officers today have degrees when they join, whereas few, if any, of the participants in this study would have had a degree. In addition, the inventories were filled in by course students who had a clear understanding that the scores were only being used for research purposes. It might be that, if this type of measure was to be used to screen out unsuitable students, they might complete the

inventories in a very different way, reflecting the type of person assumed to be suitable for the course, for example: extravert, agreeable, conscientious, stable, open.

In summary, the research only found a few statistically significant correlations which would not help the practitioner select out unsuitable candidates for the course. The literature review and other research presented in the thesis may suggest using more targeted, specific, designed measures to investigate areas like Ambition and Locus of Control.

6.8 Practitioner Issues

The main finding was that the *Non-verbal* psychometric test taken, on average, about thirteen years before the course started predicted the **Course Grade** with a correlation of 0.20. Table 6.3 shows a contingency table of this result with the correct and incorrect decisions based on the mean of the *Non-verbal test* as the cut-off score.

Table 6.3: Non-verbal Test scores and Pass-Fail Result.

		SMCC Pass-Fail Result			
		Fail	Pass	Total	Pass Rate
Non-verbal Test Scores	39 and over	10	40	50	Pass rate 80.0%
	38 and under	14	24	38	Pass rate 63.2%
	Total	24	64	88	Pass rate 72.7%

The pass rate for students above the cut-off is 80.0 percent compared with 63.2 percent below this. However, using this cut-off means that 37.5 percent of successful students on the course (24 out of 64) would be rejected if this method were to be used for screening. Whilst the result is interesting, if this example is typical of the prediction of validity coefficients of 0.20 which are often quoted in the literature, it has little practical use. Flin and Slaven (1996) found similar results when they tried to use personality inventories to predict emergency command performance. The meta-analytic literature only gives correlations; it does not set out real-world examples of selection like this.

Selection and assessment measures have been developed to offer objective ways of matching people to jobs which have benefits for both employer and employee. Occupational psychologists and other professionals have to rely on published research to justify the methods they use themselves and to recommend these methods to other people. The literature review in this thesis outlined various meta-analytical studies which claimed to show that both cognitive tests and personality measures had moderate to high validity coefficients for predicting job performance and training success. The literature review suggests that meta-analysis may provide transferable selection tools for

users. For the assessment of cognitive ability, where it is clear that a particular job requires a particular ability or aptitude, then cognitive tests can be used and recommended.

The studies supporting personality inventories, particularly those related to the Big Five, appear to have over-compensated for the inherent unreliability in the type of measure and the outcome criteria. As Morgesen (2007) notes, 'We must not forget that personality tests have very low validity for predicting overall job performance. Some of the highest reported validities in the literature are potentially inflated due to extensive corrections or methodological weaknesses' (p. 1030). This misleading information has subsequently been published in review chapters and books likely to be read by occupational psychologists and other personnel in the human resource field who are responsible for designing and running assessment and selection procedures. Practitioners need this information to help them make fair decisions about the people they assess. The increased use of personality tests by organizations for assessment and selection purposes is worrying, given that the real, underlying correlations give such a weak prediction.

The practitioner who reads the journals is often faced with whole journal volumes, sometimes double issues, devoted to the topic of personality, eg, *Human Performance* (1998) and *International Journal of Selection and Assessment* (2001). Mitroff (1974) writing about NASA Apollo moon scientists notes that: 'If you want to get anybody to believe your hypothesis you've got to beat them down with numbers; you've got to hit them again and again over the head with hard data until they are stupefied into believing it' (p. 144). There seems to be quite a bit of this going on in the personality theory field from interested parties, e.g. Hogan (2005a, 2005b), who is test developer and described by Furnham (2008) as writing 'with typical gusto, confidence and bravado' (p. 134). Much of the writing is in a proselytizing style, assuming that there really ought to be links between personality measures and performance. Hough and Dilchert (2010) comment 'that focusing exclusively on factor-level personality traits in the prediction of heterogeneous work-related criteria can be counterproductive for a science aiming to explain the relationship between personality constructs and work-related constructs' (p. 300).

Gray (2003, p. 3) notes that articles have appeared 'throughout the years' in the British Psychological Society publication for practitioners *Selection and Development Review* and that: 'All have the same lament' that there seems to be no personality measure

which 'can fully or largely explain or predict management performance'. Gray notes that this is no surprise since management performance is situational and different jobs need different temperaments. Cook and Cripps (2005) give the results from Barrick et al. (2001) in the form of a histogram of the Big Five personality dimensions showing the validity coefficients and note that: 'Even allowing for the limitations of selection research, personality tests cannot offer a better prediction of work performance than a correlation of 0.23. They cannot even reach the '0.3 barrier', let alone break it' (p. 74).

One main point from this research and from the literature review is that just because selection measures have been repeatedly shown to distinguish between those individuals who will perform well in initial training and in their early career and those who do not perform so well, it does not necessarily follow that these same measures will predict success later in a person's career.

6.9 Conclusion

This research was undertaken to investigate if the results from the psychometric tests and other assessment centre measures taken at the Admiralty Interview Board could predict the outcome for students on the Submarine Command Course which they attended, on average, thirteen years after they completed the assessment centre. The results showed that the *Non-verbal* test was correlated with the ***Course Grade*** at 0.20 and that the difference between the means for the pass and fail groups was statistically significant with a small to moderate effect size. However, any attempt to use such a measure to screen out candidates who would be unsuccessful on the course would lead to an unacceptable level of misclassification. The result is of theoretical interest, but this has limited practical application.

In addition, the *Non-verbal* test was correlated with the *Total Tactical* score (0.30) given by the course commanding officer and several important aspects of performance including Professional Knowledge (0.35), Tactical Awareness (0.29), Practical Ability (0.31) and Instinctiveness (Safety) (0.23). The *Verbal test* score predicted the *Total Admin* score (0.23) and Staff Work (0.25) though other negative correlations were found with performance grades. One of the four *Board dimensions,* Leadership Potential predicted the Practical Ability grade (0.36) and the *Professional Aptitude Predictor* (biodata) score correlated with the Use of English grade.

None of the dimensions measured by the **OCEAN** personality inventory were correlated with the ***Course Grade***, nor with the *Total Score* nor with the three *subtotals* of the

Twenty Quality Grades. The *trait version* of *Openness* correlated -0.34 with the *Total Tactical* score. It was only with the *individual Quality Grades* that correlations were found, in particular with *Extraversion* which was linked with Command Presence. At the ***subcomposite*** level of the **OCEAN dimensions**, one ***subcomposite*** *Neuroticism* (Envious and Jealous) correlated with the *Total Score* as well as the *Total Tactical* and *Total Personality* scores.

Only two of the twenty-eight scales offered by the **Occupational Stress Indicator**, 'Ambition' and 'Broad view of Type A', correlated with the ***Course Grade*** with correlations of 0.43 and 0.23, respectively. These two scales also showed significant differences in means between the students who passed and failed the course with a medium effect size. 'Ambition' also correlated with the *Total Score* as well as the *Total Tactical* and *Total Personality* scores. The OSI Locus of Control scale 'Individual influence' showed a correlation with the *Total Tactical* score.

Given that the types of selection measure used assessment centres can predict aspects of job performance and training success at a junior level, it may be important that these measures are not used to try to select homogeneous groups of middle and senior staff both in the armed services and elsewhere when the literature and this research show poorer prediction at this level. The literature review shows that practitioners, who have to devise selection processes and advise organizations about selection methods, need to keep up to date with the literature and scrutinize journal and book articles on the validity of selection measures. Practitioners should be very careful about the claims made about the prediction of these measures the seemingly high correlations reported in the literature are far from perfect in predicting actual passes and failures in real-life selection situations.

The fact that only one out of the four psychometric tests was able to predict the course grade after an average of thirteen years is disappointing, but this was a rare opportunity to look at such data. Only large organizations can conduct this sort of research and very few of these organizations can keep such a large proportion of their staff for such a length of time. The access to results and grades on this prestigious course, which was a real-life test of decision making in a very pressurised and stressful environment, was a privileged opportunity and may not be easy to repeat in the future.

References

Ackerman, P.L. and Heggestad, E.D. (1997). Intelligence, personality and interests: evidence of overlapping traits. *Psychological Bulletin, 121,* 219-225.

Allport, G.W. and Odbert, H.S. (1936). Trait names: a psycho-lexical study. *Psychological Monographs, 47,* 1-171.

Anderson, N. and Cunningham-Snell, N. (2000). Personnel Selection. In N. Chmiel (Ed.). *Work and Organizational Psychology* (pp. 69-99). Oxford: Blackwell.

Anderson, N. and Goltsi, V. (2005). Negative psychological effects of selection methods: construct formulation and an empirical investigation into an assessment center. *International Journal of Selection and Assessment, 14,* 236-255.

Anderson, N., Payne, T., Fergusson, E. and Smith, T. (1994). Assessor decision making, information processing and assessor decision strategies in a British assessment centre. *Personnel Review, 23,* 52-62.

Ansbacher, H.L. (1951). The history of the leaderless group discussion technique. *Psychological Bulletin*, *48,* 383-391.

Arthur, W. Jr. and Day, E.A. (2011). Assessment centers. In S. Zedeck (Ed.). *APA Handbook of Industrial and Organizational Psychology: Volume 2 Selecting and Developing Members for the Organization* (pp. 205-235). Washington, DC: The American Psychological Association.

Arthur, W. Jr., Day, E.A., McNelly, T.L. and Edens, P.S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56,* 125-154.

Barrick, M.R. and Mount, M.K. (1991). The Big Five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44,* 1-26.

Barrick, M.R., Mount, M.K. and Judge, T.A. (2001). Personality performance at the beginning of the new millennium: what do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9-30.

Bartone, P.T., Eid, J., Johnsen, B.H., Laberg, J.C. and Snook, S.A. (2009). Big Five personality factors, hardiness, and social judgment as predictors of leader performance. *Leadership and Organizational Development Journal, 30,* 498-521.

Bartram, D. (1995). The predictive validity of the EPI and 16PF for military flying training. *Journal of Occupational and Organizational Psychology, 68,* 219-236.

Bartram, D. and Dale, H.C.A. (1991). Validation of the MICROPAT battery of pilot aptitude tests. In P.L. Dann, S.H. Irvine and J.M. Collis (Eds.). *Advances in Computer-Based Human Assessment* (pp. 149-170). Dordrecht: Kluwer Academic Publishers.

Bass, B.M. (1960). *Leadership Psychology and Organizational Behavior.* New York: Harper and Row.

Bass, B. M. (1998). *Transformational Leadership: Industrial, Military, and Educational Impact*. Mahwah, NJ: Erlbaum.

Beadle, I.W. (1997). *Psychometric Testing on the Submarine Command Course* (Report CHS/CR97073). Farnborough, Hants: Defence Evaluation and Research Agency.

Bertua, C., Anderson, N. and Salgado, J.F. (2005). The predictive validity of cognitive ability tests: a UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78,* 387-409.

Bono, J.E. and Judge, T.A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology, 89,* 901-910.

Borman, W.C., White, L.A., Pulakos, E.D. and Oppler, S.H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76,* 863-872.

Brinner, R.B. and Reynolds, S. (1999). The costs, benefits, and limitations of organizational level stress interventions. *Journal of Organizational Behaviour, 20,* 647-664.

Burch, G.St.J. and Anderson, N. (2008). Personality as a predictor of work-related behavior and performance: recent advances and directions for future research. In G.P. Hodgkinson and J.K. Ford (Eds.). *International Review of Industrial and Organizational Psychology (Vol. 23)* (pp. 261-305). Chichester: John Wiley.

Campbell, J. P., Gasser, M. B. and Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.). *Individual Differences and Behavior in Organizations* (pp. 258-299). San Francisco: Jossey-Bass.

Campbell, J., McCoy, R., Oppler, S. and Sager, L. (1993). A theory of performance. In N. Schmitt and W. Borman (Eds.). *Personnel Selection in Organizations* (pp.35-70). San Diego, CA: Jossey-Bass.

Campbell, J.P., McHenry, J.J. and Wise, L.L. (1990). Modelling job performance in a population of jobs. *Personnel Psychology, 43,* 313-333.

Carroll, J.B. (1993). *Human Cognitive Abilities: a survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Cattell, R.B. (1989). *The 16PF: personality in depth.* Champaign, IS: IPAT.

Chernysenko, O.S., Stark, S. and Drasgow, F. (2011). Individual differences: their measurement and validity. In S. Zedeck (Ed.). *APA Handbook of Industrial and Organizational Psychology: Volume 2 Selecting and Developing Members for the Organization* (pp. 117-151). Washington, DC: The American Psychological Association.

Cheung, F.M., Leung, K., Zhang, J., Sun, H., Gan, Y., Song, W. and Xie, D. (2001). Indigenous Chinese personality constructs: is the Five-Factor model complete? *Journal of Cross-Cultural Psychology*, *32,* 407-433.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 111,* 155–159.

Cohen, S. and Edwards, J.R. (1998). Personality charateristics as moderators of the relationship between stress and disorder. In W.J. Neufeld (Ed.). *Advances in the Investigation of Psychological Stress.* New York: Wiley.

Collis, J.M. (1997). *The OCEAN Personality Inventory: Administrator's Guide and Test Manual* (Report CHS/CR97072). Farnborough, Hants: Defence Evaluation Research Agency.

Cooper, C.L. (1986). Job distress: recent research and the emerging role of the clinical occupational psychologist. *Bulletin of the British Psychological Society, 39,* 325-331.

Cooper, C.L. (1998). *Theories of Organizational Stress.* Oxford: Oxford University Press.

Cooper, C.L., Dewe, P.J. and O'Driscoll, M.P. (2001). *Organizational Stress. a review and critique of theory, research, and applications.* London: Sage.

Cooper, C.L., Sloan, S.J. and Williams, S. (1988). *Occupational Stress Indicator: Management Guide.* Windsor: NFER-Nelson.

Cortina, J.M. and Ingerick, M.J. (2005). Personality in industrial/organizational psychology: not much more than cheese. In G. P. Hodgkinson and J. K. Ford (Eds.). *International Review of Industrial and Organizational Psychology (Vol. 20)* (pp. 119-148). Chichester: John Wiley.

Cook, M. and Cripps, B. (2005). *Psychological Assessment in the Workplace: a manager's guide*. Chichester: Wiley.

Costa, P.T., Jr. and McCrae, R.R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4,* 5-13.

Cox, T. (1978). *Stress.* London: MacMillan.

Cox, T. and Mackay, C.J. (1981). A transactional approach to organizational stress. In E. N. Corlett and J. Richardson (Eds.). *Stress, Work Design and Productivity* (pp. 75-95). Chichester: Wiley.

Den Hartog, D.N. and Koopman, P.L. (2001) Leadership in organizations. In N. Anderson, D.S. Ones, H.K. Sinangil and C. Viswesvaran (Eds.). *Handbook of Industrial, Work and Organizational Psychology, Volume 2: Organizational Psychology* (pp. 166–187). Thousand Oaks, CA: Sage.

DERA. (1999). *The AIB Board Members' Handbook*. Farnborough, Hants: Defence Evaluation and Research Agency (unpublished).

Dewe, P. and Cooper, C.L. (2007). Coping research and measurement in the context of work related stress. In G. P. Hodgkinson and J. K. Ford (Eds.). *International Review of Industrial and Organizational Psychology (Vol. 22)* (pp. 141-191). Chichester: John Wiley.

Digman, J.M. (1990). Personality structures: emergence of the five-factor model. *Annual review of Psychology, 41,* 417-40.

Dobson, P. and Williams. A. (1989). The validity of the selection of male British Army officers. *Journal of Occupational Psychology, 62,* 313-325.

Downes, C. (1991). *Special Trust and Confidence: the making of an officer*. London: Frank Cass [cited in Flin and Slaven, 1996].

Drakeley, R. (1989). Biographical data. In P. Herriot (Ed.). *Assessment and Selection in Organizations: methods and practice for recruitment and appraisal* (pp. 439-453). Chichester: Wiley.

Drakeley, R., Herriot, P. and Jones, A. (1988). Biographical data, training success, and turnover. *Journal of Occupational Psychology, 61,* 145-152.

Drascow, F. (2003). Intelligence and the workplace. In W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.). *Handbook of Psychology: Volume 12 Industrial and Organizational Psychology* (pp. 107-130). Hoboken, NJ: John Wiley and Sons.

Drucker, P.E. (1955). *The Practice of Management.* London: Heinemann.

Drucker, P.E. (1974). *Management: Tasks, Responsibilities, Practices.* London: Heinemann.

Dulewicz, V. and Higgs, M.J. (2005). Assessing leadership dimensions, styles and organizational context, *Journal of Managerial Psychology, 20*, 105-123.

Dudley, M., Orvis, K.A., Justin E.L. and Cortina, J.M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40-57.

Dyer, F.J. (1984). *Gordon Personal Profile - Inventory: an interpretative guide.* New York: Harcourt, Brace and Jovanovich.

Edwards, J.E. and Morrison, R.F. (1994). Selecting and classifying naval officers: the paradox of greater specialization in broader arenas. In M.G. Rumsey, C.B. Walker and J.H. Harris (Eds.). *Personnel Selection and Classification* (pp. 69-84). Hillside, NJ: Lawrence Erlbaum.

Evers, A., Frese, M. and Cooper, C.L. (2000). Revisions and further developments of the Occupational Stress Indicator: LISREL results from four Dutch studies. *Journal of Occupational Psychology, 73,* 221-240.

Eysenck, H.J. (1992). Four ways five factors are <u>not</u> basic. *Personality and Individual Differences, 13,* 667-673.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

Fletcher, C. (2005). *Final Report on a Research Study Relating to Effective Selection of Staff for Senior Posts in The Civil Service.* London: HMSO.

Flin, R and Slaven, G. (1996). Personality and emergency command. *Disaster Prevention and Management, 5,* 40-46.

French, J., Robert, D. and Van Harrison, R. (1982). The Mechanisms of Job Stress and Strain. New York: Wiley.

Friedman, H.S. and Booth-Kewley, S. (1987a). The 'disease-prone personality': a meta-analytic view of the construct. *American Psychologist, 42,* 539-555.

Friedman, H.S. and Booth-Kewley, S. (1987b). Personality, Type A Behavior, and Coronary Heart Disease: the role of emotional expression. *Journal of Personality and Social Psychology, 53,* 783-792.

Friedman, M. (1996). Type A Behavior: its diagnosis and treatment. New York: Plenum.

Frome. M.R. (2003). Work family balance. In J.C. Quick and L.E. Tetrich (Eds.). *Handbook of Health Psychology* (pp. 143-162). Washington, DC: American Psychological Association.

Furnham, A. (2008). *Personality and Intelligence at Work: exploring and explaining individual differences at work.* Hove: Routledge.

Gaugler, B.B., Rosenthal, D.B., Thornton, C.C. and Benson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 74,* 397-402.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Gardner, K. and Williams, A.P.O. (1973). A twenty-five year follow up of an extended interview selection procedure in the Royal Navy. *Occupational Psychology, 47,* 1-13, 149-161.

Ghiselli, E.E. (1973). The validity of aptitude tests in personnel psychology. *Personnel Psychology, 26,* 461-447.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48,* 26-34.

Goldberg, D. and Williams. P. (1988). *A User's Guide to the General Health Questionnaire.* Windsor: NFER-Nelson.

Goleman, D. (1995). *Emotional Intelligence: Why it can matter more than IQ.* New York: Bantam Books.

Gray. M.J. (2003). Personality and performance. *Selection and Development Review*, *19*, 3-5.

Guion, R.M. and Gottier, R.F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 8,* 135—164.

Hampson, S.E. (1999). State of the Art: Personality. *The Psychologist, 12*, 284-288.

Hannah, S.T., Uhl-Bien, M., Avolio, B.J. and Cavarretta, F.L. (2009). A framework for examining leadership in extreme contexts. *The Leadership Quarterly, 20,* 897-919.

Hardinge, N.M. (1997). Selection of military staff. In N. Anderson and P. Herriot (Eds.). *International Handbook of Selection and Assessment* (pp. 161-81). Chichester: Wiley.

Hart, P.M., & Cooper, C.L. (2001). Occupational stress: toward a more integrated framework. In N. Anderson, D.S. Ones, H.K. Sinangil and C. Viswesvaran (Eds.). *Handbook of Industrial, Work and Organizational Psychology, Volume 2: Organizational Psychology* (pp. 93–114). Thousand Oaks, CA: Sage.

Hermelin, E., Lievens, F. and Robertson, I.T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: a meta-analysis. *International Journal of Assessment and Selection, 15,* 405-411.

Hermelin, E. and Robertson, I.T. (2001). A critique and standardization of meta-analytical validity coefficients in personnel selection. *Journal of Occupational and Organizational Psychology, 74,* 253-277.

Herriot, P. and Anderson, N. (1997). Selecting for change: how will personnel and selection survive. In N. Anderson and P. Herriot (Eds.). *International Handbook of Selection and Assessment* (pp. 1-34). Chichester: Wiley.

Herzberg, F.I. (1987). One more time: how do you motivate employees? *Harvard Business Review*, *65,* 109-120.

Hogan, R. (1986). *Hogan Personality Inventory Manual.* Minneapolis, MN: National Computer Systems.

Hogan, R, (2005a). In defence of personality measures. *Human Performance, 18,* 331-341.

Hogan, R, (2005b). Comments. *Human Performance, 18,* 405-407.

Hogan, R. and Hogan, J. (2004). Assessing leadership: a vision from the dark side. *Selection and Development Review, 20,* 3-15.

Hogan, J. and Lessser. M. (1996). Selection of personnel for hazardous performance. In J.E. Driskell and E. Slater (Eds.). *Stress and Human Performance* (pp. 195-256). Mahwah, NJ: Erlbaum.

Hogan, R. and Shelton, D. (1998). A socioanalytical perspective on job performance. *Human Performance, 11,* 129-144.

Holt, J. (2006). The Apprentice – the next big thing in assessment. *People and Orgainisations@Work, Winter, 17* [The British Psychological Society].

Hough, L.M. (1992). The 'Big Five' personality variables – construct confusion: description versus prediction. *Human Performance, 5,* 139-155.

Hough, L.M and Dilchert, S. (2010). Personality: its measurement and validity for employee selection. In J.L. Farr and N.T. Tippins (Eds.). *Handbook of Employee Selection* (pp. 299-319). New York: Routledge.

Hough, L.M. and Furnham, A. (2003). Use of personality variables in work settings. In W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.). *Handbook of Psychology: Volume 12 Industrial and Organizational Psychology* (pp. 131-169). Hoboken, NJ: John Wiley and Sons.

Hunter, J.E. and Hirsh, H.R. (1987). Applications of meta-analysis. In C.L. Cooper and I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology (Vol. 2)* (pp. 321-357). Chichester: Wiley.

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

Iles, P. (1999). *Managing Staff Selection and Assessment.* Buckingham: Open University Press.

Ilies, R., Gerhardt, M.W. and Le, H. (2004). Individual differences in leadership emergence: integrating meta-analytic findings and behavioral genetics estimates. *International Journal of Selection and Assessment, 12,* 207-219.

Jansen, P.G.W. (1997). Assessment dimensions and situations. In P. Jansen and F. de Jongh (Eds.). *Assessment Centres: A Practical Handbook* (pp. 21-31). Chichester: Wiley.

Jansen, P. and de Jongh, F. (1997). *Assessment Centres: A Practical Handbook.* Chichester: Wiley.

Jex, S.M. (1998). *Stress and Job Performance: Theory, Research and Implications for Managerial Practice.* Thousand Oaks, CA: Sage.

Johnsen, B.H., Eid, J., Pallesen, S., Bartone, P.T. and Nissestad, O.A. (2009). Predicting transformational leadership in Naval Cadets: effects of personality hardiness and training. *Journal of Applied Social Psychology, 39,* 2213-2235.

Johnson, C. and Blinkhorn, S. (1994). Desperate measures. *The Psychologist, 4,* 167-170.

Johnson, C.F., Wood, R. and Blinkhorn, S.F. (1988). Spuriouser and spuriouser: the use of ipsative personality tests. *Journal of Occupational Psychology, 61,* 153-162.

Jones. A. (1991). The contribution of psychology to military officer selection. In R. Gal and H. Mangelsdorff (Eds.). *Handbook of Military Psychology* (pp. 63-80). Chichester: John Wiley.

Jones, A. and Harrison, E. (1982). Predictions of performance in initial officer training using reference reports. *Journal of Occupational Psychology*, *55,* 35-42

Jones, A., Herriot, P., Long, B. and Drakeley, R. (1991). Attempting to improve the validity of a well-established assessment centre. *Journal of Occupational Psychology, 64,* 1-21.

Judge, T.A., Ilies, R. and Colbert, A.E. (2004). Intelligence and leadership: a quantitative review and test of theoretical propositions. *Journal of Applied Psychology, 89*, 542-552.

Judge, T.A., Bono, J.E., Ilies, R. and Gerhardt, M.W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology, 87,* 765-780.

Judge, T.A., Piccolo, R.F. and Kosalka, T. (2009). The bright and dark sides of leader traits: a review and theoretical extension of the leader trait paradigm. *The Leadership Quarterly, 20,* 855-875.

Kaiser, R.B., Hogan, R. and Craig, S.B. (2008). Leadership and the fate of organizations. *American Psychologist, 63,* 96-110.

Kelly, G.A. (1954). *The Psychology of Personal Construct (Vol. 1).* New York: Norton.

Kingston, N. (1971). *Selecting Managers: A survey of current practice in 200 companies.* British Institute of Management Survey, Report No. 4.

Kobasa, S.C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Journal of Personality and Social Psychology. 37,* 1 -1 1.

Kobasa, S.C., Maddi, S.R. and Zola, M.A. (1983). Type A and hardiness. *Journal of Behavioral Medicine, 6,* 41-51.

Kyllonen, P.C. (1996). Is working memory capacity Spearman's g? In I. Dennis and P. Tapsfield (Eds.). *Human Abilities: Their Nature and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kyllonen, P.C. and Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

Landsbergis, P.A., Schnall, P.L., Belkic, K.L., Baker, D., Schwartz, J.E. and Pickering, T.G. (2003). The workplace and cardiovascular disease: relevance and potential role for Occupational Health Psychology. In J.C. Quick and L.E. Tetrich (Eds.). *Handbook of Health Psychology* (pp. 265-287). Washington, DC: American Psychological Association.

Lazarus, R.S. and Folkman, S. (1984). Stress, Appraisal and Coping. New York: Springer.

Legge, K. (1995). HRM: rhetoric, reality and hidden agendas. In J. Storey (Ed.). *Human Resource Management: A Critical Text* (pp. 33-59). London and New York: Routledge.

Lord, W. and Rust, J. (2003). The Big Five revisited: where are we now? *Selection and Development Review*, *19*, 15-18.

Lyne, K.D., Barrett, P.T., Williams, C. and Coaley, K. (200). A psychometric evaluation of the Occupational Stress Indicator. *Journal of Occupational and Organizational Psychology, 73,* 195-220.

Mael, F.A. (1994). If past behavior really predicts future, so should biopdata's [sic]. In M.G. Rumsey, C.B. Walker and J.H. Harris (Eds.). *Personnel Selection and Classification* (pp. 273-291). Hillside, NJ: Lawrence Erlbaum.

Matthews, G. (1997). The Big Five as a framework for personality assessment. In N. Anderson and P. Herriot (Eds.). *International Handbook of Selection and Assessment* (pp. 475-492). Chichester: Wiley.

Matthews, G, Dreary, I.J. and Whiteman, M.C. (2009) *Personality Traits* (3 ed.). Cambridge: Cambridge University Press.

McClelland, D.C. (1987). *Human Motivation*. Cambridge: Cambridge University Press.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. and Maurer, S.D. (1994). The validity of employment interviews: a comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599-616.

McGrath, J.E. (1976). Stress and behavior in organizations. In M. Dunnette (Ed.). *Handbook of Industrial Organizational Psychology* (pp. 1351–1395). Chicago: Rand McNally Co.

McHenry, J.J., Hough, L.M., Toqiam, J.L., Hansan, M.A. and Ashworth, S. (1990). Project A validity results: the relationship between predictors and criteria domains. *Personnel Psychology, 43,* 335-354.

McQitty, L.L. (1957). Elementary linkage analysis for isolating orthogonal types and typical relevances. *Educational and Psychological Measurement, 17,* 207-229.

Milgram, A.N. (1991). Personality factors in military psychology. In R. Gal and H. Mangelsdorff (Eds.), *Handbook of Military Psychology* (pp. 556-572). Chichester: John Wiley.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Mitroff, I.I. (1974). *The Subjective Side of Science*. Amsterdam: Elsevier.

Morgeson, F.P., Campion, M.A., Diboye, R.L., Hollenbeck, J.R., Murphy, K, and Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personal selection. *Personnel Selection, 60,* 1029-1049.

Murphy. K. (1997). Meta-analysis and validity generalization. In N. Anderson and P. Herriot (Eds.), *International Handbook of Selection and Assessment* (pp. 323-342). Chichester: Wiley.

Murphy, K. (2003). Validity Generalization: A critical review. Mahwah, NJ: Erlbaum.

Murray, H.A. (1981). *Endeavors in Psychology: selections from the personology of Henry A. Murray* (E.S. Shneidman, Ed). New York: Harper and Row.

Norman, W.T. (1969). 'To see ourselves as ithers see us!': relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research. 4,* 417-443.

Ones, D.S., Dichert, S., Viswesvaran, C. and Saldago, J.F. (2010). Cognitive abilities. In J.L. Farr and N.T. Tippins (Eds.). *Handbook of Employee Selection* (pp. 255-275). New York: Routledge.

Ones, D.S., Viswesvaran, C. and Dilchert, S. (2005). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson and O. Voskuijl (Eds.). *The Blackwell Handbook of Personnel Selection* [online]. Oxford: Blackwell Publishing.

Ones, D.S., Viswesvaran, C. and Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679–703.

Orasanu, J.M. and Backer, P. (1996). Stress and military performance. In J. E. Driskell and E. Slater (Eds.). *Stress and Human Performance* (pp. 89-125). Mahwah, NJ: Erlbaum.

Oswald, F.L. and Hough, L.M. (2011). Personality and its assessment in organizations: theoretical and empirical developments. In S. Zedeck (Ed.). *APA Handbook of Industrial and Organizational Psychology: Volume 2 Selecting and Developing Members for the Organization* (pp. 153-184). Washington, DC: The American Psychological Association.

Quillian-Wolever, R.E. and Wolever, M.E. (2003). Stress management at work. In J.C. Quick and L.E. Tetrich (Eds.). *Handbook of Health Psychology* (pp. 355-378). Washington, DC: American Psychological Association.

Raven, J.C. (1963). *Standard Progressive Matrices*. London: Lewis.

Ree, M.J. and Earles, J.A. (1991). Predicting training success: not much more than *g. Personnel Psychology, 44,* 321-332.

Ree, M.J., Earles, J.A. and Teachout, M.S. (1994). Predicting job performance: not much more than *g. Journal of Applied Psychology, 79,* 518-524.

Reilly, R.R. and Chao, G.T. (1982). Validity and fairness of some alternative selection procedures. *Personnel Psychology, 35,* 1-62.

Robertson, I.T. and Makin, P.J. (1986). Management selection in Britain: a survey and critique, *Journal of Occupational Psychology*, *59,* 45-57.

Robertson, I.T. and Smith, M. (1989). *Advances in Selection and Assessment.* Chichester: Wiley.

Robertson, I.T., Baron, H., Gibbons, P., MacIver, R. and Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology, 73,* 171-180.

Rodger, A. (1952). *The Seven Point Plan*. London: NIIP.

Roe, R.A. (2005). The design of selection systems: context, principles, issues. In A. Evers, N. Anderson and O. Voskuijl (Eds.). *The Blackwell Handbook of Personnel Selection* [online]. Oxford: Blackwell Publishing.

Rotter, J.B. (1975). Some problems and misconceptions related to the construct of internal versus external locus of control of reinforcement. *Journal of Counselling and Clinical Psychology, 43,* 26-67.

Russell, M.T. and Karol, D.L. (1994). *16PF5: Administrator's Manual*. Windsor: NFER.

Salgado, J.F. (1997). The Five-Factor Model of personality and job performance in the European Community. *Journal of Applied Psychology, 82,* 30-43.

Salgado, J.F. (1998). Big Five Personality dimensions and job performance in army and civil occupations: a European perspective. *Human Performance, 11*, 271-228.

Salgado, J.F. and Anderson, N. (2002). Cognitive and GMA testing in the European Community: issues and evidence. *Human Performance, 15*, 75-96.

Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C. and De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: a European Community meta-analysis. *Personnel Psychology, 56,* 573-605.

Schmidt, F.L. and Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529-540.

Schmidt, F.L. and Hunter, J.E. (1981). Employment testing: old theories and new research findings. *American Psychologist, 36,* 1128-1137.

Schmidt, F.L. and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, F.L. and Hunter, J.E. (2002). Meta-analysis. In N. Anderson, D.S. Ones, H.K. Sinangil and C. Viswesvaran (Eds.). *Handbook of Industrial, Work and Organizational Psychology, Volume 2: Organizational Psychology* (pp. 51-70). Thousand Oaks, CA: Sage.

Schmidt, F. L., Ones, D. S. and Hunter. J. E. (1992) Personnel selection. *Annual Review of Psychology, 43,* 627-70.

Schmitt, N.W., Arnold, J.D. and Nieminen, L.R.E. (2010). Validation strategies for primary studies. In J.L. Farr and N.T. Tippins (Eds.). *Handbook of Employee Selection* (pp. 51-71). New York: Routledge.

Schmitt, N., Gooding, R.Z., Noe, R.A. and Kirsch, M. (1984). Metaanalyses [sic] of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-442.

Searle, R.H. (2003). *Selection and Recruitment: a critical text.* Milton Keynes: Open University/Palgrave.

Seegers, J. (1997). What is an assessment centre? P. Jansen and F. de Jongh (Eds.). *Assessment Centres: A Practical Handbook* (pp. 3-17). Chichester: Wiley.

SHL. (1999). Occupational Personality Questionnaire (OPQ 32) Manual and Users Guide. Thames Ditton: SHL Group.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. Tokyo: McGraw-Hill Kogakusha.

Silvester, J. and Dykes, C. (2007). Selecting political candidates: a longitudinal study of assessment centre performance and political success in the 2005 UK General Election. *Journal of Occupational and Organizational Psychology, 80,* 11-25.

Slaven, G. and Flin, R. (1997). Selecting managers for a hazardous environment. *Disaster Prevention and Management, 6,* (336-342).

Sonnetag, S. and Frese, M. (2003). Stress in organizations. In W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds.). *Handbook of Psychology: Volume 12 Industrial and Organizational Psychology* (pp. 453-491). Hoboken, NJ: John Wiley and Sons.

Sternberg, R.J. (1985). *Beyond IQ: a Triarchic Theory of Human Intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1997). Tacit Knowledge. In N. Anderson and P. Herriot (Eds.). *International Handbook of Selection and Assessment* (pp. 201-213). Chichester: Wiley.

Tetrich, L.E., Perrewe, P.C. and Griffin, M. (2010). Employee work related health, stress, and safety. In J.L. Farr and N.T. Tippins (Eds.). *Handbook of Employee Selection* (pp. 531-549). New York: Routledge.

Tett, R.P., Jackson, N.J. and Rothstein, M (1991). Personality measures as predictors of job-performance. *Personnel Psychology, 44,* 703-742.

Theorell. T. (2003). To be able to exert control over one's own situation: A necessary condition for coping with stressors. In J.C. Quick and L.E. Tetrich (Eds.). *Handbook of Health Psychology* (pp. 201-209). Washington, DC: American Psychological Association.

Thurstone, L.L. (1934). The vectors of the mind. *Psychological Review*, 41, 1-31.

Tupes, E.C. and Christal, R.E. (1992). *Recurrent personality factors based on trait ratings.* Reprint of original technical report in a Special Issue on the Five-Factor Model: Issues and Applications, *Journal of Personality*, *60,* 225-251.

Wood, S.A. (2003). Five reasons to be optimistic about the future of personality assessment as a selection tool. *Selection and Development Review, 19,* 12-14.

Woodruffe, C. (1993). *Assessment Centres: Identifying and Developing Competence*, (2 ed.). London: Institute of Personnel and Development.

Woodruffe, C. (2005). Emotional factors as selection criteria. . In A. Evers, N. Anderson and O. Voskuijl (Eds.). *The Blackwell Handbook of Personnel Selection* [online]. Oxford: Blackwell Publishing.

Vernon, P.E. and Parry, J.B. (1949). *Personnel Selection in the British Forces.* London: University of London Press.

Vernon, P.E. (1961). *The Structure of Human Abilities* (2 ed.). London: Methuen.

Viswesvaran, C. (2002). Assessment of individual job performance: A review of the past century and a look ahead. In N. Anderson, D.S. Ones, H.K. Sinangil and C. Viswesvaran (Eds.). *Handbook of Industrial, Work and Organizational Psychology Volume 1: Personnel Psychology* (pp. 110-126). Thousand Oaks, CA: Sage.

Viswesvaran, C. and Ones, D.S. (2002). Perspectives on modelling job performance. *International Journal of Selection and Assessment, 8,* 216-226.

Viswesvaran, C. and Ones, D.S. (2010). Employee selection in changing times. In G. P. Hodgkinson and J. K. Ford (Eds.). *International Review of Industrial and Organizational Psychology (Vol. 25)* (pp. 169-226). Chichester: John Wiley and Sons.

Von Glinow, M.A., Driver, M.J., Brousseau, K. and Prince, J.B. (1983). The design of a career-oriented HR system. *Academy of Management Review. 8,* 22-32.

Young, M. and Dulewicz, V. (2003). Command, leadership, and management competencies: predicting superior performance in the Royal Navy. *Henley Working Paper 0317.* Henley: Henley Management College.

Young, M. and Dulewicz, V. (2008). Similarities and differences between leadership and management; high performance competences in the British Royal Navy. *British Journal of Management, 19,* 17-32.

Zimprich, D., Allemand, M. and Dellenbach, M. (2000). Openness to Experience, fluid intelligence, and crystallized intelligence in middle-aged and old adults. *Journal of Research in Personality, 43,* 444-454.

Zuckerman, M., Kuhlman, D.M., Joireman, J., Teta, P. and Kraft, M. (1993). A comparison of the three structural models of personality: the Big Three, the Big Five, and the alternative five. *Journal of Personality and Social Psychology, 65,* 757-769.

Appendices

List of Appendices   Page

Appendix A: Tests of Cognitive Ability

Psychometric tests which attempt to measure a person's ability or aptitude have a long history in both education and psychology stretching back to the end of the Nineteenth Century. When Freud was developing his psychodynamic theory of personality, other people in Europe and the United States were starting the testing movement. The success of the mass testing of armed service recruits in the United States during the First World War led to their adoption in the United Kingdom in the Second World War. Rust and Golombok (1989) ask what psychometric tests measure and state:

> Psychological and educational tests carry out a form of measurement but, unlike physical measures such as length or weight, there is considerable confusion over what they measure and how they are able to do so. One particular problem is that what is measured is not a physical object but an intervening construct or a hypothetical entity (p. 26).

There is no standard definition of the term intelligence. Vernon (1961) notes that:

> It is psychologists again who, although they have been testing intelligence with some success for over forty years, have failed to reach any agreed definition as to what they are measuring … In a famous symposium published in 1921, thirteen[32] psychologists gave thirteen different views (p. 3).

Boring (1923) got around the problem by defining intelligence as the thing that intelligence tests measure! Whilst this may seem a flippant comment, Sternberg (1990) argues that it is best to understand the very different theories of intelligence as 'metaphors as a way of viewing the mind' (p. 3). He notes that 'There may be as many different definitions of intelligence as there are people who are asked to define it' (p. 33). It might be worth looking at a few of the major debates on the issue; but this is only skimming a very thick surface. Spearman invented factor analysis (more on this topic later) to help him analyse the underlying aspects of people's scores on different types of ability and aptitude tests. In 1904 Spearman concluded that people's scores on these apparently very different tests were consistent across the tests. His factor analysis showed one general factor, which he labelled 'g', and a factor which was 'specific' to the particular test (Spearman, 1927).

Thurstone (1924) developed the methods of factor analysis further than Spearman and rotated the axes to get a simpler structure, but one which would 'preclude the identification of a general factor' (Carroll, 1993, p. 55). Thurstone found seven 'primary' factors (correlated multiple factors) which were eventually named: Verbal, Reasoning, Number, Spatial, Perceptual Speed, Memory, and Word Fluency. Cronbach (1990) notes that Thurstone's use of the term primary 'deliberately suggested that the group factors combine in various proportions to produce any complex intellectual process, just as green, red, and blue spotlight can be mingled to produce any other hue or white' (p. 383).

The two basic approaches of Spearman and Thurstone, which were often represented as the 'British' or 'American' approach to intelligence, continued until after the Second

---

[32] Sternberg (1990) actually lists the 'fourteen experts' in this symposium which included Thorndike, Terman and Thurstone.

World War. More recent American psychologists like Sternberg (1997) have argued that 'there is good evidence to suggest that measures of general intelligence (so called 'g') provide the single best predictor of success across jobs' (p. 201). The concept of g and of general intelligence have been controversial, mainly due to the problems of adverse impact with various groups, and the arguments of some researchers that there is a large genetic component of intelligence, for example Jensen (1980). Ironically, it was Jensen's championing of g in the United States which resurrected interest in the concept of 'general mental ability' and the use of the abbreviation 'GMA' in the reports of many recent meta-analytical studies.

A hierarchical model of intelligence was proposed by Vernon (1961). He notes that this was suggested by Burt and the model is usually shown as an upside-down tree diagram with g at the top under which are two 'major group factors'; one v:ed (verbal:educational) and the other k:m (spatial:mechanical). This structure is given in Figure A.1 and shows some of the minor group factors at the lower level labelled under each major factor such as verbal reasoning, numerical reasoning, etc. These in turn link to specific factors which the tests might measure. For example, numerical reasoning would be composed of specific factors such as addition, division, algebra, geometry, etc. Vernon (1961) also produced a series of diagrams to show how this hierarchy changes and integrates with other domains (like education, psychological tests and occupations) depending on the domains. Vernon comments that this scheme was only approximate and that the 'strict hierarchical picture of mental structure is an over-simplification' (p. 25).

Figure A.1: Vernon's hierarchical model of intelligence.
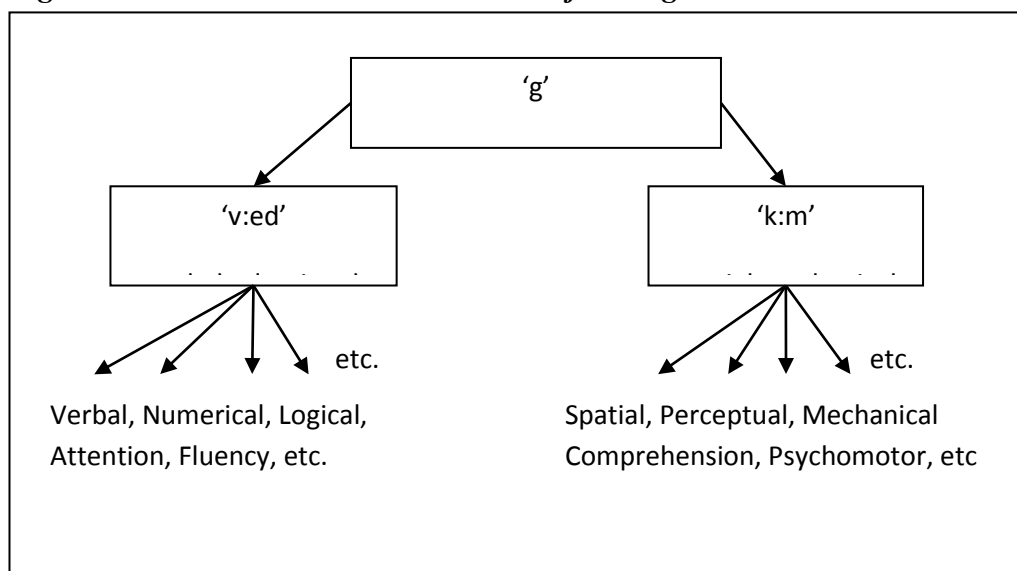
Other early theories of intelligence which have been influential are Cattell's idea that the general intelligence factor can be divided into 'fluid' and 'crystallized' intelligence. Cattell (1987) links fluid intelligence to the basic abilities, reasoning and mental processes, whereas crystallized intelligence is that developed through interaction and experience. Guilford (1967) proposed a theory of the structure of the intellect which

meant that any test measuring intelligence would have a type of 'content' on which the person performed a type of 'operation' and this would result in an outcome or 'product'. This structure is often shown as a cube with content, operations and products along the facets. According to Guilford, there are five kinds of contents (visual, auditory, symbolic, semantic, behavioral), five kinds of operations (cognition, memory, divergent production, convergent production, evaluation) and six kinds of products (units, classes, relations, systems, transformations, and implications) and since the dimensions are independent, there may be as many as 150 different factors of intelligence.

More recently Sternberg (1985) has produced a 'triarchic' theory of intelligence. Sternberg's three components are: analytical intelligence referring to the problem-solving abilities brought to bear on the problem; creative intelligence, involving the ability to deal with new situations using past experiences, current ability and skills; and practical intelligence, which is the ability to adapt to a changing environment and to apply the other two components. Carroll (1993) has produced a 'three-stratum' theory which is not unlike Vernon's hierarchical model but does not recognize the v:ed or k:m factors at the second level. Carroll notes that this level 'contains perhaps as many as a dozen broad factors with varying generality over the cognitive domain' (p. 638). His theory does recognize the broad factor g at the top level. Other theories, like the 'theory of multiple intelligences' devised by Gardner (1983) and of 'emotional intelligence' by Goleman (1995) have had some influence recently on professional practice, though support for their theories has still not been well established.

In practical terms the difference between theoretical approaches to cognitive testing has much less impact than for different personality theories on the type of testing or outcome for the candidate. Whatever the theory, a range of abilities or aptitudes are suggested which could make up an overall assessment of a person's ability to learn and the tester can choose from a range of available tests. In the field of personality, as will be shown later, the theory and test (or inventory) are more integrated. With lower level jobs, or even highly skilled technician type jobs, a job analysis will reveal that a few specific tests can help match the person to the job. With higher order jobs, like management jobs or recruiting new graduates to fulfil a wide range of roles, a test battery covering a wider range of abilities and aptitudes may be required.

One final point on this subject: Kyllonen and Christal (1990) have proposed that there is a link between intelligence and information processing theory and that there is a very high correlation between reasoning task scores and working memory capacity. This led Kyllonen and Christal to point out, as their paper title puts it, that: 'reasoning ability is (little more than) working-memory capacity?!' (p. 389).

References

A References list for material in this Appendix is given the end of Appendix B.

Appendix B: Personality Theories and Inventories

In Appendix A the difficulty of defining intelligence was mentioned. There may be even less point in trying to perform this task for the concept of personality. Before the Second World War, Gordon Allport (1937) found almost fifty different usages of the term personality in the psychological literature. Hall and Lindzey (1978) note that in personality theory:

> Once the individual has created or adopted a given theory of personality, their definition of personality will be rather clearly limited by the theory. Thus, we submit that *personality is defined by the particular empirical concepts that are a part of the theory of personality employed by the observer*. Personality consists concretely of a set of scores or descriptive terms that describe the individual being studied in terms of the variables or dimensions that occupy a central position within the particular theory utilized (p. 9, original emphasis).

Most modern theories of personality have been developed fairly atheoretically and rely on a mathematical technique called factor analysis to establish the personality dimensions. This technique has already been mentioned in the intelligence theory field and was developed by Spearman, Thurstone, Guilford and others, to produce and refine their own theories. In the personality field too, factor analysis is used to develop the personality inventories from a large number of possible inventory items. These items are typically trait words ('shy', 'nervous', 'creative', etc.) or self-descriptions ('I often feel nervous in new company', 'I like to complete the work I have set myself each day', etc.). A large number of people rate the proposed inventory items and this data is then factor analysed. Items which 'load' on the resultant factors are kept for the next stage of the research and those items which do not 'load' are discarded or re-written. This process may go through several iterations before the final version of the inventory is produced. It can be seen that this type of procedure may exclude some important personality traits which only had a few items in the original set.

The mathematical technique of factor analysis enables researchers to reduce large amounts of data to give a more parsimonious description in terms of a small number of factors or dimensions, but the apparent attraction of numerical precision hides many pitfalls. When using factor analysis it is important to ensure that the data being used meets certain underlying requirements (normality, skew, kurtosis, etc.) and that a sample size of several hundred is used (Bennett and Bowers, 1976) because at this stage the confidence limits of the underlying correlation coefficients become more stable.

Many decisions have to be made when using factor analysis and some of these decisions are arbitrary rather than rigorous, requiring intuition rather than scientific analysis. Often the factor analyst will try several methods and then chose the method which best fits the purpose he or she wants. For example, basic decisions have to be made about what kind of factor analysis to use (e.g., principal components or principal factors), how many factors to extract, whether the factors are orthogonal or oblique, whether to rotate the factors and which analytical method of rotation will be used (e.g., varimax, quartimax, etc.).

At the beginning of this section on personality measures Hall and Lindzey (1978) were quoted indicating that the definition of personality varied depending on the theorist's perspective. Most of the different personality frameworks used today depend on the factor analytic method adopted by the personality researchers, in the same way that Spearman and Thurstone produced varying structures for intelligence. Two of the most important personality theories until the late twentieth century were those of Eysenck and Cattell. Eysenck proposed first a two factor theory and then enlarged this to three; Cattell, on the other hand, found a sixteen factor solution. Eysenck used factor analysis to refine his questionnaire items, whereas Cattell used factor analysis to generate the factors from various data sources including a massive list of trait words. This same trait list (Allport and Odbert, 1936) has continued to be used by later theorist.

Eysenck developed a theory of personality which, he claimed, was able to account for some of the variance in human personality and this is firmly based on a theory of biological determinants which are mainly inherited. The dimensions Eysenck found were Extraversion and Neuroticism; later on Psychoticism was added as the third dimension. Eysenck (1976; Eysenck and Eysenck, 1985) is at pains to point out that the three dimensions of personality postulated in his theory are not conceived as being the only dimensions of personality, nor do they account for the whole of personality.

The main difference between Eysenck's theory and Cattell's (1987, 1989) sixteen personality factors theory rests on the type of factor analysis employed. Eysenck endeavoured to keep his factors 'orthogonal'. This means that the mathematical axes are at right angles to each other and so the factors (personality dimensions) are not correlated with each other. Those questionnaire items which are not orthogonal and load on more than one factor are removed. Cattell, on the other hand, believed that the factors themselves might be correlated and so used an oblique rotation method for the axes (not at right angles) to give more factors which are correlated with each other.

Another problem with factor analysis is what you do with the factors once they are identified. Factors are mathematical entities but their meaning has to be explained to other people and be given a label. This label is usually a word which summarises a number of items which have a significant factor loading on the particular factor; for example, 'extraversion'. In reality the factors extracted are hypothetical, mathematical constructs, but once a label is given to a construct it can become reified as a concrete entity. Eysenck relied on traditional words in psychology borrowed from Jung (1928) who had coined the words 'extravert' and 'introvert' sixteen years earlier. Cattell, on the other hand, was very aware of the problems of labelling so he originally used neologisms for his factor labels. Later on, Cattell reverted to using everyday terms since the words he made up were difficult for other researchers and practitioners to follow. In fairness to Cattell, it must be noted, that he refers to the factors by their identifying letters rather than to the label. However, this does not stop users of his personality inventory from applying these factor labels literally and rigidly.

Hall and Lindzey (1978) laid down criteria for a good personality theory. One of these was that the theory should generate research. The theories of both Eysenck and Cattell have met this criterion. Although Eysenck and Cattell were at one time the best known personality theorists, other researcher produced personality measures using factor analysis which proved useful in research. For example:

- Guilford - ten orthogonal dimensions (Guilford, 1959);
- Comrey - six factors later increased to eight and renamed (Comrey, 1962; Comrey and Jamison, 1966);
- Norman – 'five relatively orthogonal, easily interpreted, personality factors': extraversion, agreeable, conscientiousness, emotional stability, culture (Norman 1963 and 1969).

References for Appendices A and B

Allport, G.W. (1937). *Personality: a psychological interpretation*. New York: Holt.

Allport, G.W. and Odbert, H.S. (1936). Trait names: a psycho-lexical study. *Psychological Monographs, 47,* 1-171.

Bennett, S. and Bowers, D. (1976). *An Introduction to Multivariate Techniques for Social and Behavioural Sciences.* London: Macmillan.

Boring, E.G. (1923). Intelligence as the tests test it. *New Republic, 35,* 35-37.

Carroll, J.B. (1993). *Human Cognitive Abilities: a survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Cattell, R.B. (1987). *Intelligence: its structure, growth, and action*. New York: Elsevier.

Cattell, R.B. (1989). *The 16PF: personality in depth.* Champaign, IS: IPAT.

Comrey, A.L. (1962). A study of thirty-five personality dimensions. *Educational and Psychological Measurement, 22,* 543-552.

Comrey, A.L. and Jamison, K. (1966). Verification of six personality factors. *Educational and Psychological Measurement, 26,* 945-953.

Cronbach, L.J. (1990). *Essentials of Psychological Testing*. New York: Harper Collins.

Eysenck, H.J. (1976). *The Measurement of Personality.* Lancaster: MTP Press.

Eysenck, H.J. and Eysenck, M.W. (1985). *Personality and Individual Difference*: A Natural Science Approach. New York: Plenum.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Goleman, D. (1995). *Emotional Intelligence: Why it can matter more than IQ.* New York: Bantam Books.

Guilford, J.P. (1959). *Personality.* New York: McGraw-Hill.

Guilford, J.P. (1967). *The Nature of Human Intelligence.* New York: McGraw-Hill.

Hall, C.S. and Lindzey, G. (1978). *Theories of Personality* (3 ed.). New York: John Wiley and Sons.

Jensen, A. (1980). *Bias in Mental Testing,* New York: Free Press.

Jung, C.G. (1928). *Psychological Types.* Reprinted in B. Semeonoff (Ed.). *Personality Assessment.* Harmondsworth: Penguin, 1966.

Kyllonen, P.C. and Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

Norman, W.T. (1963). Towards 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. (1969). 'To see ourselves as ithers see us!': relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research. 4,* 417-443.

Rust, J. and Golombok, S. (1989). *Modern Psychometrics: the science of psychological assessment*. London: Routledge.

Spearman, C. (1927). *The Abilities of Man.* London: Macmillan.

Sternberg, R.J. (1985). *Beyond IQ: a Triarchic Theory of Human Intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1990). *Metaphors of Mind: conceptions of the nature of intelligence.* Cambridge: Cambridge University Press.

Sternberg, R.J. (1997). Tacit Knowledge. In N. Anderson and P. Herriot (Eds.). *International Handbook of Selection and Assessment* (pp. 201-213). Chichester: Wiley.

Thurstone, L.L. (1924). *The Nature of Intelligence.* Westport, CT; Greenwood Press [1973 reprint].

Vernon, P.E. (1961). *The Structure of Human Abilities* (2 ed.). London: Methuen

Appendix C: Cooper's Model of Organizational Stress

The basic model by Cooper (1986) presented as Figure 3.1 (p. 39) is expanded below. [Reproduced with permission from the British Psychological Society]

Sources of Stress at Work
- Intrinsic to the job
    - Poor physical work conditions
    - Work overload
    - Time pressures
    - Responsibility for lives
- Role in the organization
    - Role ambiguity/conflict
    - Image of occupational role
    - Boundary conflicts
- Career development
    - Over promotion
    - Under promotion
    - Lack of security
    - Thwarted ambition, etc
- Relationships at work
    - Poor relations with boss, subordinates, or colleagues
    - Difficulty in delegating responsibility, etc
- Organizational structure and climate
    - Little or no participation in decision making
    - Restrictions on behaviour (budgets, etc)
    - Office politics
    - Lack of effective consultation, etc.

The Individual
- Level of neuroticism
- Tolerance of ambiguity
- Type a behaviour

Home-work Interface
- Family problems
- Dual-career marriages
- Life crises

Individual Symptoms of Stress
- High blood pressure
- High cholesterol
- Heart rate
- Smoking
- Depressive mood
- Escapist drinking
- Job dissatisfaction
- Reduced aspirations, etc

Outcomes include: disease, chronic heart disease, mental ill health

Organizational Symptoms of Stress
- High labour absenteeism
- High labour turnover
- Industrial relations difficulties
- Poor quality control

Outcomes include: prolonged strikes, frequent and severe accidents, chronically poor performance

Appendix D: The Admiralty Interview Board Process

Before candidates arrive at the Board they will have been through a long process which includes contact with a young serving careers officer. They will have been encouraged to visit a unit and will have been briefed about attending the Admiralty Interview Board, including watching a video on the subject[33]. Candidates who wish to join the Royal Marines will have passed a gruelling three day physical Board at Lympstone, near Exmouth in Devon.

The candidates arrive on the afternoon of the first day. This enables candidates to travel to the Board, situated in Gosport in Hampshire, from most parts of the country. The candidates are shown to their room (their cabin!) and then meet the other candidates and are briefed about the next two days. They fill in a biographical questionnaire which gives the Board an up-to-date record of their qualifications and achievements. It also asks about positive and negative aspects of service life and about participation in sports, clubs and hobbies. Much of this information feeds into a biodata measure.

The first morning is taken up with testing the whole group of candidates (twelve people) all in one room. Before the tests start the candidates are welcomed and briefed by a Board President, a serving Royal Navy Captain. The first two tests are a test of general knowledge followed by the four tests which make up the psychometric test battery (a verbal test, a non-verbal test, a clerical accuracy test, and a numerical test). Next the candidates complete a spatial awareness test which identifies whether candidates can picture and manipulate views in their heads (they are not allowed to draw during this test). This is followed by a diagnostic written communication skills test which can identify errors in grammar, spelling, etc. Finally, after a break, they tackle an essay and a précis. A visiting headteacher or university lecturer who takes part in the full as a Board assessor in the process marks these two pieces of work.

After lunch the candidates rehearse the two tasks they will perform in front of the assessors on the following morning: the Discussion Planning exercise and Gym tasks. The Gym tasks involve taking equipment and people over obstacles. Two out of the four Gym exercises take place over a large tank of water. The candidates get a chance to try out these exercises and learn to swing on ropes and use cantilevers. There are also briefings on safety. The candidates wear overalls and helmets with numbers during these exercises.

In the Discussion Planning exercise a group of four candidates is given a scenario where something has goes wrong and they have to decide what to do. For example, they are leading a climbing expedition on an island and someone falls and breaks their leg. They have to construct a plan given various options. The exercise always includes a map and calculations of times, speeds and distances. The candidates are allowed time to study the scenario and develop a plan on their own before they get together with the rest of the

---

[33] The video used to brief candidates about the Admiralty Interview Board procedure is available on YouTube.

group to discuss their individual plans and agree a group plan. After this process they are quizzed fairly insistently by the Personnel Selection Officer, both individually and as a group. Finally, they each give a one minute summary to the Board on their plan. This completes the second day.

The actual assessment of candidates by Board members is completed during the following morning when the assessors, the Board Members, watch the Gym tasks and Discussion Planning exercise. They observe and rate the candidates on various dimensions for each exercise. After each exercise the Board members give a set of marks for each candidate and share their thoughts about candidates with the Board. If the marks on these tasks are not within certain limits the Board Members will discuss these further to try to resolve the differences. As the final part of the assessment process the four candidates are interviewed individually. One interview takes place with a Personnel Selection Officer and the other interview is with the other three members of the Board.

When the interviews are finished the Board Members look at the various pieces of information collected on each candidate and discuss them. During this final process the Board members use an information presentation sheet for each candidate on which they transfer the marks given to a candidate throughout the process. Only the Personnel Selection Officer, who is one of the four Board members, knows the psychometric test results during the early stages of the assessment process. The Personnel Selection Officer is trained to interpret the psychometric tests and does not disclose these results to the other three Board members until just before the final interview. This is to avoid prejudicing the way the candidate may be viewed and assessed during the Gym and Planning exercises. An overall composite test score is calculated from the four tests which make up the test battery. A biodata score is also calculated (by computer) based on the biographical details which candidates gave on their first afternoon.

The composite test score and the biodata score have been shown to be the best single predictors of success in future training. On the information presentation sheet these figures are put into the largest boxes to illustrate their predictive success. These two scores feed in to the Effective Intellect dimension of the final assessment, which is the best predictor of success in training. Other characteristics of effective intellect which were observed in other parts of the assessment process, such as the essay, précis, Gym and Planning exercise and interview can moderate the effective intellect score, but not by much without good reason. On occasions the Board has seen highly intelligent candidates (with a PhD, say) who have scored top mark on both the Composite Test Score and biodata score. However, these candidates have been totally ineffective in the Gym and Planning exercises. The Board members problem is to decide what the Effective Intellect mark should be. Have they seen any indications that this highly intelligent, but otherwise ineffective candidate (in this practical situation), can be trained to use their natural intelligence and succeed in training?

The other three dimensions are Leadership Potential, Character and Personality (moderate predictors) and Service Motivation (weak predictor). These again are built up from marks on the exercises and interview. Using the three strongest predictor dimensions a Final Board Mark is agreed. There are again, as in the rest of the process, strict procedures and protocols in place for arriving at a Final Board Mark. The final mark is arrived at just before lunchtime.

At this particular assessment centre the candidates are seen individually immediately after the process by the President of each Board and given their results and some feedback. Unsuccessful candidates can also write in for written feedback. Many unsuccessful candidates are encouraged to try again after a year, but those who have done really badly and show no indication that they might improve are discouraged from reapplying. About half the candidates pass the Board, but they then have to wait to see if they have a high enough score to be selected when all the candidates' scores are compared at a meeting about a month before the next training intake is due.

Candidates for pilot and observer roles also have their flying aptitude scores taken into account. These flying aptitude scores consist of a series of psychomotor and reaction time tests taken at the Royal Air Force assessment centre at Cranwell in Lincolnshire. Candidates for flying roles must pass these aptitude tests before they can be assessed at the Admiralty Interview Board. The flying aptitude scores are combined with the Final Board Mark in an arithmetical weighting process.

After lunch the Board members write up their notes on the candidates they have seen. One reason for this is so that feedback can be given to any candidates who write in for help. The Personnel Selection Officer collates the candidates' main marks, which are entered into the computer database. This information is used for a wide variety of management tasks such as validity research and answering Parliamentary Questions. The Board members then spend the rest of the afternoon doing their "homework" which is reading and making notes from the biographical questionnaires for the candidates they will see the following day.

Whilst this gives a flavour of the assessment centre it should be noted that candidates are assessed for different specialisms in the Royal Navy. Usually a Board has four candidates all with roughly the same qualification and of the same age who are applying for the same specialism[34]. Although this only applies to about ten percent of cases, a candidate can be assessed at the same time to be considered for a career as a pilot, warfare officer, engineer officer, supply and secretarial officer or engineering training manager. The psychometric tests and the Gym and Planning exercise are differentially predictive for these different groups. For example, the spatial test is a better predictor for flying duties and warfare roles. For engineers and warfare officers the non-verbal test is a better predictor than for supply and secretarial officers. The clerical test is a better predictor for the latter group, and so on.

---

[34] If women candidates are being assessed they will always be on a board with another woman candidate rather than being the only female.

Appendix E: Marking Guide for Final Board Mark Assessment

Category	Board mark average	Final mark declared	"Big Box" Marks for Effective Intellect, Leadership Potential and Character and Personality
OUTSTANDING (High potential)	980 800	980 950 920 880 850 820	2x9, 1x8 1x9, 2x8 3x8 No 7s 2x8, 1x7 1x8, 2x7 3x7 No 6s
STRONG (Good potential)	790 600	780 750 720 680 650 620	2x7, 1x6 1x7, 2x6 3x6 No 5s 2x6, 1x5 or 1x6 2x5
ADEQUATE (Reasonable potential)	599 550	580 560	1x6, 2x5 or 3x5 3x5 No 4s
(Fair potential with some shortcomings likely to be overcome in normal training)	549 500	540 520	2x5, 1x4 or 1x5, 2x4
DOUBTFUL (Distinct shortcomings which may cause problems in training)	499 450	480 460	1x5, 2x4 or 3x4 No 3s
VERY DOUBTFUL (Major shortcomings likely to cause major problems in training)	449 400	440 420	
UNACCEPTABLE NOW (Limited potential)	399 200	380 350 320 280 250 220	1x4, 2x3 3x3 2x3, 1x2 1x3, 2x2 3x2 No 1s
ENTIRELY UNSUITABLE (Negligible potential)	199 0	180 150 120 080 050 020	2x2, 1x1 1x2, 2x1

Appendix F: Background and Results for Phase One of the Research

[Some of this material (including the three Tables) is reproduced with permission from the Defence Science and Technology Laboratory]

Several ideas were considered about the types of test to be employed in the research including the NASA Cockpit Management Attitudes Questionnaire (Helmreich, 1984) and the DMT (Defence Mechanism Test). It is claimed that the DMT (Kragh, 1960) can detect poor judgement in stressful situations using a tachistoscopic presentation of pictures with emotional content The DMT had been used successfully in the Swedish Air Force to reduce crashes due to pilot error. In addition, MICROPAT (Bartram and Dale, 1991), a computerised battery of tests to measure psychomotor and information processing was considered since computers with this system installed were readily available. MICROPAT is used to select armed service pilots in the UK and civilian pilots in many countries. An indicator of workplace stress, the Occupational Stress Indicator, developed by Cooper *et al.* (1988) was also reviewed.

It was established after contacting the course instructors that the period of time available for testing was limited to a ninety-minute session sometime during the first few days of the course. It would also be very difficult to provide a suitable environment for using some of the more esoteric testing ideas like DMT and MICROPAT. So it was decided that simple paper and pencil measures should be used. The following three measures were selected:

4. <u>A spatial awareness test 'Directions and Distances'</u> which had been introduced into the initial entry selection procedures for Royal Navy Officers at the Admiralty Interview Board in September 1994. A trial of this measure at the Admiralty Interview Board had shown that this spatial test predicted the results of navigational tests at Britannia Royal Naval College and in subsequent specialist navigation training. As a result it became part of the standard testing procedures. Although the test is a paper and pencil test, the four parallel versions of the test were produced using computer algorithms to generate the test items (Dennis *et al.* 2002). The test has 15 items and the test taker is allowed fifteen minutes. Accurate timing of this test is essential.

5. <u>The 16 Personality Factor Questionnaire</u> (Cattell, 1989) which has been used extensively in research, clinical psychology and occupational psychology since 1949 (and is available in over forty languages). The test produces sixteen dimensions or scales of personality. A shortened version of the inventory, Version C, was used for this research. The 16PF (as this inventory is usually referred to) takes about thirty minutes to complete, but there is no time limit.

6. <u>A Locus of Control Inventory</u> (Rotter, 1975). Locus of control is a variable or construct which is believed to account for a variety of behaviours. External locus of control refers to a person's belief that outcomes in a person's life, or the world at large, are not determined by the individual's action but by external factors. On the other hand, internal locus of control suggests that the individual can control events

by their own actions. Research has indicated that people who believe that they can control events may interpret these events as challenging rather than stressful. The inventory takes about ten minutes to complete, but there is no time limit.

Testing and other data collection for this first phase of the research started in March 1993 and was completed in November 1995

All SMCC course members had attended the Royal Navy's assessment centre, the Admiralty Interview Board, before joining the Royal Navy as an officer, so the researcher had selection data available for each course member. This included the results of psychometric and other testing, a Composite Test Score, a grade for performance on the four practical group tasks each led by a different candidate (called the Gym task mark), a mark for performance in a group Discussion Planning exercise, the four Board dimensions and the Final Board Mark. Biodata scores and some GCE grades were available.

The course instructor, who is also the Commanding Officer of the course, makes a decision on whether a particular student passes or fails the course. This can happen at any stage of the course, though it is most likely to be towards the end of the course. It is often the case that the student realises that they are not able to cope with a particular situation and the commanding officer has to take over command of the submarine. Initially, only the individual student's pass or fail result for the course was available to the researcher, but to aid the research the two commanding officers produced an overall Course Grade (A to F) for each student. This included giving grades to students on the course which had recently finished.

In 1996, after three years and six courses 36 sets of data which were available with 28 sets for the spatial test and inventories. The data were analysed using the Statistical Package for the Social Sciences (SPSS). The results showed that the spatial ability test did not predict success on the course. It was noted that the spatial test (a paper and pencil test) had a high correlation with the Verbal ability selection test which the students had taken at the AIB some years before. It appeared likely that the spatial test was measuring verbal rather than spatial ability. In addition, the students on the course had an average of thirteen years' experience of navigation which could have made them fairly equal on this skill.

Only one of Cattell's sixteen personality factors predicted the Pass-Fail result. An analysis of the difference in means between the pass and fail groups showed that students who reported themselves as 'Group-oriented' rather than 'Self-sufficient' were more likely to pass the course [U = 44.5, p = 0.04; two-tailed]. Another 16PF factor 'Suspicious' as opposed to 'Trusting' did not have a significant difference in means for the pass and fail groups, but the scores were associated with the overall A to F Course Grade with a correlation of 0.42 [p = 0.03; two-tailed]. The 'Suspicious' end of this factor includes traits such as hard to fool, distrustful and sceptical, deliberate in their actions, unconcerned about other people and being poor team members. This suspicious

label does not conform to the usual view that Royal Navy officers are friendly, open, cheerful, adaptive team players. There also appeared to be a contradiction between this 'Suspicious' factor and the other predictive 'Group-oriented' factor found. However, whilst these results were statistically significant the contingency tables drawn for these factors showed a lot of misclassification making these results theoretically interesting, but not useful in the practical situation. Table F.1, for example, shows the contingency table for the 'group-oriented versus self-sufficient' factor (Q2). It might be theoretically possible to achieve a 90% pass rate on the course by using the Q2 factor as a selection device and exclude eight unsuitable students whilst only passing one of these, but you would disqualify eight other students who would have been successful on the course. So whilst factor Q2 appears to do a good job by identifying most of the failures it misclassifies almost half the successful candidates.

Table F.1: 16PF Factor Q2 (Group-oriented – Self-sufficient) and the Pass-Fail Result.

		SMCC Pass-Fail Result			
		Fail	Pass	Total	Pass rate
Personality Factor Q2	3 and under	1	11	12	Pass rate 91.7%
	4 and over	8	8	16	Pass rate 50.0%
	Total	9	19	28	Pass rate 67.9%
Note: Higher scores on this factor are for those reporting themselves as Self-sufficient.					

The theory behind the Locus of Control measure suggests that those who attribute their success and failures to internal rather than external factors would be more likely to pass the course. These findings were confirmed with a statistically significant difference between those who passed or failed the course [$U = 51$, $p = 0.04$; one-tailed]. The mean score for those passing the course was 11.8 and for those failing the course was 14.0. The correlation of 0.25, though significant, meant that the prediction was far from perfect. Again, an analysis of the misclassified students, shown in Table F.2, illustrates that using this measure as a selection device would have denied good students the chance of taking the course. In particular, of the six students who would have been excluded from the course by the Locus of Control score, two were graded A and three graded B on the overall SMCC Course Grade.

Table F.2: Locus of Control scores and the Pass-Fail Result.

		SMCC Pass-Fail Result			
		Fail	Pass	Total	Pass rate
Locus of Control Scores	13 and under	2	13	15	Pass rate 86.7%
	14 and over	7	6	13	Pass rate 46.2%
	Total	9	19	28	Pass rate 67.9%

An analysis of the historical selection data from the Admiralty Interview Board proved to be interesting. The results of the Non-verbal test, taken on average 13 years before the course, and the recorded GCE Mathematics grade predicted the overall grading on

the course with statistically significant correlations of 0.39 and 0.51, respectively. Table F.3 shows the actual breakdown of scores of the Non-verbal test and the course results. Again, whilst this is a useful theoretical result the table shows that a lot of good candidates would be rejected, even with a correlation of nearly 0.4 [N= 28]. Non-verbal ability is one of the components of spatial ability (Carroll, 1993). The scores on the spatial test taken on the SMCC course correlated (0.37) with the verbal and non-verbal tests and the Composite Test Score taken, on average, thirteen years before, at significant levels.

Table F.3: Non-verbal test scores and Pass-Fail Result.

		SMCC Pass-Fail Result			
		Fail	Pass	Total	Pass rate
Non-verbal Test Scores	38 and over	4	18	22	Pass rate 81.2%
	37 and under	6	8	14	Pass rate 57.1%
	Total	10	26	36	Pass rate 72.2%

The even higher correlation of the Maths GCE grade as a predictor meant that a Grade A, B or C gave an approximate probability of passing the course of 83%, 71% and 57%, respectively [N= 20]. An attempt was made to analyse the cognitive and personality aspects together using linear regression, but no significant findings were found.

In a report (Beadle, 1997) the researcher concluded that psychometric testing might not be the answer to the problems of failure on the course, and that training, development and appraisal procedures should be examined as ways of reducing failure, but a further investigation with more up-to-date measures might be worthwhile. It was recommended that a different measure of personality should be trialled, with fewer dimensions, such as the use of Five Factor Model of personality. At the time the original research report was being written, a United States services Five Factor Model had become available to Defence psychologists in the UK, who re-named the personality inventory 'OCEAN' (Collis, 1997). In addition, to examine the influence of locus of control further and also look at the some of the wider issue of stress, like Type A Behaviour, the use of the Occupational Stress Indicator (OSI), developed by Cooper *et al.* (1988) was revisited as a possible measure to trial.

References

Helmreich, R.L. (1984). Cockpit management attitudes. *Human Factors, 26,* 63-72.

Kragh, U. (1960). The Defense Mechanism Test: A new method for diagnosis and personnel selection. *Journal of Applied Psychology, 44,* 303-309.

The other references for this Appendix can be found in the main References list pp. 111-119.

Appendix G: AIB Test Statistics

Although sample sizes are not available the minimum number of cases would be about 500 (A year's entry to Dartmouth).

Reliability estimates for AIB Tests

	Split-half	Internal consistency	Test-retest
Verbal test	.97	.83 to .95	
Non-verbal test	.94	.84	
Numerical test		.85	
Instructions test			.92*
* This test is quite complex and so not amenable to split half or internal consistency methods so the test-retest method was the only way to examine reliability.			

Intercorrelations between AIB Tests

	Verbal test	Non-verbal test	Numerical test
Non-verbal test	.36		
Numerical test	.41	.49	
Instructions test	.44	.36	.43

Average scores and standard deviations of AIB Tests

	AIB reliability and validity studies		Present study	
	Mean score	Standard deviation	Mean score	Standard deviation
Verbal test	88.7	16.1	98.8	15.3
Non-verbal test	37.4	6.3	39.2	6.0
Numerical test	23.0	6.8	24.1	6.9
Instructions test	124.5	18.6	131.4	16.9

These figures are taken from an unpublished source: DERA. (1999). *The AIB Board Members' Handbook.* Defence Evaluation and Research Agency, Farnborough, Hants. The figures are averaged over several studies. Most of this Handbook was written by the author of this thesis.

Appendix H: The Predictors Used in Scoring the Biodata (PAP)

The table below shows which biodata items are used in the Professional Aptitude Predictor (PAP) and which are, individually, the stronger and weaker predictors of BRNC Professional Marks based on written examinations.

Strong Predictors:	**Weak Predictors:**
Maths GCSE level grade	Type of 'A' levels
Physics GCSE level grade	No. of years at last school
English GCSE level grade	Age
Number of GCSEs	Marital Status
GCSE level grade points	Number of spare time activities
'A' level grade points	Type of Spare Time Activities pursued
Number of 'A' level subjects failed	Number of games teams played for
Student status	Number of part-time jobs held
Type of degree	Influential information source
Number of full-time jobs held	Main attraction
Current employment status	Number of GCSE subjects failed

Appendix I: Submarine Command Course Assessment

SUBMARINE COMMAND COURSE ASSESSMENT

Name:

Quality grading: 0 – Inadequate
                              1 – Some Weaknesses
                              2 – Satisfactory/Average
                              3 – Good

TACTICAL	**GRADE**	**PERSONALITY**	**GRADE**
Professional Knowledge		Leadership	
Tactical Awareness		Stamina	
Practical Ability		Command Presence	
TWS Technical Knowledge		Bearing and Dress	
Navigation and Pilotage		Intelligence	
Instinctiveness (Safety)		Common Sense	
Courage		Maturity	
Caution		Charm/Sense of Humour	
ADMINISTRATION	**GRADE**	Honesty	
Management Ability			
Staff Work			
Use of English		**Total:** (Average 40)	

OVERALL PERFORMANCE

A - Exceptional   (Very promising career prospects)

B - Good   (Promising career prospects)

C - Average   (Likely to be reliable)

D – Just Pass   (Some shortcomings)

E - Fail   (Unable to fulfil SMCC requirements)

F – Severe Fail   (Unlikely to succeed in General Service)

Appendix J: Results Tables Not Included in the Chapter 5

Section 1. Difference in means between Pass and Fail groups and the various OCEAN dimensions.

	Pass	N	Mean	Standard Deviation
O	1.00	13	129.8308	22.30110
	2.00	39	122.4872	20.39075
C	1.00	13	177.4538	21.22721
	2.00	39	183.3333	17.46927
E	1.00	13	184.2308	21.77596
	2.00	39	180.3590	22.44465
A	1.00	13	193.4538	22.84683
	2.00	39	199.2308	23.95880
N	1.00	13	122.1923	22.79395
	2.00	39	125.1795	25.53725
OT	1.00	13	51.0769	8.17987
	2.00	39	47.7949	9.38975
CT	1.00	13	101.8077	13.41258
	2.00	39	106.2051	10.95488
ET	1.00	13	82.5385	12.62679
	2.00	39	81.0256	12.18819
AT	1.00	13	103.8077	13.10779
	2.00	39	107.3077	13.29059
NT	1.00	13	43.3462	7.78723
	2.00	39	42.3077	9.88660
OSD	1.00	13	78.7538	16.49003
	2.00	39	74.6923	13.52805
CSD	1.00	13	75.6462	8.85782
	2.00	39	77.1282	7.80748
ESD	1.00	13	101.6923	9.67285
	2.00	39	99.3333	11.20698
ASD	1.00	13	89.6462	10.52747
	2.00	39	91.9231	11.85301
NSD	1.00	13	78.8462	16.77596
	2.00	39	82.8718	17.54720

	O	C	E	A	N
Mann-Whitney U	192.000	218.500	228.500	211.500	245.500
Asymp. Sig. (2-tailed)	.194	.459	.597	.375	.866

	OT	CT	ET	AT	NT
Mann-Whitney U	205.000	209.000	242.500	198.500	241.000
Asymp. Sig. (2-tailed)	.305	.347	.816	.245	.791

	OSD	CSD	ESD	ASD	NSD
Mann-Whitney U	204.000	226.000	219.500	229.000	223.500
Asymp. Sig. (2-tailed)	.295	.560	.472	.604	.526

Section 2. Pass and Fail groups and the various OCEAN subcomposites.

	Pass	N	Mean	Std. Deviation
OsubP	1.00	13	51.5385	10.41326
	2.00	39	48.1538	10.05129
OsubSI	1.00	13	14.6308	5.97263
	2.00	39	12.4615	3.96595
OsubCr	1.00	13	23.4615	2.93301
	2.00	39	21.3846	4.04329
OsubR	1.00	13	29.7231	6.87000
	2.00	39	30.4615	5.95530
OsubCu	1.00	13	10.4769	3.13957
	2.00	39	10.0256	3.27259
CsubED	1.00	13	88.2385	9.42103
	2.00	39	90.0513	7.52522
CsubHW	1.00	13	31.8846	4.44482
	2.00	39	32.0000	3.89331
CsubO	1.00	13	57.3308	11.12230
	2.00	39	61.2821	8.52226
EsubSB	1.00	13	71.3462	10.56209
	2.00	39	68.2564	11.38956
EsubT	1.00	13	31.8846	6.40362
	2.00	39	31.6410	4.90696
EsubSA	1.00	13	25.6538	3.09155
	2.00	39	25.2308	3.21560
EsubA	1.00	13	19.2692	2.14685
	2.00	39	18.8205	2.64448
EsubS	1.00	13	36.0769	5.05736
	2.00	39	36.4103	4.20349
AsubWS	1.00	13	62.8462	7.57018
	2.00	39	62.9231	8.14792
AsubF	1.00	13	32.4000	6.08660
	2.00	39	35.1795	4.48314
AsubC	1.00	13	35.5385	4.19554
	2.00	39	35.8718	4.93206
AsubCI	1.00	13	36.0462	5.58594
	2.00	39	38.4872	5.97306
AsubH	1.00	13	26.6231	3.08711
	2.00	39	26.7692	3.20741
NsubNSO	1.00	13	32.0962	9.11246
	2.00	39	32.3590	9.16942
NsubW	1.00	13	31.5962	7.35239
	2.00	39	33.6154	8.05120
NsubI	1.00	13	47.4423	10.39338
	2.00	39	47.6154	10.30487

	OsubP	OsubSI	OsubCr	OsubR	OsubCu
Mann-Whitney U	201.000	193.500	177.000	240.500	238.000
Asymp. Sig. (2-tailed)	.267	.204	.104	.783	.742

	CsubED	CsubHW	CsubO
Mann-Whitney U	219.500	243.500	210.500
Asymp. Sig. (2-tailed)	.472	.832	.363

	EsubSB	EsubT	EsubSA	EsubA	EsubS
Mann-Whitney U	207.500	239.500	226.500	229.000	245.500
Asymp. Sig. (2-tailed)	.331	.767	.566	.601	.865

	AsubWS	AsubF	AsubC	AsubCI	AsubH
Mann-Whitney U	237.500	200.500	224.500	183.500	248.500
Asymp. Sig. (2-tailed)	.735	.262	.537	.139	.915

	NsubNSO	NsubW	NsubI	NsubEJ
Mann-Whitney U	253.000	223.500	241.500	233.500
Asymp. Sig. (2-tailed)	.992	.526	.800	.671

Section 3. Correlation of the OCEAN subcomposites and the Course Grade

Course Grade		
OsubP	Correlation Coefficient	-.090
	Sig. (2-tailed)	.528
	N	52
OsubSI	Correlation Coefficient	-.166
	Sig. (2-tailed)	.239
	N	52
OsubCr	Correlation Coefficient	-.196
	Sig. (2-tailed)	.163
	N	52
OsubR	Correlation Coefficient	.102
	Sig. (2-tailed)	.471
	N	52
OsubCu	Correlation Coefficient	.151
	Sig. (2-tailed)	.285
	N	52

Course Grade		
CsubED	Correlation Coefficient	.069
	Sig. (2-tailed)	.626
	N	52
CsubHW	Correlation Coefficient	.124
	Sig. (2-tailed)	.381
	N	52
CsubO	Correlation Coefficient	.165
	Sig. (2-tailed)	.238
	N	52

Course Grade		
EsubSB	Correlation Coefficient	.116
	Sig. (2-tailed)	.411
	N	52
EsubT	Correlation Coefficient	.131
	Sig. (2-tailed)	.356
	N	52
EsubSA	Correlation Coefficient	.062
	Sig. (2-tailed)	.662
	N	52
EsubA	Correlation Coefficient	.156
	Sig. (2-tailed)	.271
	N	52
EsubS	Correlation Coefficient	.124
	Sig. (2-tailed)	.381
	N	52

Course Grade		
AsubWS	Correlation Coefficient	-.035
	Sig. (2-tailed)	.805
	N	52
AsubF	Correlation Coefficient	.063
	Sig. (2-tailed)	.655
	N	52
AsubC	Correlation Coefficient	.017
	Sig. (2-tailed)	.905
	N	52
AsubCI	Correlation Coefficient	.149
	Sig. (2-tailed)	.293
	N	52
AsubH	Correlation Coefficient	-.078
	Sig. (2-tailed)	.581
	N	52

Course Grade		
NsubNSO	Correlation Coefficient	-.006
	Sig. (2-tailed)	.965
	N	52
NsubW	Correlation Coefficient	.117
	Sig. (2-tailed)	.409
	N	52
NsubI	Correlation Coefficient	.045
	Sig. (2-tailed)	.749
	N	52
NsubEJ	Correlation Coefficient	.253
	Sig. (2-tailed)	.070
	N	52

Section 4. PAP

	Pass	N	Mean	Std. Deviation
PAP	1.00	7	141.1429	14.62223
	2.00	20	136.7500	17.06913

PAP	
Mann-Whitney U	58.500
Asymp. Sig. (2-tailed)	.524

Course Grade		
PAP	Correlation Coefficient	-.084
	Sig. (1-tailed)	.338
	N	27

PAP		
Total	Correlation Coefficient	-.089
	Sig. (1-tailed)	.327
	N	28
TotTactical	Correlation Coefficient	-.206
	Sig. (1-tailed)	.146
	N	28
TotAdmin	Correlation Coefficient	.299
	Sig. (1-tailed)	.061
	N	28
TotPers	Correlation Coefficient	-.034
	Sig. (1-tailed)	.431
	N	28

Section 5: The Four Board Dimensions

	Pass	N	Mean	Standard Deviation
Effective Intellect	1.00	8	5.2500	.46291
	2.00	21	4.8214	.81832
Leadership Potential	1.00	8	4.6250	.74402
	2.00	21	4.9048	.62488
Character and Personality	1.00	8	5.2500	.46291
	2.00	21	4.9881	.70921
Service Motivation	1.00	8	5.5000	.53452
	2.00	21	5.2262	.88000

	EI	*LP*	*CandP*	*SM*
Mann-Whitney U	55.000	64.000	65.000	68.000
Asymp. Sig. (2-tailed)	.122	.275	.304	.407

Course Grade		
EI	Correlation Coefficient	-.222
	Sig. (1-tailed)	.124
	N	29
LP	Correlation Coefficient	.231
	Sig. (1-tailed)	.114
	N	29
CandP	Correlation Coefficient	-.110
	Sig. (1-tailed)	.285
	N	29
SM	Correlation Coefficient	-.288
	Sig. (1-tailed)	.065
	N	29

		Total	Total Tactical	Total Admin	Total Personality
EI	Correlation Coefficient	-.102	-.083	.223	-.139
	Sig. (1-tailed)	.295	.332	.118	.233
	N	30	30	30	30
LP	Correlation Coefficient	.246	.236	-.064	.214
	Sig. (1-tailed)	.095	.105	.368	.128
	N	30	30	30	30
CandP	Correlation Coefficient	.027	-.004	-.090	.026
	Sig. (1-tailed)	.444	.492	.317	.445
	N	30	30	30	30
SM	Correlation Coefficient	-.093	-.113	-.225	-.022
	Sig. (1-tailed)	.312	.275	.116	.454
	N	30	30	30	30