UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

DETECTING SOCIALLY DESIRABLE RESPONSES IN PERSONALITY INVENTORY

by

PRIYALATHA GOVINDASAMY

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Arts

July 2011

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

TABLE OF CONTENTS

Acknowledgement ii
Table of Contents iii
List of Tables vi
List of Figures viii
Abstrak ix
Abstract xi

1. CHAPTER 1 INTRODUCTION

1.0 Introduction 1
1.1 Background of study 1
1.2 Problem statement 5
1.3 Purpose of the study 8
1.4 Research objectives 9
1.5 Research questions 10
1.6 Significance of the research 11
1.7 Limitations of the study 12
1.8 Definition of personality dimension 13

2. CHAPTER 2 LITERATURE REVIEW

2.0 Introduction 15
2.1Personality Assessment 15
2.2 Application of Personality Assessment 17
2.3 Definition of Social Desirability 20
2.4 Social Desirability in Personality Assessment 21
2.4.1 Person as contributing factor to social desirability in personality assessment 22
2.4.2 Personality test format as contributing factor to social desirability in personality assessment 24
2.5 Methods of Reducing Socially Desirable Responding 27
2.5.1 Standard Guidelines in Minimizing Socially Desirable Responding 27
2.5.2 Forced Choice Format 29
2.5.3 Randomized Response 29
2.5.4 Equal Positive and Negative Items 30
2.6 Methods used for detecting socially desirable responding 30
2.6.1 Social Desirability Scales 31
2.6.2 Detecting social desirability from item responses 34
2.7 Theoretical Framework 40
2.7.1 Conceptual Framework 43

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS NIVERSITI P

3. CHAPTER 3 METHODOLOGY

3.0 Introduction 44
3.1 Research Design 44
3.2 Instrument
3.2.1 Selection of Instrument 45
3.2.2 Description of Instrument 46
3.2.3 Factorial Structure of the Instrument 49
3.4 Pilot study 55
3.5 Sampling 56
3.6 Investigation and procedure 59
3.7 Data analysis 60

4. CHAPTER 4 FINDINGS

4.0 Introduction 66
4.1 Comparing mean score of honest and socially desirable groups
4.1.1 Comparing the mean score deviations and differences between the honest and socially desirable groups 67
4.1.2. Distribution of non-fitting response of the Rasch model 70
4.2 Identifying personality dimension that are susceptible to social desirability
4.2.1 Factor Analysis 71
4.2.2 Test information function 76
4.2.2 (a) Test information function for the five dimensions in IPIP 77
4.3 Identifying items that are prone to socially desirable responding 80
4.3.1 Differential Item Functioning with Mantel-Haenszel Chi-square 81
4.3.2 Differential Item Functioning using Cumulative Log Odd Ratio 82
4.3.3 Differential Test Functioning 84
4.3.4 Differential Step Functioning 85

5. CHAPTER 5 CONCLUSION, DISCUSSION, IMPLICATIONS AND RECOMMENDATIONS

5.0 Introduction 89
5.1 Conclusion 89
5.1.1 Research question 1 89
5.1.2 Research question 2 90
5.1.3 Research question 3 91
5.2 Discussion 93
5.3 Implications 98
5.4 Recommendations 99

6. REFERENCES 101

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

7. APPENDICES

Appendix A Permission letter of IPIP: Malay version
Appendix B Permission letter: MRSM
Appendix C List of modified Malay IPIP items
Appendix D Questionnaire: Cover page of honest group
Appendix D (i) Questionnaire: Socially desirable group
Appendix EDIF output: Mantel-Haenszel approach
Appendix E (i) DIF output: Common log odd ratio
Appendix F List of item difficulties

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

LIST OF TABLES

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

Table 3.1 Classification of IPIP items accordingly to the dimension and types of items 48

Table 3.2Total Variance Explained 50

Table 3.3Table of Rotated Component Matrix 53

Table 3.4 Distribution of demographic factors 58

Table 3.5 Differential Step Functioning Form 65

Table 4.1 Mean and standard deviation for honest, socially desirable and combined groups 67

Table 4.2 Mean, standard deviation of five dimensions of IPIP inventory 68

Table 4.3Score distribution for honest and socially desirable groups 69

Table 4.4 Summary of non-fitting responses according to the groups 71

Table 4.5 Total variance explained for honest and socially desirable groups 72

Table 4.6Comparing item loading for honest and socially desirable group 74

Table 4.7 Differential Item Functioning (DIF) with Mantel-Haenszel Chi-square 81

Table 4.8 Differential Item Functioning from cumulative Log Odd Ratio 83

Table 4.9 Item difficulty parameter between honest and socially desirable groups 83

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

Table 4.10 Results of Differential Test Functioning (DTF) for overall and five dimensions of IPIP 84

Table 4.11Differential Step Functioning Results for DIF items 86

Table 4.12 Summary of DIF items with DSF magnitude and interpretations 87

Table 4.13 Frequency of response categories between honest and socially groups 88

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

LIST OF FIGURES

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Figure 2.1 Conceptual Framework 43

Figure 3.1 Scree plot to show the number of factors 51

Figure 4.2 Test information function of Openness for honest and socially desirable groups 76

Figure 4.3 Test information function of Conscientiousness for honest and socially desirable groups 77

Figure 4.4 Test information function of Extraversion for honest and socially desirable groups 77

Figure 4.5 Test information function of Agreeableness for honest and socially desirable groups 78

Figure 4.6 Test information function of Neuroticism for honest and socially desirable groups 78

Figure 4.6 Test information function of Neuroticism for honest and socially desirable groups 79

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

MENGESAN RESPONS IDAMAN SECARA SOSIAL DALAM INVENTORI PERSONALITI

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Abstrak

Kajian ini dijalankan untuk mengesan respon idaman sosial daripada respon-respon yang berbeza. Selain itu, kajian ini juga memeriksa item dan dimensi-dimensi personaliti yang mudah terjejas kepada kebolehinginan sosial. Dalam kajian eksperimental ini, 521 pelajar sekolah diuji dengan borang soal-selidik *International Personality Item Pool (IPIP*) sebanyak dua kali di bawah arahan menjawab secara jujur dan dengan kebolehinginan sosial. Respon-respon dari pentadbiran pertama diklasifikasikan sebagai kumpulan jujur dan respon-respon dari pentadbiran kedua dikategorikan sebagai kumpulan kebolehinginan sosial. Kepincangan skor respon dari min dan analisis keserasian digunakan untuk mengesan respon idaman sosial. Keberbezaan fungsi ujian diaplikasikan dalam mengenal pasti item-item yang cenderung untuk kebolehinginan sosial. Analisa diteruskan dengan keberbezaan fungsi ujian untuk memeriksa kesan keberbezaan fungsi item pada keseluruhan ujian dan pada dimensi-dimensi personaliti yang ditaksir. Keberbezaan fungsi langkah diaplikasikan untuk mengenal pasti kategori respon yang menyumbang kepada keberbezaan fungsi item. Kajian ini juga menjalankan analisa faktor dan lengkuk fungsi informasi ujian untuk memeriksa dimensi-dimensi personaliti yang rentan untuk kebolehinginan sosial. Kumpulan kebolehinginan sosial melaporkan taburan skor yang tinggi pada dua sisihan piawai di atas min. Bersamaan dengan itu, kumpulan kebolehinginan sosial menggambarkan peratusan respon yang tidak padan dengan modal Rasch pada nilai 2.0 logits lebih tinggi

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

berbanding kumpulan jujur. Perbandingan struktur-struktur faktor melaporkan perbezaan di antara kumpulan jujur dan kebolehinginan sosial. Lengkuk fungsi informasi ujian bagi dimensi keterbukaan menggambarkan perbezaan maklumat di antara kumpulan jujur dan kebolehinginan sosial. Enam item dikesan untuk keberbezaan fungsi item dan tiga daripada item tersebut mewakili dimensi keterbukaan. Kehadiran keberbezaan fungsi item meyebabkan perubahan dalam ujian personaliti dan dalam dimensi keterbukaan. Analisis keberbezaan fungsi langkah melaporkan kategori-kategori respon berfungsi secara berbeza mengikut item. Kajian menyimpulkan bahawa, sisihan skor yang besar dan respon yang tidak padan dalam inventori personaliti sebagai petunjuk kepada respon kebolehinginan sosial. Sementara itu, parameter kesukaran item yang rendah dan pemilihan kategori respon menggambarkan tanda-tanda kebolehinginan sosial. Lantarannya, proses penyaringan individu yang berpotensi menjawab dengan kebolehinginan sosial serta penyingkiran item-item yang cenderung ke arah kebolehinginan sosial mampu mengatasi masalah kebolehinginan sosial dalam ujian yang berkepentingan tinggi.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

DETECTING SOCIALLY DESIRABLE RESPONSES IN PERSONALITY INVENTORY

Abstract

This study attempted to detect the socially desirable responses within differential responses. Besides, this study also examined items and personality dimensions that are vulnerable to social desirability. In the experimental design, a sample of 521 students was tested twice with the International Personality Item Pool (IPIP) under honest and socially desirable instructions. Responses from the first administration were classified as honest group responses and those from the second administration were grouped as socially desirable responses. The mean dispersion and fit analysis were applied in detecting socially desirable responses. Differential Item Functioning (DIF) was used in identifying items that are prone to social desirability followed by Differential Test Functioning (DTF) which examines the DIF effects to the test and to each personality dimension assessed. Differential Step Functioning (DSF) was applied to determine the contributing steps in polytomous responses to DIF. Factor analysis and item information function curves were used to examine the personality dimensions that are susceptible to social desirability. The socially desirable group reported a distinctively higher score distribution at two standard deviations above the mean. Correspondingly, the socially desirable group has a higher percentage of non-fitting responses with values more than 2.0 logits. Differences of factor structures of IPIP dimensions were found between the honest and socially desirable response groups. Test information function for the Openness dimension illustrated a differential information function between the honest and socially desirable groups. Six items were flagged for Differential Item Functioning

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

(DIF) and three of the items represented the Openness dimension. In addition, the presence of DIF caused changes in the test and in the Openness dimension for the socially desirable groups. The DSF analysis reported response categories function differently according to the items. The study concluded that high score and non-fitting responses in a personality inventory are indications of socially desirable responding. Meanwhile, low difficulty item parameter and endorsement of response categories are also signs of social desirability. Therefore, screening of the potential socially desirable responding individual and the elimination of items that are prone to social desirability would help to arrest the problem of social desirability in high-stake testing.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

CHAPTER ONE

INTRODUCTION

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

1.0 Introduction

Chapter one focuses on the background of the study and problem statements which leads to the purpose of this study. Based on the purpose of the study, the research objectives and questions were stipulated. This chapter also reports the significance and limitations of this study. The chapter concludes with the definition of the five personality dimensions.

1.1 Background of the study

Testing and evaluation started as early as 2200 B.C in China in the form of civil service examinations (Aiken & Marnat, 2006). The Chinese civil service examination was used to determine whether government officials were capable in performing their tasks. This Chinese civil examination leads to the development of civil examinations in countries like Britain, France and Germany. Later in the 19th century, psychiatrists and psychologists of mental disorders developed clinical assessment techniques and tests to assess their patients (Aiken & Marnat, 2006). The great impact in the test development was when Alfered Binet constructed the first mental test called the Binet Intelligence Test. The Binet intelligence test was used to predict the scholastic achievement of an individual. The success of measuring intelligence was then applied to the Military through the Army alpha and beta tests. These military tests were constructed to measure the cognitive ability of their

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

recruitment candidates (Boyle, Matthews & Saklofske, 2008). Besides cognitive ability, Military test also measures the candidate's personality during the recruitment selection (Boyle et al., 2008). According to Boyle et al. (2008), the psychological testing instruments used in military recruitment gave an insight into the suitability of a recruit together with the cognitive abilities of the recruit and thus facilitated the decision making process of selecting a recruit into the Military. Among the early personality inventories were the Personal Data Sheet and Thurstone Personality Schedule.

Over the years, various personality tests were developed and among them are the Minnesota Multiphasic Personality Inventory (MMPI), California Psychological Inventory (CPI), Sixteen Personality Factor Questionnaire (16 PF) and the Revised NEO personality inventory (NEO-PI-R). Using personality to explain how human behavioural traits react within their working environment results in its importance in predicting about an individual (Detrick, Chibnall & Luebbert, 2004). It gives an idea of how individuals interact with others and their reaction towards things and ideas. The predictive nature of personality testing resulted in its wide application in the industrial, educational and in military context (Horst, 1968). Inference about an individual from the personality inventory provides extra information which would be utilized in the selection of appropriate candidates (Carrigan, 2007). Hence, the personality inventory is relied upon when screening for job applicants in employment selection (Li & Bagger, 2007).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

The study by Detrick et al. (2004) proved that NEO-PI-R has predictive validity with respect to police academic performance. The predictive capacity of personality test emerges as an aid in decision making and has evolved in the Malaysian context. The significance of personality is noted as a component in employment and the educational setting. This is seen in the application of personality elements in the University Science Malaysia (USM) entrance examinations known as the Malaysian University Selection Inventory (MUnSyI) and in the entrance examinations for teacher trainees known as the Malaysian Educators Selection Inventory (MEdSI). In addition, personality elements are included in the recruitment of graduate employees into the Malaysian civil service. Even though the personality test were used under different setting, the purpose was the same; to predict an individual's attributes to aid in decision-making.

However, the use of personality tests was questioned by the Civil Rights Act of 1964 and Equal Employment Opportunity Commission (EEOC) of the United States over the validity and discriminatory issues in pre-employment (Carrigan, 2007). Doubts arose as many personal disclosures were used in the selection process and whether it contributed to any bias issues. To arrest the problem, American Psychological Association (APA), American Counselling Association (ACA), National Board of Certified Counsellor (NBCC), Standards for Educational and Psychological Testing and National Council on Measurement in Education govern and draw up the guidelines to help professionals administer the tests ethically (Drunmond & Jones, 2006). Drunmond and Jones (2006) added that the guidelines were drawn due to the impact of the personality test on society and decision makings which saw a need for a standard setting procedure in testing and assessment. Ethical measures were

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

considered in order to minimize any bias decision from the outcome of the personality testing.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

In psychological testing, the quality of the test and ethical conduct of helping professionals were strictly monitored to reduce the biasness in decisions made. However, the test-taker's attitude and honesty are not being controlled. Studies have shown that, people like to demonstrate good impression in a psychological testing (Dunn, 2009). In cases where the test scores become a determinant for a person's future the individual has the tendency to manipulate and answer dishonestly on the test (Aiken & Marnat, 2006). Fox and Meijer (2008) also pointed out that people often respond untruthfully on personal or sensitive questions in psychological or educational assessments. The problem of untruthful responding in personality testing opens to debate over its use for the selection process (Dilchert, Ones, Viswesvaran, & Deller, 2006). Since personality items have no absolute correct answers, it makes the items easily faked (Horst, 1968). Meantime, the purpose of the test encourages a person's tendency to fake the response (Dunn, 2009). The pressure and personal need of an individual to pursue a better life would motivate the test takers to fake a test. Irrespective of the reasons, all of these would affect the validity and utility of the personality inventory (Charles, 2003).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

1.2 Problem statement

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Personality assessment was used to obtain data about people to reach a conclusion on clinical, legal, educational and guidance, educational and vocational selection and also for research purposes (Holt, 1971). Application of the personality instrument in the working context predicts one's personal attributes which then evaluates over the suitability of the candidate for the specific job (Fitting the bill, 2009). This is because, employee's personality traits are found to impact on their behaviour and performance at their workplace (Nek Kamal bin Yeop Yunus, 1997). The ability of the personality test to predict an applicant's capability and suitability to the workplace has resulted in taking into consideration the results of the tests for the selection procedures of employees (Kumaresan, Aizat Mohd Nasurdin & Ramayah, 2005).

The application of personality in the job screening process is found in the Public Service Department examination while recruiting a grade 41 category officer into the civil service (*Suruhanjaya Perkhidmatan Awam*, 2005). The personality of a job applicant is appraised to determine the suitability of the candidate concurrently with the position applied by the candidate into civil service (*Suruhanjaya Perkhidmatan Awam*, 2005). Meanwhile, personal attributes of students are assessed to match the corresponding choice of courses applied for in the university. The university selection tests that include personality elements are used to determine the suitability of students in their selection of course in the university (*Ujian Penetapan Khas*, 2008). University Science Malaysia (USM) conducts a test called the Malaysian University Selection Inventory (MUnSyI) to facilitate the placement of students into different courses of study in the university (*Ujian Penetapan Khas*, 2008). Besides

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

MUnSyI, the Malaysian Educators Selection Inventory (MEdSI) is designed spesifically for filtering applicants for the teacher education training program in Malaysia (MEdSI, 2009). Applicants of teacher education programs are required to pass the MEdSI and subsequently called to attend an interview before they are chosen (Arifin Bin Hj Zainal, Asmawati Binti Desa, Hazalizah Binti Hamzah & Nachiappan, 2009). Moreover, the *Maktab Rendah Sains MARA* (MRSM) also sets *Ujian Kecenderungan Kemasukan Maktab* (UKKM) test as pre-requisites for students to be selected into MRSM (*Syarat Kemasukan Ke MRSM*, 2010).

The MUnSyI, MEdSI and UKKM tests are self-reported assessment instruments with the personality element included as a common domain to be assessed. In addition, a reality show called Nescafe Kisck-Start in Malaysia used the LEONARD Personality inventory to help provide additional information when selecting the 16 semi-finalist candidates (Emotional signs, 2004). This indicates the personality features can be used to predict the likelihood of a candidate and provide useful information to be used in the selection process (Fitting the bill, 2009). The predictive feature of personality assessment leads to its applications in various decision making contexts (Kline, 1976).

The findings of personality assessments in school entrance exams, university placements, job screening and competitions showed its application in high-stake testing. All these indicate the increasing reliance of personality assessments both in school in school and workplace. Therefore, similar to graduates looking for employment, students too face with high-stake tests such as UKKM, MUnSyI and MEdSI that determine their future undertakings in life. In all these tests, personality

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

assessment was included to gather information about them for decision making. However, the use of self-reported personality tests in competitive environments is susceptible to biased responding (Hirsh & Peterson, 2008). Moreover, when selections are meant for the best, it could place pressure on the test-taker (Nunnally, 1975). The intention to portray the best of themselves in an impressive way often leads to the manipulation of responses in a test (Lanyon & Goodstein, 1982). The motivation to distort is caused by the intended goal in a person and the implication of the implication of the test to them (Detrick & Chibnall, 2008). It is human nature for individual to portray the best impression of themselves to meet their desired goals (Iddekinge, McFarland & Raymark, 2007). Besides, an individual has the potential towards approval-seeking attitude (Leite & Beretvas, 2005). Individual's approval-seeking attitude is to meet their desired goal and to portray the required characteristics that match the purpose of the test. The approval seeking is manifested by responding to the statements in the personality tests in a socially desirable manner which becomes the concern when the results are used for making important decisions (Arthur, Woehr & Graziano, 2001).

In addition, research showed that self-reported measures often inflate an individual's score compared to rating by others (Yang, Bagby & Ryder, 2000). This is because individuals have the tendency to fake personality tests to avoid any personal disclosures (Dunn, 2009). Therefore, obtaining valid and reliable information about a person depends greatly on the cooperation from the individual (Fox & Meijer, 2008). This is because the test-taking attitude of the job applicants and students affects the actual test performance and influences the validity of selection (McCarthy & Goffin, 2003). Individuals responding in a socially desirable

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

manner try to present a good image rather than projecting their actual self (Horst, 1968). The socially desirable responses become a great concern to many as it reduces the validity of the personality measured (Crowne & Marlowe, 1960). This raises the question on the accuracy of prediction and becomes a threat to the validity of the personality inventory (Kroner, Mills, Yessine & Hemmati, 2004). Therefore, information obtained on individuals become unreliable and would cause a deserving individual to lose his or her opportunity in employment and education to an unqualified candidate.

1.3 Purpose of the study

Kline (1976) stated that the predictive ability of personality assessments leads to its application in various decision-making contexts. Yet, the accuracy of the personality inventory is doubtful and opened to criticism. Ellingson, Sackett and Hough (1999) mentioned that in a standard selection process, it is difficult to assess an applicant's true scores. In such circumstances, selection is done based on the observed scores which could be the true or faked responses. As stated earlier, an individual's interest in the test and his willingness determine the response patterns in the personality inventory. Therefore, it is essential to examine the score outputs as well as the scoring pattern to determine any aberrant scores. The analyses of score outputs could trace individuals who tend to respond in a socially desirable manner and determine the respondents who tend to fake or respond in a socially desirable way. Meanwhile, score output analyses would also help in extracting items that are prone to faking. The detection of deviant response patterns and items with faking possibilities could be re-evaluated prior to the selection process. Moreover, detection

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

of such responses would help to identify only the appropriate candidates in the selection process and reduce wrong decisions. Therefore, this study intended to examine the score dispersion of honest and socially desirable groups from the mean and person fit analysis in detecting the socially desirable responses. Differential item functioning (DIF), item difficulty parameter and differential step functioning (DSF) were used in identifying items that are prone to social desirability. Meanwhile, personality dimensions that are prone to social desirability were examined through factor analysis, test information function and differential test functioning (DTF). This would help to identify honest responses and ensure that right decisions are made about individuals to minimize errors of selecting individuals who provide socially desirable responses.

1.4 Research Objectives

In personality tests, the rightness of answers differs from one to another (Horst, 1968). This is because, individuals are different and act contradictory at times (Alex, 1975). Besides that the personality test's biased responses of individuals are caused by their intention to gain social approvals (Stocke & Hunkler, 2007). Thus, the social gaining approvals lead to difficulties in assessing an individual's personal attributes rather than thinking on how they should respond (Horst, 1968). In addition to that, personality tests with no correct answers aids in responding items in a socially desirable manner (Horst, 1968). Meantime, the transparency of the item meanings helps to clearly discriminate items which require socially desirable responses (Detrick & Chibnall, 2008). Therefore, it becomes essential to determine whether personality measures functions equivalently across different group of applicants

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

(Mitchelson, Wicher, LeBreton & Craig, 2009). Thus, this study was initiated to detect individuals who distort their responses in a socially desirable manner by comparing the different response patterns. This study also focused on identifying items that are prone to socially desirable responses. Finally, this study also attempted to identify dimensions in the personality inventory that are prone to socially desirable responses. Based on the above objectives, these following research questions were proposed.

1.5 Research Questions

1. To what extent do the responses of the group that distorted responses in a socially desirable manner differ from that of the honest group?
2. What are the characteristics of the items that are prone to socially desirable distortions?
3. What are the personality dimensions that are vulnerable to socially desirable responding?

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

1.6 Significance of the research

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The application of personality measures is gaining importance in many aspects of our life (Aiken & Marnat, 2006). In tandem with this rapid growth, there are upcoming studies underlying socially desirable response effects on personality tests. Therefore, this study is an initiative to venture and explore the response patterns of individuals to detect those distorted in a socially desirable manner.

The findings from this study would help test-users to be aware of the issue of social desirability and take measures to prevent socially desirable responses. It also notifies the test-users not to rely solely on personality inventory when making inference about an individual. This is due to the vulnerability of the personality inventories.

Study also hopes to open the minds of test-takers to not accept the results of personality inventory per se as it does not measure the overall characteristics of a person. The use of statistical approaches in detecting socially desirable responses would provide evidence that even in personality testing faking can be detected. Therefore, it is hope that the detection of socially desirable responses will help to motivate individuals to provide honest responses in future. In addition, detection of socially desirable responding would help to differentiate between honest and dishonest respondents and ensure a fair chance for all in any selection process.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

1.7 Limitations of the study

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The accuracy of detecting socially desirable responses relies on the cooperation of the respondents. The pilot study conducted revealed that even individuals asked to respond in a socially desirable manner did not respond as instructed. This indicates that respondent participation and willingness to respond is the major limitation of this study. This is because it is difficult to monitor the respondents and direct them to respond in the way needed for the research.

Study also fails to provide an exact scenario that underlines the importance of the necessity for participants to portray their best characteristics. The instructions to respond in a socially desirable manner and scenarios presented when administrating the test may not be sufficiently important to the participants. Therefore, failure in setting the appropriate environment at the experimental design would be another limitation of this study.

The choice of instrument is identified as another limitation of this study. It was difficult to obtain established instruments as they were expensive while some authors were reluctant to grant permission on the use of their inventory. Meanwhile, efforts to obtain locally developed personality instruments were also not successful. Also due to the small sample size, the data analysis using the Item Response Theory (IRT) model is limited to the 1-parameter logistic model which is run with the WINSTEP computer program that is available in the university.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

1.8 Definition of Personality Dimensions

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

In this study, the International Personality Item Pool (IPIP) inventory was used in studying socially desirable responding. IPIP was developed based on the Big five personality model that comprised of five dimensions. The following are the five dimensions and their definitions adapted from Pervin and John (2001, p. 257).

Openness

Assesses proactive seeking and appreciation of experience for its own sake, toleration for and exploration of the unfamiliar.

Conscientiousness

Assesses the individual's degree of organization, persistence and motivation in goal-directed behaviour.Contrasts dependable, fastidious people with those who are lackadaisical and sloppy.

Extraversion

Assesses quantity and intensity of interpersonal interaction, activity level, need for stimulation and capacity for job.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Agreeableness

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Assesses the quality of one's interpersonal orientation along the continuum from compassion to antagonism in thoughts, feelings and actions.

Neuroticism

Assesses the adjustments versus emotional stability. Identifies individual prone to psychological distress, unrealistic ideas, excessive cravings or urges, and maladaptive coping responses.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

CHAPTER TWO

LITERATURE REVIEW

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

2.0 Introduction

Chapter two is a collection of critical literature reviews of social desirability and personality. This chapter begins with the definition of personality assessment and its applications in real life settings. It is followed by the definition of social desirability and its presence in personality assessments. The literature review next reviews reported methods of reducing and detecting social desirability responding. Finally, the chapter concludes with the theoretical and conceptual framework of this study.

2.1 Personality Assessment

Personality is defined as an individual's unique pattern of traits (Guilford, 1954). Alex (1975) stated that personality only refers to those traits that are assumed to be fixed in an individual. The traits convey consistent and continuing information about a person (Carver & Scheier, 2000). This is why personality is referred to as clusters of related dimensions that allow describing a person's behavior, feeling and interaction with others (Aiken & Marnat, 2006). The means of gathering and organizing information about a person with the intention of wanting to learn about a person is called personality assessment (Lanyon & Goodstein, 1982). Personality assessment comprises a series of items that describe an individual's personality (Lonnqvist, 2008). Therefore, statements in the personality assessment would be anything that can be said about a person (Horst, 1968). Meanwhile, personality

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

assessment presumes that there are characteristics that illustrate the differences between people and the ways to measure it precisely (Aiken, 1999).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Personality assessment takes various forms in gathering information about a person's personality. Among them are through external impressions (observation) and through self-reported measures (Carver &Scheier, 2000). Kline (1976) mentioned that personality attributes can be measured through a questionnaire, projective and objective techniques. However, most personality assessments are in the form of self-reported inventories because the items in the inventory require individuals to describe their behavior, attitude or feelings in general (Robie, Schmit, Ryan&Zickar, 2000). Due to the descriptive nature of personality assessments, it was used extensively in clinical, counseling, business, industrial, governmental, military, educational and school contexts (Aiken, 1999).The assessment and testing of personality is to explore an individual's personal attributes under various settings for the purpose of reaching a conclusion (Holt, 1971). This is because personality variables were found to contribute and correlate to an individual's life outcome events (Magnus, Viswesvaran, Deshpande & Joseph, 2006).

The first personality assessment was the Hoodsworth's Personal Data Sheet which was applied in the U.S military context (Aiken, 1999). This was then followed by assessments like Minnesota Multiphasic Personality Inventory (MMPI) applied in a clinical setting and Hogan Personality Inventory in the employment selections (Aiken, 1999). Meanwhile, the strong scientific background of the 16 Personality Factor Questionnaire (16 PF) brought its application into the industrial, organizational, counseling and clinical, research, educational and medical settings

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

(Cattell& Mead, 2008). On the other hand, the California Psychological Inventory (CPI) which comprises common descriptions of personality is applied in job selection procedures as the items are skill oriented and lean towards interpersonal attributes (Boer, Starkey & Hodgetts, 2008).

Despite the predictive ability of the personality inventory, it can never be adequate enough to explore an individual's personal attributes (Horst, 1968). This is because humans are complex in nature and personality inventories are merely tools in assisting the gathering of information about a person and does not measure completely a person's personal attributes in total. Incompleteness of measuring a person as a whole could also explain why individuals themselves are unaware of their attributes.

2.2 Application of Personality Assessment

Personality measures were used in personnel selections, training processes and in personal development to mould competent employees (Salgado, 2005). The personality questionnaires were used to identify individuals with traits reflecting occupational success (Kirkcaldy, 2001). Therefore, the application of personality measurement convinces that appropriate candidates were selected at the selection process (Stabile, 1997). The usefulness of personality assessment has gained its popularity and is expected to contribute to a 20% increase of online testing in future (Piotrowski & Armstrong, 2006).

Findings have reported the increasing reliance of personality testing in the uniformed forces can be seen. The necessity of appraising personality attributes was

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

found in the United States Army's Special Operations Target Interdiction Course (Girard & Scholtz, 2005) whereby the personal characteristics of the army officers were assessed to determine their suitability for special operations. Meanwhile, the Navy Computer Adaptive Personality Scales (NCAPS) of the United States were used in assessing individuals over their suitability to the job in the navy (Underhill & Lords, 2002). There were also studies suggesting the implementation of personality testing in the United States fire fighter's selection criteria (Younts, 2003). Besides that, the Canadian Forcers were suggesting the application of psychological screening which includes personality at the selection of Snipers (Girard & Scholtz, 2005).

Besides that, personality testing was used by NASA reviewing the suitability of an astronaut candidate and adding them in their selection of astronauts (Musson, Sandal & Helmreich, 2004). In competitive environments, more information about a person is required in making decisions (Oswald, Schmit, Kim, Ramsay & Gillespie, 2004). Therefore, non-cognitive elements such as personality are used to predict individual's suitability during the selection of students into the higher learning (Oswald et al., 2004).

Psychological measures are also used in diagnosing the problems of an individual in schools and institutions and to apply interventions for problematic students (Wrobel & Lachar, 1998). Similarly, the Royal Australian Air Force (RAAF) uses personality measures in diagnosing the student's problems and assisting them in overcoming their difficulties during their training (Elliot, 1997).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

In Malaysia, the Public Services Department (PSD) is the largest employer of the nation. PSD conducts general assessments in the process of recruiting a graduate candidate for Grade 41 in the civil service (*Suruhanjaya Perkhidmatan Awam*, 2005). This assessment consists of two sections where section I comprises items that embrace general knowledge of Malaysia and its surroundings, problem solving ability, writing and comprehension skills in *Bahasa Malaysia* and the English language, Section II assesses the non-cognitive aspects, that is on personality measurement. Personality is appraised to determine the suitability of a candidate for the job that the candidate is applying. This highlights the importance of personal attributes in recruitment in addition to cognitive abilities. Even after the recruitment, the Public Services Department still enhances an employee's personality in order to ensure a success-oriented culture in the civil service (*Surat Pekeliling Perkhidmatan*, 2009).

Personality assessments are also being used to determine the suitability of students as regards to their choice of course when applying for a place in the university (*Ujian Penetapan Khas*, 2008). University Science Malaysia (USM) administered the Malaysian University Selection Inventory (MUnSyI) to facilitate the placement of students into the appropriate courses of study (*Ujian Penetapan Khas*, 2008). MUnSyI encompasses five distinctive domains including career interest, personality, integrity, emotional intelligence and patriotism. The MUnSyI self-reported measure outcomes are used to place students into different courses in the university.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Besides the MUnSyI test, the Malaysian Educators Selection Inventory (MEdSI) is another test administered to candidates applying for teacher degree programs in Malaysia (MEdSI, 2009). The self-reported MEdSI measures four domains which are career interest, personality, integrity value and emotional intelligence. MEdSI is very specific in nature because it is designed particularly for teacher candidates. Therefore, the four domains would assess candidates over the qualities required in the teaching profession. Nonetheless, candidates of MEdSI are required to pass the MEdSItestin order to proceed to the interview session before they are recruited into the teaching program (Arifin Bin HjZainal et al., 2009).

2.3 Definition of Social Desirability

Social desirability is termed as the tendency to respond to self-reported items in a way that presents the individual in a good light rather than to respond in an accurate and truthful manner (Holtgraves, 2004). This has led to a systematic responding to items on some basis other than the specific item content (Robinson, Shaver & Wrightsman, 1991). Socially desirable responding was described as the way individuals project themselves in socially desirable image (Cervellione, Lee & Bonanno, 2009). Richman, Kiesler, Weisband and Drasgow (1999) defined social desirability as respondent's propensity in manipulating responses in a socially desirable manner under different test conditions, mode and administration. Therefore, social desirability can be concluded as the tendency for individuals to distort their responses according to the test conditions with the intention to present themselves in socially desirable and acceptable images.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Social desirability is associated to impression management. The term, impression management is defined as consciously and intentionally distorting responses towards positively desirable ways (Charles, 2003). Impression management was also known as motivational distortion which takes the form of faking good that refers to socially desirable responding (Hakstian & Ng, 2005). Meanwhile, impression management was also described as one's purposeful response to create the most positive social image (Robinson et al., 1991). Stocke and Hunkler (2007) defined impression management as respondents giving biased answers with the intention of obtaining social approval from others. In addition, impression management was classified as a voluntary projection of individuals on how others want to perceive them in a positive way (Salgado, 2005). Impression management's underlying concepts of intentional distortion of responses to portray socially desirable image for other's acceptance does show similarity between the social desirability and impression management terms. Although the terms differ, the meaning of social desirability and impression management are the same.

2.4 Social Desirability in Personality Assessment

The literature review found that the opportunity to fake, personal characteristics of individuals and situational factors contribute to faking (Magnus et al., 2006). Research on person and personality test format that contribute to socially desirable responding are discussed in detail under each following subtopics.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

2.4.1 Person as a contributing factor to social desirability in personality assessment.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Morgeson, Campion, Diboye, Holleabach and Schmitt (2007) in their study reported faking as a result of person and situation interaction. This means the degree of faking differs between individuals and the situational demands, whereby the situations of the job applicants provoke the applicant's motivation to distort their responses. Indeed, job applicants were found to be more motivated to fake their responses compared to the incumbents of the job (Day, 2008). In a study by Harvey, Wilson and Hansen (2005), instructions to fake the personality test to the Troopers officers found score elevations compared to their responses in a honest condition. In addition, when test users are viewed in applicant condition, they tend to elevate their scores significantly compared to the honest condition (Harvey et al., 2005).

According to Hakstian and Ng (2005), job applicants distort the responses to increase the likelihood of getting the job they have applied for and the distortion occurring in employment is referred to as employment related motivational distortions. This is because respondents manipulate their answers to gain social approval from others (Stocke & Hunkler, 2007). However, socially desirable responding is linked to the test taking situations (Kroner et al., 2004). In this instance the test administration and implications influence the occurrence of socially desirable responding. Studies have found the discrepancy between the observed and true nature of a respondent in a motivating setting as intentional distorting or as faking (Dilchert et al., 2006).

Meanwhile in Edith, Stoop and Meijer's (2000) study, it was mentioned that a person's ability can be invalidated due to the familiarity of the questions as well as

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

guesses of answers and the pre-knowledge of some items. Moreover, the familiarity of items could easily determine the social values placed on the scale item which leads to the likelihood of socially desirable responding (Dilchert et al., 2006). Therefore, the variety of personality testing found online and in magazines exposes the test-taker over the likelihood of the items in personality testing. Moreover, free and accessible personality inventories attract more individuals to test and infer their characteristics. The great exposure on personality inventories lets the test-takers map out items that would respond to a particular construct in a personality testing.

In the study by Magnus et al. (2006), relationship was examined between social desirability with emotional intelligence, over claiming and self-esteem variables. Their study found that lower self-esteem individuals were more likely to be engaged with socially desirable responding. While, individuals with greater emotional intelligence also had a tendency to socially desirable responding. It is believed that lower self-esteem individuals try to respond in a socially desirable manner with the intention of gaining social approval from others. However, emotionally intelligent persons are well adjusted individuals and their socially desirable responding could be their actual self-description.

The study by Iddekinge et al. (2007) mentioned that impression management was found in the form of verbal statements, non-verbal behaviors and modification of appearance during an interview session. Explanations of past experience and competence in Behavioral Descriptive Interview (BDI) approach in interviews are more likely to use self-promoting impression management (IM). Even though the structure of the interview and the interviewer hints of the interviewee in portraying

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

the best of interviewees, the degree of impression management is moderated by the situational factors.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Meantime, Robie, Tuzinski and Bly's (2006) survey on assessor's beliefs and practices in recruitment related to faking revealed that candidates desire to present themselves in the best light as the reason behind their faking. In their study, they reported that the recruitment assessor believed faking is a threat to the validity of a personality inventory yet agreed that it does not affect the overall assessment of an individual. This is due to the test content and the outcomes of the test that influence the reactions to the personality tests. As a conclusion, the test implications and consequences motivate an individual to distort the response in a socially desirable manner with the help of the past experiences of the individual.

2.4.2 Personality test format as contributing factor to social desirability in personality assessment.

Measures of non-cognitive variables are found to be vulnerable to response biases (Rennie, 1982). This is because research has demonstrated evidence of faking of non-cognitive self-reported measures (Alliger & Dwight, 2000). According to Kroner et al. (2004), self-reported inventories are vulnerable to alternative interpretations which then become a threat to the validity of the inventories. A study by Yang et al. (2000) found that a self-reported inventory has higher mean differences compared with others rating on individuals using the same inventory. It is supported that the mean for responses changes according to the faking instructions given to the respondents (Viswesvaran & Ones, 1999). Viswesvaran and Ones (1999) in their study also reported that respondents are able to elevate their responses to half

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

a standard deviation when asked to fake. This is because individuals have the tendency to endorse items on interest highly to match the purpose of the test.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The study by Yang et al. (2000) found that NEO PI inventory was susceptible to response bias. The response bias in NEO PI was examined and found that those respondents who were asked to fake good scored lower on Neurotic and higher on Agreeableness and Conscientiousness dimensions. Inversely, the fake bad respondents scored low in Agreeableness and Conscientiousness and higher in Neurotic dimension. This finding shows the ability of the respondents to identify the items representing different dimensions in a personality inventory. It could be due to the familiarity of extraversion and neuroticism dimension in personality which makes the items representing the dimension recognizable and therefore easily manipulated according to the situational demands (Kirkcaldy, 2001).

Meantime, Steffan, Kroner and Morgan (2007) in their clinical setting found that prior knowledge of a symptom helps individuals in dissimulating malingering effects. Prior knowledge of the symptom would help to determine the desirable and undesirable symptoms of a specific medical condition. Advancement over the knowledge then would guide them to respond to items assuming to be the best answers which would eliminate them from any medical conditions. Besides that, the transparent meaning of items clearly indicates the most positive answers to the candidate which then enables the candidate to fake the answer easily (Morgeson et al., 2007). Holtgraves (2004) mentioned that the items ambiguity also contributes to the problem of socially desirable responses. It was stated that items with high ambiguity makes the test less stable under repeated measures (Ferrando, Lorenzo &

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Molina, 2001). When test items are ambiguous it tends to differ in perception across individuals (Zickar & Ury, 2002). When items were responded differently under different occasions, this would lead to differences in tapping the individual's characteristics. The lack of consistency in the scoring would then question the reliability of the test. Moreover, lack of resources in checking and verifying the characteristics of the applicant does contribute to faking in personality measures (Isaacson, Griffith, Kung, Lawrance & Wilson, 2008).

Besides items, Underhill and Lords (2006) underlined that even that scaling format of an inventory is vulnerable to faking and less predictive in behavior. Jackson, Wroblewski and Asthon (2000) reported that the traditional single stimulus (Likert scale) items have scores that are higher than expected. The continuous items are more prone to faking because in continuous items the individual has the opportunity to fake to the maximum positive and minimum negative behaviors (Graham, McDaniel, Douglas & Snell, 2002). That shows that even the format of scales influences the individual's responses in an inventory. It is suggested that forced choice format items are fake resistant and better behavior predictors compared to the Likert scales (Underhill & Lords, 2006). Besides that, the common personality inventory that is in application is rather general in context (Robie et al., 2000). Problem arises when the general context personality items are administered for a specific purpose, for example for recruitment. When a general personality inventory is used for a specific purpose it would fail to tap the criteria characteristics.

In addition, the paper and-pencil administered tests are found to provide opportunities for the test takers to distort their responses (Cruz & Dipboye, 2003). Research has found less social desirability when computerized testing being

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

conducted (Richman et al., 1999). When ambiguous items, constructs and general context of personality inventory are administered in decision making, then the accuracy of the inference about an individual is doubtful. In conclusion, the nature of the items, type of scales and context specification items contributed to socially desirable responses in personality inventory.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

2.5 Methods used in Reducing Socially Desirable Responding

Several approaches have been suggested for reducing social desirability effects in personality inventory. The approaches include the standard guidelines in test administrations, forced-choice format, randomized responses format and equal distribution of positive and negative items.

2.5.1 Standard Guideline in Minimizing Socially Desirable Responding

The American Psychological Association (APA) and American Counselor Association (ACA) clearly draw guidelines in the administration of psychological testing. These guidelines help to minimize occurrence of biases in applying psychological measurement in assessing an individual. These ethical concerns are essential due to the impact created on the functioning of the individual. The result has the potential or is vulnerable in various aspects of the individual's life (Welfel, 2002). The following subtopics are the extracted APA and ACA guidelines in the test administration. They are regarding test constructions, client welfare and explaining test results, obsolete test and outdate test results and diversity in testing. The subtopic shows the extent to which test construction and test results are monitored from any

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

biases issues. When bias potentials are confronted at the onset of the construction of the test, the test output would provide a valid test. Once the inventory is out of bias, then it would minimize the potential contributions of the social desirability when administrating the test. The following is the APA and ACA guidelines with its explanations.

2.5.1 (a) Test construction

ACA : Section E.12 and APA : Principle section 2.03

This section requires test constructors to apply standardized scientific procedures, professional knowledge in the development, publication and utilization to eliminate bias.

2.5.1 (b) Client welfare and explaining test results

ACA : Section E.1.b and APA : Principle section 2.09

This section ensures proper explanation is given to the person assessed as the person has the right to know.

2.5.1 (c) Obsolete Test and Outdated Test Results

ACA : Section E.11 and APA : Principle section 2.07

The section highlights that outdated test and results that do not serve the purpose of current decision or purpose will not be used for decision making.

2.5.1 (d) Diversity in Testing

ACA : Section E.8 and APA : Principle section 2.04 C

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

This section is an attempt to determine the population prior to administration to ensure that the test is applicable and able to measure the construct that has been looking for.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

2.5.2 Forced choice format

Aiken (1999) advocated forced choice format rating as an applicable method of reducing faking. The forced choice format provides respondents with a set of paired items in which each has the same social desirability effects (Horst, 1968). The pair of forced choice items represents different dimensions measured under a personality inventory. Individuals will be instructed to select the items that best reflect themselves and through this the potentials to fake are reduced. The forced choice format makes individuals select either response therefore minimize the chances of faking in the test.

2.5.3 Randomized response (RR)

Test-taking attitude is stated as a concern in the selection of methods in testing (McCarthy & Goffin, 2003). This is because the willingness of a respondent to respond to a test will affect the way a respondent responds. Fox and Meijer (2008) in their study stated that randomized response (RR) technique gains cooperation from respondents and reduces their motivation to falsify their responses. In this randomized response (RR) method, respondents will perform the randomization and only the respondent will know the questions. Meanwhile, the administrator will only receive the responses. These randomization responses would maintain the privacy and anonymity of the responses. Therefore, the pressure of being evaluated or

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

observed by others would be eliminated and would increase the willingness of the respondents to respond honestly. Meanwhile, randomized format of items in a test eliminates the cues about the items and their construct (Wolford, 2009). This means scattered items of the construct across the tests minimizes the chance of guessing the items and the underlying construct which then reduces the possibilities of faking by the respondent.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

2.5.4 Equal Positive and Negative Items

Rennie (1982) proposed that an equal number of positively and negatively keyed items in a personality test would minimize the faking possibilities. Summation of item scores from positive and negative items would cancel out the effects of biased responses. The negatively keyed items would cross-check the positive responses of an individual from the same domain in a personality test. Therefore, the cross-checking and reversely coded values of negative items would eliminate the bias responses from the inventory.

2.6 Methods used for Detecting Socially Desirable Responses

Various methods have been applied in reducing socially desirable responses in personality assessment. However, the methods in reducing socially desirable responses did not guarantee that respondent will provide honest responses. Hence, it becomes difficult to differentiate the honest and socially desirable responses from their observed responses. Therefore, methods of evaluating the observed responses were reported useful in detecting socially desirable responses. Among them is the social desirability scale and observed item responses.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

2.6.1 Social Desirability Scales

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Social desirability was studied as early as the 1950's and viewed as motivated, purposeful behavior residing in the respondent's goals and belief on being evaluated (Crowne & Marlowe, 1960). The vulnerability of questionnaires towards socially desirable understanding led to the application of social desirability scales in psychological testing (Leite & Beretvas, 2005). Social desirability scales were used to clarify the credibility of a person's response to the items in the questionnaire (Lonnqvist, 2008). Social desirability scale ensures that the phrasing of survey questions does not encourage socially desirable responses (Leite & Beretvas, 2005). Meantime, high scores on the social desirability scales becomes the evidence of socially desirable responding (Lonnqvist, 2008).

Earlier, social desirability scales were embedded within the psychological inventories. For example, in Minnesota Multiphasic Personality Inventory (MMPI), three validity scales were identified , namely (1) K-scale to detect defensiveness, (2) Lie- scale to detect faking good responses and (3) F-scale which examines the carelessness and confusion of the participants (Crowne, 1979). The good impression scale in the California Psychological Inventory (CPI) assists in detecting faking good responses among job applicants (Boer et al., 2008). The Comrey Personality Scale also has built in scales called validity scales that detects random responses and the social desirability responding is detected through the response bias scale (Comrey, 2008).

Even though social desirability scales were incorporated within the inventories, yet in 1957 Edwards Social Desirability Scales (EDS) became the first social

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

desirability scale that stood as an independent measure (Stober, 2001). This EDS scale is a compilation of 39 items of high or low social desirability ratings from MMPI (Paulhus, 1991). Later in the 1960's the Marlowe-Crowne Social Desirability Scale (CMSD) was developed based on the need for approval basis (Crowne, 1979). Paulhus (1991) described the CMSD measurement consisting of 33 true-false items about behavior that is desirable but rare or undesirable but common. The absence of psychopathology effects and ability of detecting socially desirable responses saw the CMSD scale being used widely (Stober, 2001). Over the years, Paulhus (1984) proposed Balanced Inventory Desirable Responding (BIDR) that consists of two subscales and they are (1) self deception enhancement (SDE) and (2) Impression Management Dimension (IM) (Paulhus, 1991).

The self-deception enhancement measures the unintentional propensity to portray oneself in a favorable manner while impression management as intentionally portraying the best of oneself. In addition, Paulhus (1991) also listed the RD-16 instruments and Children's Social Desirability Scale as other inventories that assist in detecting socially desirable responding. Besides detecting socially desirable responding, the social desirability scales were also used to determine the discriminate validity of a psychological instrument (Beretvas, Meyers & Leite, 2002). Beretvas et al. (2002) also added that the social desirability scale was used in factor analysis to determine factors that differ from social desirability scales and helped in deleting elevated responses of the respondents.

Besides the respondent's motivation to distort the responses, item transparency which means item clarity and the presentation of item content facilitates individuals

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

to distort their personality items (Detrick & Chibnall, 2008). This highlights the fact that not all items are likely to be responded in a socially desirable manner. Inflation of scores and social desirable responses reflects error in the scores and would affect the reliability of the score outputs. The degree of reliability for a test score is essential since the personality scores are correlated with social desirability scales (Li & Bagger, 2007). Empirical evidence of low correlation between social desirability scales with a survey output reflects low impact of socially desirable effects on the survey questionnaires (Leite & Beretvas, 2005). Li and Bagger (2007) found that low correlation between Balanced Inventory of Desirable Responding and Personality construct as a sign of lack of bias in a personality inventory. While, correlations between the desirable responses and social desirability scale report the presence of socially desirable responding (Lonnqvist, 2008).

Even though external social desirability scales were used to detect socially desirable responses yet its accuracy is still debated. It is because social desirability is developed based on the "need for approval" construct (Crowne & Marlowe, 1960). Crowne and Marlowe (1960) stated that social desirability scales describe items reflecting the good-bad dimension which are likely to be untrue of most people. Items on the social desirability scale indicate the possibility of respondents to distort the responses in order to get approval from others. What extent would an individual take to distort the social desirability scale? In addition, the determinant of correlation coefficient among personality tests and social desirability scales is an empirical evidence of social desirability existence. But, how far is the personality score reliable? It would also be possible that low correlation of personality and social desirability is due to poor reliability (Li & Bagger, 2007). Therefore, empirical

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

evidence of correlation coefficient will not be substantial enough in determining and detecting socially desirable responses.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

2.6.2 Detecting social desirability from item responses

Faking by applicants on personality measures are defined as deliberate alteration of responses (Griffith, Peterson, Quist & Evan, 2008). Meantime in socially desirable conditions, respondents have a tendency to inflate their scores in traits that respondents believed to portray a good image of themselves (Salgado, 2005). Thus, individual scores can be regarded as valuable information when examining for the presence of socially desirable responding. The mean difference approach was applied in detecting the presence of socially desirable responses. The study by Griffth, Chmielowski and Yoshita (2007) examined the mean score differences in a personality inventory between the honest and faking good responding groups. In their study, the honest group response was set as a baseline of score distribution to compare the faking good's score distributions. Hereby, the faking good group scores fell between 30% to 50% of the confidence interval and respondents with scores within this range were categorized as fakers.

Sellbom and Bagby (2008) studied the mean comparisons in the NEO-Personality Inventory (Revised) scores for differential responding groups. The scores of personality inventory were compared between the valid reporting, invalid overreporting and invalid underreporting and significant differences in the mean scores between the groups were found. Mean differences were also found between the job applicants and incumbents on an employment based inventory (Griffth et al.,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

2007). This is because job applicants are more motivated to fake the responses than the incumbents (Day, 2008). Thus explains the job applicant's desire and the implications of the tests as the reasons for faking the inventory. Besides, the meta-analysis investigation by Birkeland, Manson, Kisamore, Brannick and Smith (2006) reported standardized mean differences between the job and non-applicants scores. In the meta-analysis, 33 studies of comparing job applicants and non-applicants personality scales scores were analyzed and differences in standardized mean average for the Big 5 dimensions were found. The result of the analysis indicated that the standardized mean score for applicant is significantly higher than non-applicants in four of the five dimensions. They are Extraversion, Emotional Stability (Neuroticism), Conscientiousness and Openness.

However, only individuals with the intention to distort will fake their responses in a test (Douglas, McDaniel & Snell, 1996). This explains the individual's contribution in response alternation which is also subject to change between different individuals. The verbal protocol analysis showed that individuals selectively analyze the social values of the items and its implications before responding to the items (Robie, Brown &Beaty, 2007). The initial stage of verbal protocol analysis approach is to administer the questionnaire under the instruction that the best respondents will be rewarded. While attempting the questionnaire, respondents were asked to say out loud whatever that strikes their mind at the point of answering the questionnaire. Then, these responses are recorded and coded into groups (example: faking, honest) and compared to the scores from the questionnaire. The difficulty in conveying what is in the mind of the respondents and subjectivity in interpreting the statements may results in inconsistency of classification as faking or honest group. Bearing in mind

UNIVERSITI PENDIDIKAN SULTAN IDRIS

that a test determines the future, it would be possible for respondents to respond with caution and think over their responses for the test. The verbal protocol analysis showed the tendency of respondents to distort responses in a socially desirable manner. This is because socially desirable responding respondents not only respond to item content but also reflect on the items of social values (Tan & Grace, 2008).

Responding to items according to the social values showed that individuals are selective in distorting their responses in a socially desirable manner. The selectiveness of item selections explains that respondents do not fake every item but only items that are related to their situations and demands (Ziegler, Toomela & Buhner, 2009). The variances in the score distribution could lead to variances in the construct measurement across different groups of people (Meade, Ellington & Craig, 2004). Therefore, it is essential to examine the response patterns of every individual and this is possible according to the Item Response Theory (IRT).

In IRT, an individual's estimated trait level in a test is derived from the responses of the individual and the properties of the administered items (Embretson & Reise, 2000). From the measured individual trait ability and item parameters, it would be possible to map out an expected pattern of responses for an individual. However, it is only an expected response pattern for a specific individual ability relative to the item parameters. In reality, a person's observed response would not be exactly the same as expected scores from the model. Therefore, it is essential to examine the differences between the observed and expected pattern scores of an individual and the discrepancy determined and coined as person misfit (Ferrando & Chico, 2001). This explains the inconsistency in a person's response patterns for an

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

estimated ability and item parameters (Day, 2008). The inconsistencies are studied through appropriateness measurement to identify the misfitting respondents (Schmitt, Chan, Sacco, McFarland & Jennings, 1999).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

In other words, the appropriateness indices aid in detecting scores that function differently in the large data sets (Meade et al., 2004). Meanwhile, the appropriateness measurement ideas were applied in Brown and Harvey's (2003) study to detect intentionally distorted responses. Knowing the person's ability and item parameters, it is possible to draw out an expected pattern of responses from a particular individual and when the observed score is compared, it is assumed to be approximately near to the expected score. Deviation from the expected score would create an inconsistent pattern of responses. Besides that, an individual's misfitting response patterns could be examined through a person response curve (PRC) (Day, 2008). The person response curve illustrates the relationship between the endorsement item responses to the increasing item difficulties. In PRC, a negative slope explains a decreasing response endorsement towards increasing item difficulties. Therefore, a less negative slope indicates misfitting response patterns as their responses are not according to the level of difficulty. The less negative slopes indicate the misfitting response patterns. The person fit indicates the extent to which the item response model fits the data for each person (Meade et al., 2004). In addition, the Mix-Model of IRT categorizes similar response patterns within the groups (Zickar, Gibby & Robie, 2004). The mix-model IRT model helps to determine the variability of responses within a group.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Besides person fit, item response theory also explains the functioning of items across different groups. In IRT, differential item functioning (DIF) identifies items which function differently between groups with same ability (Emberetson & Reise, 2000). The detection of DIF explains how items are viewed and interpreted diversely for individuals with the same ability from different groups. The DIF method assesses test-takers response patterns to specific test items (Perrone, 2006). In addition, DIF aids in detecting test items that are prone to biasness whereby items are compared between groups to examine its invariances in the measurement (Perrone, 2006). Abedi, Leon and Kao (2007) studied on how items of reading assessments function differently between children with and without disabilities. Thus, the DIF method did find that the test items were advantaged to the children without disabilities as they perform at a higher level in the reading assessment. This supports that the test may not be invariant across different groups of samples. Meantime, the study by Mitchelson, Wicher, LeBreton and Craig (2009) found that 33 items of the Abridged Big Five Circumflex (AB5C) personality inventory flagged for DIF across gender and ethnicity, which is the items were responded differently across different groups. Besides that, the application of DIF in the Hogan Personality Inventory reported items functioning differently between gender and race groups (Sheppard, Han, Colarelli, Dai & King, 2006). DIF compares individuals with the same ability in different groups and shows how a group membership contributes to the changes in the item measurements.

When items function differently across groups, it leads to variances in the construct measurement between the groups (Meade et al., 2004). The measurement equivalence portrays equal probabilistic endorsement of individuals with the same

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

trait values (Waller, Thompson & Wenk, 2000). Meantime, the equivalence of factor loadings of items across groups showed measurement invariances (Tucker, Ozer, Lyubomirsky & Boehm (2006). However in a faking situation, items from different dimensions were found to be correlated and contributing as distinctive factors (Ziegler &Buehner, 2009). The inter-correlated items from different dimension and formation of new dimension showed the variances and reported the influence of faking instructions in the changes of factor structure of an inventory. The study by Vanheule, Rosseel and Vlerick (2007) examined metric measurement invariances of the Maslach Burnout Inventory for human services between hospital and welfare institution nurses. It was found that item loadings on factors are substantially different between the samples. This suggested that the meaning of items is interpreted as sample specific, where the different scores on items are due to the way they conceptualize the behavior being assessed (Nye, Roberts, Saucier & Zhou, 2008).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

2.7 Theoretical Framework

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Based on the literature review, studies have reported differences in the mean scores of personality dimensions between the honest and faking response groups. Commonly, the faking groups mean are reported higher than the honest group. The mean scores differences between groups was reported as evidence of individuals faking behaviors (Burkevich, Jenkins & Griffith, 2007). Study by Viswesvaran and Ones (1999) reported instructions to fake has inflated individuals scores of at least a half standard deviation compared to the honest responses. Hence, score inflation by respondents suggests possibility of faking respondents. It was also found that high scores in non-cognitive tests under high-stake situations were causes of response distortions (Dilchert et al., 2006). This concludes that, faking instructions causes score inflations and thus lead to high scoring in the non-cognitive tests. Therefore, the scores of individuals could be studied as a source of detecting the socially desirable responses. In this study, the scores of honest and socially desirable groups were compared. The distribution and dispersion of scores were examined and compared between the honest and socially desirable groups. The differences in score distributions between them are used as a source of identifying characteristics of socially desirable responses.

In the Rasch model every response is known as stochastic, where an individual's response selection depends on the level of trait on an occasion (Curtis, 2004). Therefore, the Rasch model aids in modeling the expected response pattern of individuals with their estimated ability and item parameters. Meanwhile, the differences between the observed responses from personality inventory and expected scores are determined as an indication of a person misfit (Ferrando & Chico, 2001).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

This study adapted the approach of person fit to determine the non-fitting response patterns by comparing respondents observed scores to the expected responses from the model. The influence of socially desirable responding was examined in the individuals observed responses and in the discrepancy between respondents observed and expected responses. In addition, scores or responses from non-cognitive measures can also be used in understanding the items and dimensions measured in an inventory. In high stake testing, respondents reported to have the tendency of suppressing their negative traits while enhancing their positive traits to portray their best image possible (Kirkcaldy, 2001). This shows the importance of selectiveness of items and traits to be responded in a socially desirable manner. The selection of items and traits for socially desirable responding implicates the differential interpretation of the items or tests between different groups. In other words, only certain items and dimensions are susceptible to bias responses.

The observed responses of honest and socially desirable groups were examined to determine whether the dimensions in a test function similarly across different instructions. This is because literature has reported only certain items and dimensions of personality inventory as susceptible to response biases. Therefore, the honest and socially desirable group responses were studied to determine the items that are susceptible to socially desirable responding. In identifying the items, the differential item functioning (DIF) approach was used. DIF in IRT detects items that function differently between groups with the same ability level (Embretson & Reise, 2000). This study applied DIF in detecting personality items that function differently between the honest and socially desirable groups. The DIF was used in detecting items that are perceived and responded differently due to the impact of socially

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

desirable responding. Hence, the test information functions for all the five dimensions of the personality inventory were compared between the honest and socially desirable groups. The comparison of information function between the honest and socially desirable groups was examined to determine whether the dimension provides invariant information across different situations. This is because traits are differently biased and scores of certain traits are inflated more than others (Hirsh & Peterson, 2008). Therefore, the comparison between the information function for honest and socially desirable groups showed the changes in obtainable information for the same ability level at different instructional settings.

The effects of socially desirable instructions were also studied through factor analysis. Factor analysis was used to examine the differences in factor structures of International Personality Item Pool (IPIP) between the honest and socially desirable group. The evaluation of factor solutions and item loading of dimensions were compared between the groups. With that, the changes in the factor structure reflected the impact of socially desirable responding and the affected dimension. Therefore, the affected dimension could be identified as the dimension vulnerable to socially desirable responding. Based on the arguments above, the following conceptual framework was developed.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

2.7.1 Conceptual Framework

The review of literatures has stated differences in the response provided by the honest and socially desirable group. In this study honest and socially desirable groups response is assumed to provide differential response pattern due to the influence of responding instructions. The response from honest and socially desirable groups was compared in identifying the characteristics of socially desirable responses. Therefore, the score dispersion and non-fitting response pattern was examined in detecting the socially desirable responses. Meanwhile, factor analysis, test information function and differential item functioning were used in identifying items and personality dimensions that are affected by the socially desirable responding. The following diagram conceptualizes the execution of the theoretical framework for this study.

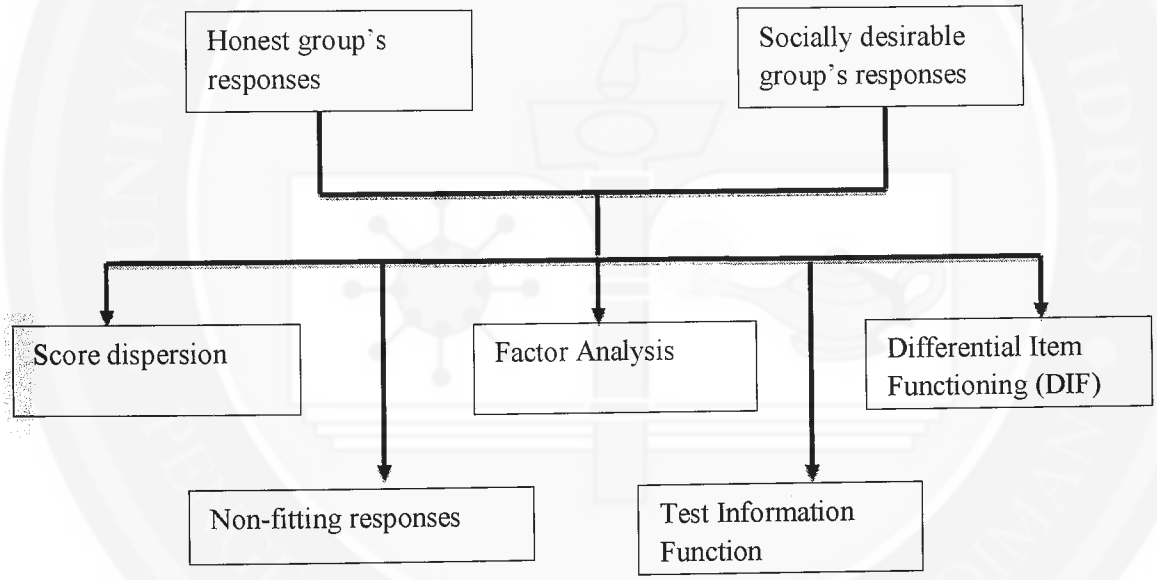


Figure 2.1 Conceptual Framework

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

CHAPTER THREE

METHODOLOGY

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

3.0 Introduction

In chapter three, the research methodology is explained in detail for the execution of the research objectives. Firstly, the research design is explained follow by the instrumentation process and the factor analysis resultsof the instrument. Findings from the pilot study and factor analysis results were used to validate the instrument used in this study. The selection of the sample, and the setting up of the honest and socially desirable group were described and concluded with the explanation of the methods of data analysis.

3.1 Research design

An experimental repeated measure design was applied in studying the socially desirable responses. In this experimental design, the sample was tested twice through survey technique. The sample was asked to respond to a personality questionnaire by describing themselves. During the first administration, the respondents were asked to respond honestly to personality inventory. After a three-week interval, the same personality inventory was administered again to the same sample. But at the second administration, the sample was asked to respond in a socially desirable manner. Hence, the socially desirable group is purposely created for the purpose of creating a differential pool of responses for the study by instructed the respondents to do so under a simulated conditions. The responses from the first administration were classified as the honest group and responses from the second administration as

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

socially desirable group. This experimental study was a between-group study and the procedure of repeated sample was to minimize the person characteristics variances between the honest and socially desirable groups. The content and test duration was the same across the two administrations. The responses from the honest and socially desirable groups were compared to determine individuals whose responses are distorted in a socially desirable manner.

3.2 Instrument

3.2.1 Selection of instrument

The process of selecting an appropriate instrument was tedious and time consuming. It was tedious because there were criteria set before selecting a suitable instrument. In the process of selecting a personality test, the purpose of the test and its context of application were examined. Next, the items, response format, norms, scoring and interpretations were studied extensively. Meantime, the empirical evidence of reliability and validity of personality inventories were also used as criteria for the process of selecting a personality inventory. Due to financial constraint, the cost of the personality instrument was also one of the factor taken into consideration. Besides that, failure in obtaining permission to use locally developed inventories like *Sidek'sInventori Personality Warna (IPW)* lead to the use of IPIP for this study. Based on the mentioned criteria, Inventory Personality Item Pool (IPIP) was chosen.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

IPIP is an open source instrument and can be downloaded free at www.ipip.org. IPIP has the advantage of easy administration as it is a self-reported instrument with 5-point Likert scale response style. Being a public domain inventory, there are easily available evidence of proven convergent and discriminant validity (Lim & Ployhart, 2006). Furthermore, items in IPIP are short and use only simple words in all the statements. As IPIP did not have any norms set, it is practical and applicable in a variety of situations. Scoring of the inventory only requires the summation value of the response to every item. The interpretation is based on the total score of each dimension.

3.2.2 Description of instrument

In this study, IPIP was used to assess student's personality. This instrument measures the five-factor model of personality (Goldberg, 1999). The IPIP inventory taps an individual's level of Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. IPIP is a pool of short items that describe personality attributes. Two shortened versions of the IPIP questionnaire are available online; one with 50 items and the other has 100 items. In this study, the version with 50 items was used and presented in a bilingual format. The Malay translation has been done by Hazalizah Hamzah for her PhD thesis (the work of her thesis is still in progress). Permission to use the Malay version was obtained from her. The permission letter granted by her is attached in Appendix A.

However, a close scrutiny of the Malay items showed some items in Malay has different meanings compared to the original English version. Hence, a simple

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

translation process that consisted of three phases was done. During phase one, the original Malay version translated by Hazalizah Hamzah was refined and reworded by checking with the original English version. The list of modified items is presented together with the English and Hazalizah's Malay version in Appendix B. During phase two, the refined Malay items were sent to a retired *Bahasa Malaysia* teacher to check the meaning of the items and the appropriateness of the language structure in Malay. Feedbacks from the teacher were used to improve the items. Upon improving the items based on the Malay language teacher's suggestions, the instrument was sent to the same teacher and two Master's students to evaluate the comparability of meaning between the Malay and the English versions. Comments and suggestions from these reviewers were used to further refine and rephrase the Malay items. Finally during phase three, the instrument was administered in a pilot study to determine the suitability of the instrument as well as to identify flaws in bilingual items presented in the instrument.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.1

Classification of IPIP items according to the dimension and types of items.

Personality Dimension	Positive Items	Negative Items
Extraversion	1, 11, 21, 31, 41	6, 16, 26, 36, 46
Agreeableness	7, 17, 27, 37, 42, 47	2, 12, 22, 32
Conscientiousness	3, 13, 23, 33, 43, 48	8, 18, 28, 38
Neuroticism	4, 14, 24, 29, 34, 39, 44, 49	9, 19
Openness	5, 15, 25, 35, 40, 45, 50	10, 20, 30

Table 3.1 illustrates the classification of items into five personality dimensions as well as grouping of positively and negatively keyed items. Every dimension is measured with ten distinctive items each. All four dimensions of personality have unequal positive and negative items except extraversion dimension where it has equal positive and negative items. Items in IPIP are scored using the 5-point Likert Scale, where "1" represents Strongly Disagree (SD) to "5" representing Strongly Agree (SA). However, in negatively scored items, the scores are reversed. The range score of each dimension is between ten to fifty and the total of each dimension were be used to determine the dominant dimension in a respondent.

The initial study of reliability of IPIP for Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness (based on Cronbach's alpha for internal consistency) were reported as .87, .91, .79, .86, and .84 respectively

(Goldberg, 1999). Another study carried out with the Malay version of IPIP reported lower internal consistency of the factors with reliability values of .78 (Extraversion), .65 (Agreeableness), .71 (Conscientiousness), .86 (Neuroticism) and .72 (Openness) (Arifin Bin Zainal, et al., 2009). In the current study, the reliability values obtained were .74 (Openness), .78 (Conscientiousness), .69 (Extraversion), .75 (Agreeableness) and .69 (Neuroticism). Two dimensions reported reliability values of closer to .70 and the rest reported values of more than .70. According to George and Mallery (1995), reliability values of more than .70 are regarded as acceptable. Since all dimensions reported reliability values of closer and more than .70, the results from this analysis are considered as acceptable.

3.2.3 Factorial structure of the instrument

The International Personality Item Pool (IPIP) is based on five dimension model with each dimensions represented by ten distinctive items. Based on the hypothesized model of IPIP, a Confirmatory Factor Analysis (CFA) was conducted on the honest responses to examine the coherence of the item loading to the dimensions as explained by the model. The default setting of Principal Component Analysis (PCA) in SPSS was performed for the extraction of factors. The Kaiser-Meyer Olkin (KMO) measure of .855 verified the sampling as adequate for analysis. Barlett's test of sphericity for honest group was 6899.049 (1225) = .000, $p<.01$ indicated that item correlation between items were significantly large for principal component analysis. Table 3.2 shows the variance explained by each factor in the analysis. The first factor showed higher percentage of variances compared to the rest of the factors. While, there is a slight differences in the percentage variances for the second and third

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

factors. Factor five has the lowest percentage of variance compared with the rest of the factors. Altogether, the five factor accounts for 36.3% variance. The low total variance by the five factors could be due to the restriction of the number of factors to be extracted from the responses to five only as suggested by the hypothetical model. Besides the total variance explained, the scree plot displayed in Figure 3.1 is used to support the identification of the distinct five factors as the original version.

Table 3.2

Total Variance Explained

	Rotation Sums of Squared Loading		
Component	Total	% of variance	Cumulative %
1	4.783	9.565	9.565
2	3.844	7.688	17.253
3	3.689	7.378	24.631
4	3.104	6.209	30.84
5	2.735	5.471	36.31

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

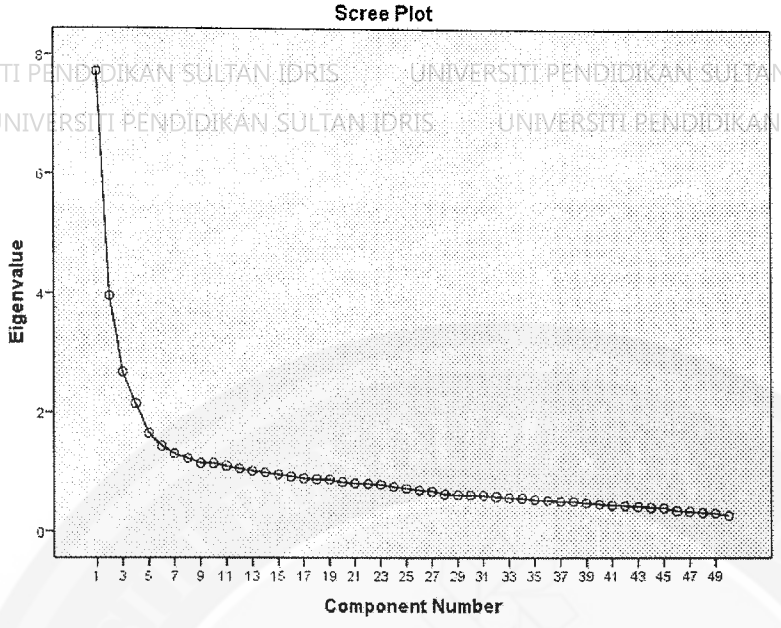


Figure 3.1

Scree Plot to show the number of factors

Varimax rotation was performed and five factors are extracted as hypothesized by the IPIP. The extracted five factors from the analysis were classified according to the original item loadings of IPIP. Rotation of factor solution reported items that did not load according to the original IPIP. Table 3.3 illustrates the rotated component matrix which classified items into appropriate factors. In Table 3.3, the loading value of items from a common personality dimensions was bold. This is to exhibit the underlying dominant items that load into a factor. Therefore, majority of the items were classified according to the factors specify by the English version IPIP. Based on the majorityitems loading, the first component is identified as Openness and subsequently followed by Conscientiousness, Neuroticism, Agreeableness and Extraversion. Therefore, the results of the Malay version IPIP show that the

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

responses exhibit the same five factors as proposed by the original English IPIP inventory. However, 13 items were found to be not loading according to their respective dimensions. Six of these items, Items 47, 27, 37, 7, 12 and 17, representing the Agreeableness dimension while four items (item 46, 6, 16, 26) were from Extraversion dimension. Two items were from Neuroticism and one representing Openness dimension.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.3

Table of Rotated Component Matrix for the 50 items

Item	Component				
	1	2	3	4	5
Item50	**.784**				
Item25	**.754**				
Item15	**.680**				
Item35	**.598**				
Item30	**.537**				
Item5	**.443**				
Item47	.421				
Item45	**.399**				
Item27	.393				
Item37	.337				
Item7	.318				
Item9	-.315				
Item20	**.313**				
Item43		**.648**			
Item48		**.614**			
Item8		**.535**			
Item18		**.491**			
Item23		**.478**			
Item3		**.468**			
Item38		**.465**			
Item13		**.464**			
Item33		**.457**			
Item28		**.444**			
Item12		.384			
Item29			**.627**		
Item44			**.622**		
Item34			**.622**		
Item4			**.621**		
Item49			**.527**		
Item14			**.502**		
Item24			**.495**		
Item39			**.493**		
Item10			-.358		
Item46			-.344		

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.3 Continues

Item	Component				
	1	2	3	4	5
Item6				.529	
Item32				**.504**	
Item16				.495	
Item22				**.490**	
Item42				**.464**	
Item2				**.431**	
Item40				-.409	
Item17				**.375**	
Item26				.374	
Item41					**.606**
Item11					**.582**
Item31					**.568**
Item36					**.511**
Item1					**.511**
Item21					**.492**

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 12 iterations.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

3.4 Pilot study

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

The pilot study was carried out at *Institut Kemahiran Belia Negara* (IKBN), Peretak where students are taking skill-based courses that prepares them for employment. In the Institute, there were altogether 18 classes and they were divided equally into honest and socially desirable responding groups. The honest and socially desirable instructed questionnaires were administered alternately between the classes. At every class, the instructions were read out to the participants and emphasized on the instruction of honest and socially desirable responding group. In addition, for the socially desirable group the researcher asked the respondents to answer thinking on how to impress the researcher. The final sample of the pilot study used for the analysis consisted of 297 students (111 males, 186 females). The item fit (MNSQ) was examined to determine the representation of the items to the five dimensions measured in the inventory. Six out of fifty items showed misfitting and low correlation to their corresponding dimension. This misfit value and correlation pointed to the flaws in the item.

The calculated internal coefficient alpha values of reliability for the five dimensions in honest responses was greater than .90. However, the socially desirable responding responses reported a minimum of .70 except for the agreeableness dimension that reported only .42. Meanwhile, the mean values for Openness, Conscientiousness, Extraversion and Agreeableness reported were higher in honest response group than socially desirable response group. In addition, honestly responding group reported lower Neuroticism mean than the socially desirable respondents. Differential item functioning revealed eight items functioning

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

differently between the honest and socially desirable responding groups. Five out of eight items were negatively scored items. The person outfit values found 66.7% of nonfitting responses were from the socially desirable responded group. The pilot study helped to test the suitability of IPIP and cross checking the translated items in Malay with the original English version. The pilot study also reported 95% of the respondents did not understand the work "keki" in Item 44 which means irritated. Based on the findings, the research design was reviewed to adopt the repeating sample with the respondents being tested twice to minimize group differences.

3.5 Sampling

In this experimental study, secondary school students were the targeted samples. Concurrent with the objective to detect socially desirable responding, the sample of high-stake situation was identified as the most appropriate population for this study. One suitable sample group was the students from *Maktab Rendah Sains MARA* (MRSM) as it sets *Ujian Kecenderungan Kemasukan Maktab* (UKKM) as one of the pre-requisites for Form Three students to be selected for entry into MRSM (*Syarat Kemasukan Ke MRSM*, 2010). The personality element was included as one of the three components administered in the UKKM test. Corresponding to the need of this study, students of Form Two and Four were targeted for this study as these two groups were the closest to the intended Form Three students. Therefore, the sample was obtained from MRSM Kuala Kubu Bharu and from Sekolah Menengah Kebangsaan (SMK) Ketari secondary schools. On the day of data collection, the MRSM has a changed in mind and allowed a much smaller sample size as was agreed originally. Due to insufficient time to secure approval, attempts were made to

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

get approval directly from the school heads nearby. One school that has given permission was SMK Ketari. Although this school differ from MRSM, the students selected share similarities in age group, academic subjects learned and also academic capabilities. This was made possible with recommendations from teachers on the student's language proficiencies and academic abilities in selecting respondents for the study. Teacher's recommendation has been used to involve students with similar characteristics as MRSM students to participate in this study. In SMK Ketari, three best academically capable classes of Form Two and Four were selected for this study. For the Form Four students, they were all from the science stream. In the case of MRSM, there were only three classes for each form, hence all of them were selected for this study. The MRSM's letter of approval to conduct the research in their institution is attached in Appendix C. Table 3.4 illustrates the demographic distribution of the sample obtained.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.4

Distribution of demographic factors.

Variable	Categories of variable	Frequency	Percentage
School	MRSM	230	44.15%
	Government School	291	55.85%
Age	14 years	212	40.69%
	15 years	45	8.64%
	16 years	264	50.67%
Gender	Male	236	45.30%
	Female	283	54.30%
Ethnic	Malay	231	44.30%
	Chinese	247	47.40%
	Indian	34	6.50%
	Others	8	1.50%
Form	Form 2	260	49.90%
	Form 4	261	50.10%

UNIVERSITI PENDIDIKAN SULTAN IDRIS

3.6 Investigation mode and procedure

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

In this experimental study, participants were assessed through the survey technique. Participants were asked to respond to IPIP and tested twice with the same inventory. During both the administration, the teachers in the schools helped to conduct the survey. In order to ensure that they do what is required, the teachers were briefed on the purpose, procedures and test administration instructions. During the first administration, the students were told to respond honestly to the personality inventory. Prior to administration of the test, the teacher read out the following instruction to the respondents.

" Dear Participants,
Thank you for participating in this research. Please complete this personality inventory as honestly as you can. There is no right or wrong answers. It is very important that you respond to the inventory by describing your actual self and not how you want others to see you. The results will be completely anonymous and processed only for research purposes. All information gathered will be kept confidential"

These responses were collected and categorized as honest group. After three weeks interval, the second administration was carried out. During the second administration, the students were asked to distort their responses in a socially desirable manner. Meantime, a scenario was created to motivate the students to distort their responses in a socially desirable manner. The instruction and scenario are as follows:

"Dear Participants,
Thank you for participating in this research. Please respond to the personality inventory as if you were being screened for the best student category award. With the desired intention to win the prestigious award, you should respond in a way that projects the best personality characteristics of yourself which make you the best

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

candidate for the award. The results will be completely anonymous and processed only for research purposes. All information gathered will be kept confidential"

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

The responses from the second administration were categorized as socially desirable group. At each administration, the students took approximately 15-20 minutes to complete the test. The cover page for the honest instruction questionnaire is attached in Appendix D. Meanwhile, a complete set of questionnaire with the cover page of socially desirable instruction is attached in Appendix D(i).

3.7 Data Analysis

The data analysis was carried out to compare the honest and socially desirable group responses from the IPIP inventory. Scores from the personality inventory were compared to identify the characteristics of the socially desirable groups which lead to the detection of socially desirable responding individuals. Firstly, the score deviations of honest and socially desirable groups were examined. Scores from the honest and socially desirable groups were combined and computed for the mean and standard deviations. Based on the mean, score dispersions were categorized accordingly by using the standard deviations. Distribution of the honest and socially desirable groups was examined at each standard deviation interval from the mean. The score distribution is used to examine the differences in score deviations between the honest and socially desirable groups.

Next, the Rasch model was applied in studying the response patterns of the honest and socially desirable groups. In the Rasch model, the probability of endorsing an item depends on the item difficulty and the person's ability. Therefore,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

individuals with higher ability will endorse option categories that are higher on the response category. Thus, the selection of responses was examined concurrent with the respondent's ability and item difficulties. Person fit examines the extent to which a respondent's response conforms to the model (Bell, 1982).

According to Bell (1982), the person fit approach evaluates the respondents abnormal response pattern through calculated standardized squared residual differences between the respondents observed and expected (from the model) scores after considering the person's ability and item difficulties. Thus, this study applied outfit statistics to detect the socially desirable responses. The outfit statistics is a mean square fit statistic which is sensitive to unexpected response pattern on items (Linacre, 2006). The fit is calculated through differencing each pair of observed and expected scores, squaring the differences, summing all the pairs and finally averaging and standardized to a unit normal distribution. Therefore, it illustrates how much does a person's observed score conforms to the expected scores from the model (Mueller, Bullock & Leierer, 2010). In order to detect socially desirable responses, the observed responses of honest and socially desirable groups were combined and compared with the expected responses from the model. The acceptable outfit values ranged between 0.7 – 1.3 (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). Therefore, outfit values outside the acceptable range are classified as non-fitting responses. Then, the frequency of honest and socially desirable groups was computed from the list of non-fitting respondents.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Subsequently, factor analysis of IPIP was conducted separately for honest and socially desirable groups. The factor structures between the two groups were compared. Therefore, a Confirmatory factor analysis (CFA) was performed with extraction of five dimensions and the item loadings were compared between the honest and socially desirable groups. Confirmatory factor analysis (CFA) was used because it is like creating a mould and testing an object against the mould to see how well the object fits (Child, 2006). Moreover, CFA require beforehand the hypothesized number of common factors for extraction (Beck, 1994). On that ground, the hypothesized five factor model of IPIP was set as a mould for factor extractions in honest and socially desirable responses. With the fixed set of personality dimensions, the changes in their factorial structure were compared between the honest and socially desirable group. This is to examine the influence of socially desirable responding instructions to the personality dimensions measured, thus, identifying the personality dimensions that are susceptible to social desirability.

In addition, the test information function graphs were examined to determine the amount of obtainable information from the test with respect to individual ability. As is indicated by the name, test information function tells how much information could be obtained from a test with respect to the ability level of the individuals (Baker, 2001). Therefore, test information function was used to determine whether honest and socially desirable responding groups provide similar information from the same test or whether the instruction to distort the responses in a socially desirable manner influences information from the test. Therefore, the test information function for the five dimensions in IPIP were compared and evaluated between the honest and socially desirable responding groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Analysis on differential item functioning (DIF) was carried out. The item level examination was to find any discrepancy in item functioning in different situations. In examining DIF, the functioning of items were compared between the honest and socially desirable groups whereby the honest group responses were referred to as the reference group and the socially desirable responses as the focal group. Firstly, the presence of DIF between the honest and socially desirable groups were analysed using the Mantel-Haenszel Chi-square. The Mantel-Haenszel Chi-square examined the DIF effect size at $p < .01$ and classified the DIF into categories proposed by the Educational Testing Service (ETS) (Linacre, 2006). According to Linacre (2006), the DIF effect size of $|DIF| \geq 0.64$ logits is categorized as "C" which means the DIF is moderate to large. While $|DIF| \geq 0.43$ logits as category "B" that means the DIF as slight to moderate. The $|DIF| < 0.43$ as category "A" classified the DIF as negligible. In order to avoid false identification of DIF items, another method, the Liu Agresti Cumulative Common Log Odds Ratio $|\hat{\alpha}_{LA}|$ by Penfield (as cited in Miller, Chahine & Childs, 2010) was used to identify DIF items. The criteria of flagging DIF adopted the effect size of $|\hat{\alpha}_{LA}| < 0.53$ as negligible, $0.53 \leq |\hat{\alpha}_{LA}| < 0.74$ as moderate and $|\hat{\alpha}_{LA}| \geq 0.74$ as large (Miller, Chahine & Childs, 2010). In addition, the positive values of Liu Agresti Cumulative Common Log Odds Ratio indicate DIF favours the reference group and negative values shows DIF favouring the focal group (Penfield, 2007c). The presences of DIF was examined through LOR Z and COX Z values, whereby LOR Z and COX Z values of less than -1.96 and more than 1.96 are defined as showing presence of DIF in items (Miller, Chahine & Childs, 2010). Additionally, the Differential Test Functioning (DTF) was applied in examining the DIF effect variances for the overall test and all dimensions for the honest and socially desirable groups. The DIF effect variance (Nu^2) values of < 0.07 is

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

categorized as small, 0.07 <(Nu^2)<0.14 as medium and (Nu^2) > 0.14 as large DIF effect variances (Brown, 2010). The DTF was examined to learn the effect of DIF items to the overall test and to the personality dimensions assessed in IPIP.

Items that flagged DIF was further analyzed for Differential Step Functioning (DSF). The DSF were evaluated to the steps in polytomous items to identify the cause of DIF in items for the reference and focus groups. The levels of DSF in steps were determined through $|\text{CU-LOR}|$ whereby $|\text{CU-LOR}| > 0.4$ is classified as moderate and $|\text{CU-LOR}| > 0.6$ as large effects. Meantime, the presence of DSF was studied by examining the Z statistics which is the ratio of DIF effect variances estimates divided by its standard error estimates. Then, the locus and consistency of causes of DIF for every DIF flagged item was studied. Therefore, the DSF measures, magnitudes and interpretations were done based on the DSF form taxonomy as given in Table 3.5. The taxonomy helps to study the step level effects of the polytomous items to the DIF contributions to the items. The examination of steps and responses lead to the evaluation of frequency of response category of the DIF items. This is to examine the differences in the endorsement of response categories between the honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.5

Differential Step Functioning Form

Consistency cause of DIF	Pervasiveness (Locus of cause of DIF)	
	Pervasive	Non-pervasive
Constant	All steps display a DSF effect equal in magnitude and sign	Only one / few steps display a DSF effect equal in magnitude sign across affected steps
Convergent	All steps display a DSF effect having the same sign but different magnitudes	Only a few steps display a DSF effect having the same sign but different magnitudes
Divergent	All steps display a DSF effects but the signs vary across steps	Only a few steps display a DSF effect and the signs vary across steps

Source: Penfield, Alvarez & Lee, in press, p. 29.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

CHAPTER FOUR

FINDINGS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

4.0 Introduction

This chapter presents the findings from the data analysis to answer the research questions of this study. The results reported on how the observed response patterns were used in studying the differences between honest and socially groups. Score deviations and fit statistics were applied in detecting the socially desirable responses. Factor analysis results were reported to examine the changes in factor structures of the IPIP for honest and socially desirable groups. In addition, the test information function and Differential Test Functioning (DTS) for overall test and each dimension of IPIP inventory were reported. Then, the differential item functioning analysis identified items that varied in meaning between the honest and socially desirable groups. Further analysis was performed for differential step functioning that may help to identify the characteristics of DIF.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

4.1 Comparing mean score of honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

4.1.1 Comparing the mean score deviations and differences between the honest and socially desirable groups.

Table 4.1 reports the mean and standard deviation descriptions for honest, socially desirable and combined data of honest and socially desirable groups. The socially desirable groups reported higher mean compared to the honest group. However, the standard deviation was the same across all three groups.

Table 4.1

Mean and standard deviation for honest, socially desirable and combined groups

	Group		
Description	Honest	Socially Desirable	Combined
Mean	162.1	163.3	162.7
Standard Deviation	18.18	18.18	18.19

The mean and standard deviations of the five dimensions of IPIP inventory are shown in Table 4.2. All five dimensions except for Agreeableness reported differences in the mean score between the honest and socially desirable groups. The mean values of Openness, Conscientiousness and Extraversion were higher in socially desirable groups compared to the honest group. However, the honest group reported higher mean values on the Neuroticism dimension than the socially desirable group. Whereas, there was no mean difference between honest and socially desirable groups for agreeableness dimension.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.2

Mean and standard deviation of five dimensions of IPIP inventory

Dimension	Description	Group	
		Honest	Socially Desirable
Openness	Mean	32.25	33.62
	SD	5.809	6.391
Conscientiousness	Mean	33.46	34.74
	SD	6.460	6.837
Extraversion	Mean	30.82	31.65
	SD	6.007	6.109
Agreeableness	Mean	35.37	35.37
	SD	5.992	6.238
Neuroticism	Mean	30.18	27.93
	SD	6.079	6.426

The score distribution of honest and socially desirable groups was studied by examining the groups score deviations from the mean. Therefore, the combined mean was set as a base and the dispersion of the scores from the mean was categorized according to the standard deviation intervals. Table 4.3 reports the distribution of scores according to the standard deviation cutoff values.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.3

Score distribution for the honest and socially desirable groups

Standard Deviation (Score range)	Group				Total
	Honest		Socially Desirable		
	n	%	n	%	
2 SD below mean (below 127)	18	66.70%	9	33.30%	27
1 SD below mean (128 – 145)	70	48.95%	73	51.05%	143
1 SD within mean (146 – 181)	365	50.56%	357	49.45%	722
1 SD above mean (182 – 199)	57	47.11%	64	52.89%	121
2 SD above mean (above 200)	11	37.93%	18	62.07%	29
Total	521		521		1042

The score distribution of socially desirable groups was found higher at one and two standard deviation above the mean. The percentage of score distribution for honest and socially desirable groups was narrower at one standard deviation above the mean. At 1SD above the mean, honest and socially desirable groups reported almost equal score distributions. The percentage of scores at 2SD above the mean indicated socially desirable groups were two times more than the honest group's. On the other hand, 66.7% of scores at two standard deviation below the mean was represented by the honest responses. The score distributions at 2SD below and above the mean clearly discriminate the dominance of groups in the score distribution.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

4.1.2 Distribution of non-fitting response of the Rasch model

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Table 4.4 lists the summary of non-fitting outfit values. The outfit values revealed individuals response pattern that do not conform to the Rasch model. Whereby, individuals observed scores from IPIP inventory differed than their estimated Rasch model score patterns. In total, 255 out of 1042 responses reported non-fitting response patterns. Majority of non-fitting responses were from socially desirable groups with 69.4%. Nonetheless, 83.5% of non-fitting responses fall within the outfit values of 1.3 to 2.0. Meantime, 78.6% of socially desirable group respondents reported to have outfit values of greater than 2.0 logits. Compared to the honest group, the socially desirable group has two times more of students with outfit values bigger than 2.0 logits. This shows distinctive dominance of non-fitting responses by the socially desirable groups in the region bigger than 2.0 logits. The results indicated high probability of non-fitting responses contributed by socially desirable groups. In conclusion, socially desirable group has the tendency to have outfit values of more than 2.0 logits.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.4

Summary of non-fitting responses according to the groups.

Outfit t-values	Group					
	Honest		Socially Desirable		Total	
	n	%	n	%	N	%
1.3 < t < 2.0	69	32.4	144	67.6	213	83.5
t > 2.0	9	21.4	33	78.6	42	16.5
Total	78	30.6	177	69.4	255	100

4.2 Identifying personality dimensions that are susceptible to social desirability

4.2.1 Factor Analysis

Confirmatory factor analysis (CFA) was applied to compare the factor structure of the IPIP for honest and socially desirable groups. A principal component analysis was conducted on 50 items with varimax rotation. The Kaiser-Meyer Olkin (KMO) measure of .855 for honest and .853 for socially desirable group verified the sampling as adequate for analysis. Barlett's test of sphericity for honest 6899.049 (1225) = .000, $p<.01$ and 6216.584 (1225) = .000, $p<.01$ for socially desirable group indicated that item correlation between items were significantly large for principal component analysis.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Next, Table 4.5 reported variance explained by each factor in the analysis for honest and socially desirable groups distinctively. The socially desirable group has higher percentage of variance for the first factor and a very small percentage of variance for the fifth factor. The percentage of variance was also slightly lower from factor 3 and 4 for the socially desirable groups.

Table 4.5

Total Variance Explained for honest and socially desirable group

	Rotation Sums of Squared Loading			
Component	% of variance		Cumulative %	
	Honest	Socially Desirable	Honest	Socially Desirable
1	9.57	10.33	9.57	10.33
2	7.69	7.69	17.25	18.01
3	7.38	6.31	24.63	24.32
4	6.21	6.03	30.84	30.35
5	5.47	3.86	36.31	34.21

Based on the hypothesized model of IPIP, five factors were extracted for both honest and socially desirable group's responses. The rotated component matrix reported the item loadings into five extracted factors. Items that load commonly for honest and socially desirable groups were bold for easier identification. Factor one for the socially desirable group reported the highest number of item loadings. There are 12 and 16 items of honest and socially desirable groups loading into factor one. Eight items were found to be loading commonly between the honest and socially desirable groups. The honest and socially desirable groups have an equal number of items loading into factor two. However, there were only two common items (item 8 and 28) shared by the honest and socially desirable groups in factor two.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

The honest group has reported 10 items loading into factor three. Eight out of the 10 items (item 29, 44, 34, 4, 49, 14, 24, 39) are items representing the Neuroticism dimension. The rest of the items represented Openness (item 10) and Extraversion (item 46) dimension. The socially desirable group has reported six items loading into factor three. There were five items representing the Extraversion dimension and one item (item 15) of Openness dimension. Furthermore, there was no common item spotted between the honest and socially desirable groups in factor three. A highest number of 15 items were found loading into factor four for the honest group. Whereas, socially desirable group reported 11 items loading into factor four. The honest and socially desirable groups shared five common items in factor four.

Result showed factor five for honest group is formed of six items representing the Extraversion dimension. The fifth factor for the socially desirable group showed loading of four items. However, there were only two (item 41 and 36) of Extraversion items found common between the groups. The review of factorial structure for the honest and socially desirable group showed differences in the item loading between the groups. The factor comparison between the honest and socially desirable group reported differences in the amount of loading, similarities in the item and the magnitude of item loading. Results showed honest and socially desirable groups did report differences in the representation of items and its magnitude of loading into a specific factor. Table 4.6 summarized the list comparing item loadings of each factor for the honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.6

Comparing item loading for honest and socially desirable group

Factor	Item No	Group	
		Honest	Socially Desirable
1	Item50	**.784**	**.556**
	Item25	**.754**	**.513**
	Item35	**.598**	**.551**
	Item5	**.443**	**.364**
	Item47	**.421**	**.461**
	Item45	**.399**	**.394**
	Item27	**.393**	**.452**
	Item37	**.337**	**.354**
	Item30	.537	
	Item7	.318	
	Item9	-.315	
	Item20	.313	
	Item48		.649
	Item33		.625
	Item 43		.592
	Item23		.532
	Item13		.517
	Item3		.513
	Item42		.493
	Item17		.456
2	Item8	**0.535**	**-.410**
	Item28	**0.444**	**-.455**
	Item43	0.648	
	Item48	0.614	
	Item18	0.491	
	Item23	0.478	
	Item3	0.468	
	Item38	0.465	
	Item13	0.464	
	Item33	0.457	
	Item12	0.384	
	Item34		.711
	Item29		.621
	Item44		.598
	Item4		.587
	Item39		.549
	Item14		.506
	Item24		.413
	Item10		-.383
	Item49		.357

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table 4.6 Continued

Factor	Item No	Group	
		Honest	Socially Desirable
3	Item29	0.627	
	Item44	0.622	
	Item34	0.622	
	Item4	0.621	
	Item49	0.527	
	Item14	0.502	
	Item24	0.495	
	Item39	0.493	
	Item10	-0.358	
	Item46	-0.344	
	Item21		.633
	Item31		.538
	Item1		.520
	Item15		.520
	Item11		.473
	Item26		.410
4	Item6	**0.529**	**.375**
	Item32	**0.504**	**.626**
	Item16	**0.495**	**.438**
	Item22	**0.49**	**.501**
	Item2	**0.431**	**.335**
	Item42	0.464	
	Item40	-0.409	
	Item17	0.375	
	Item26	0.374	
	Item38		.537
	Item30		.504
	Item12		.438
	Item18		.405
	Item20		.400
	Item7		.309
5	Item41	**0.606**	**.606**
	Item36	**0.511**	**.363**
	Item11	0.582	
	Item31	0.568	
	Item1	0.511	
	Item21	0.492	
	Item46		.570
	Item40		.385

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

4.2.2 Test Information Function

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Figures 4.1 to 4.6 illustrate the test information functions for the overall test and subsequently for five dimensions in IPIP. The test information function curve for the socially desirable responding was found to be narrowed yet provides more information on the test. In contrast, honest group responses were spread well but provided less information from the test to individual's ability.

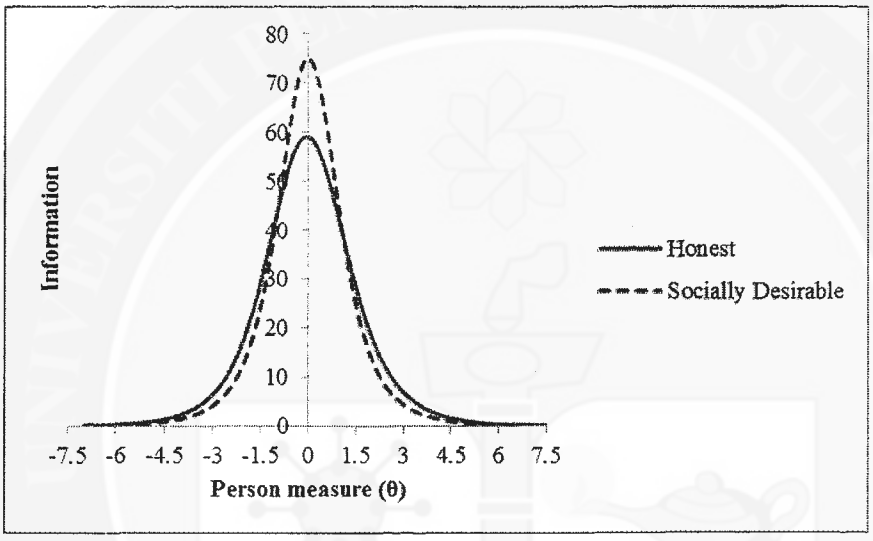


Figure 4.1

Test information function for honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

4.2.2 (a) Test Information Function for the five dimensions in IPIP.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

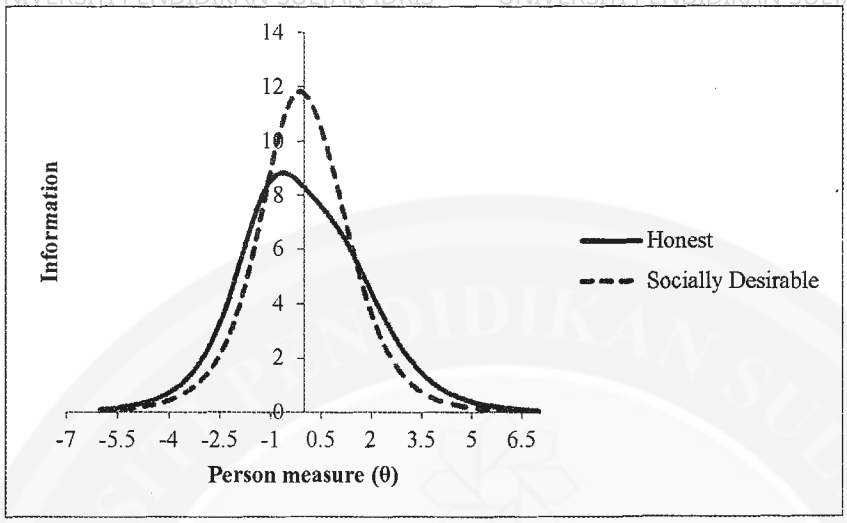


Figure 4.2

Test information function of Openness for honest and socially desirable groups

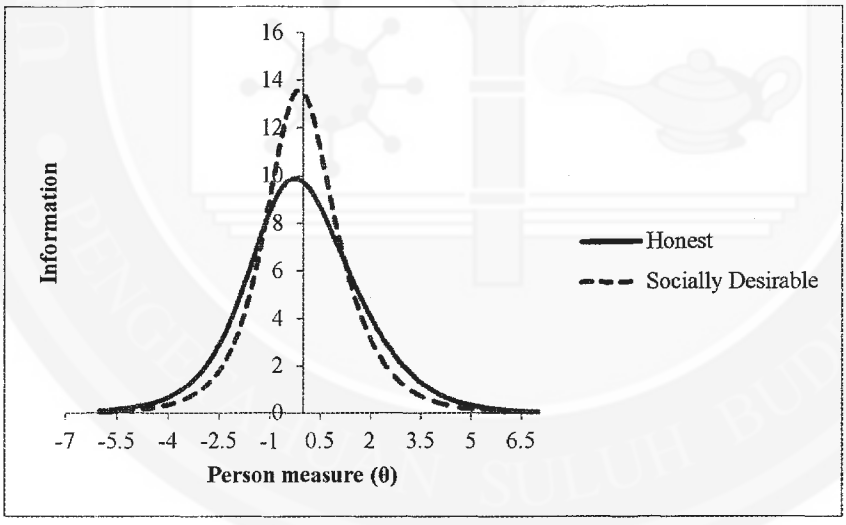


Figure 4.3

Test information function of Conscientiousness for honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

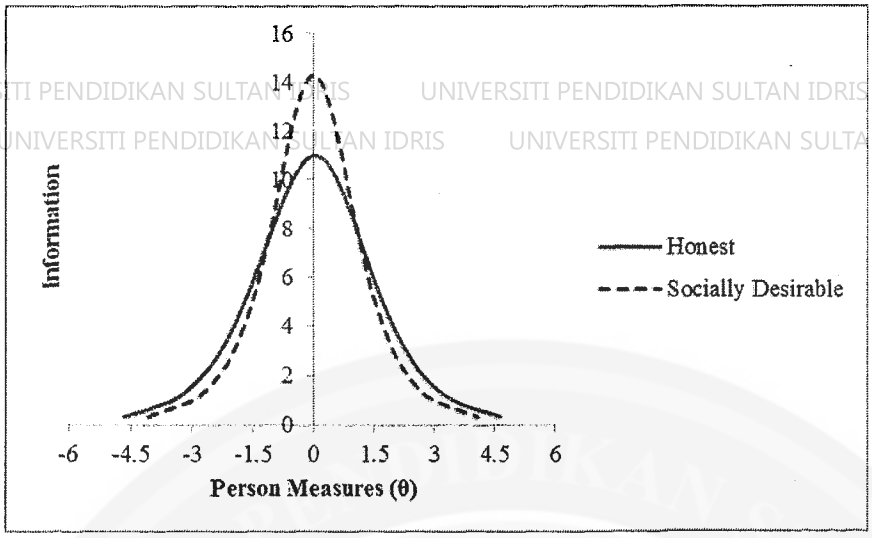


Figure 4.4

Test information function of Extraversion for honest and socially desirable groups.

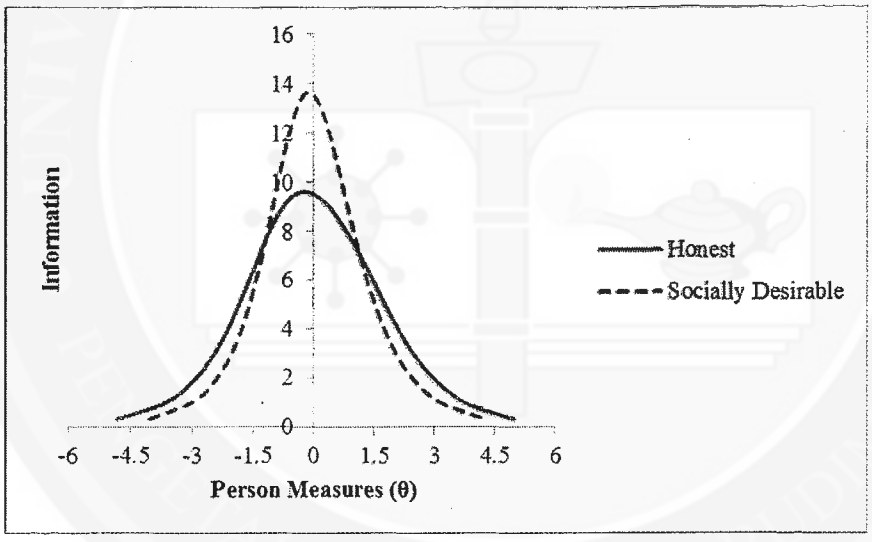


Figure 4.5

Test information function of Agreeableness for honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

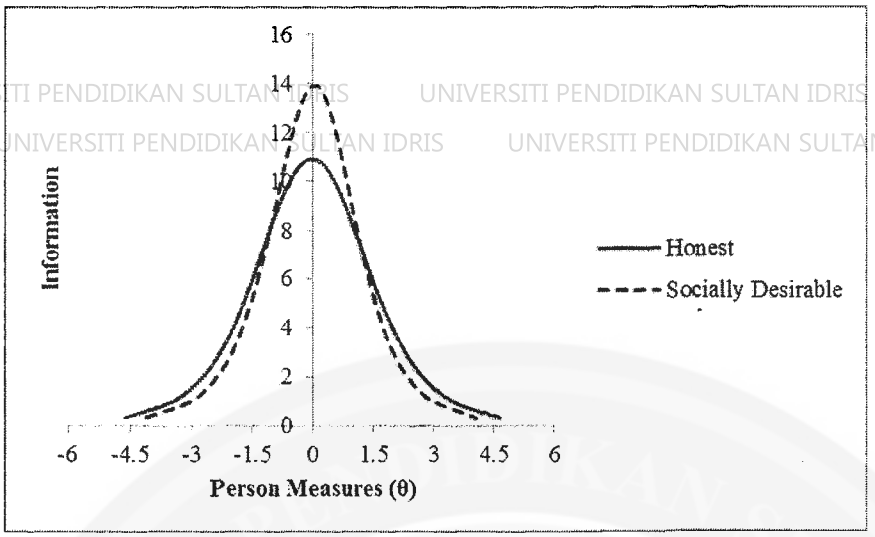


Figure 4.6

Test information function of Neuroticism for honest and socially desirable groups.

The figures for test information curves for all five dimensions in socially desirable groups showed monotonous distribution and most information about individuals was obtained at the ability level of .0. All five dimensions for socially desirable groups exhibited a narrowed curve. However, honest groups reported a wider yet non-monotonous distribution of person ability in the test. The Openness dimension for the honest group showed a negatively skewed distribution. This explains that, Openness items are easily agreeable and most of the lower ability individuals possess a high level of Openness characteristics. The differential information of Openness dimension between honest and socially desirable groups shows the impact of socially desirable responding instructions.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

4.3 Identifying items that are prone to socially desirable responding

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Table 4.7 explains the Mantel-Haenszel Chi-square values that indicated DIF items. The list of Mantel-Haenszel Chi-square for 50 items is reported in Appendix E. In total, the analysis has flagged 21 items as DIF. The analysis has reported effect size of the items at probability less than .01 for significant DIF. However, evaluation of Mantel-Haenszel Chi-square reported 42% of the test items functioning differently across the honest and socially desirable responses. Moreover, all 21 DIF flagged items were categorized as A which means DIF values below 0.43 and negligible. Mantel-Haenszel Chi-square seems to be overly sensitive in flagging DIF items. The test items was re-examined with Cummulative log Odd ratio method to avoid false identification of DIF. The cumulative log odds ratio reported only six item identified as DIF. These six items were also detected through Mantel-Haenszel Chi-square approach. Therefore, the six items were analyzed further for differential step functioning. In addition, Differential Test Functioning (DTF) was applied to examine the impact of DIF in the functioning of the test and personality dimensions between the honest and socially desirable groups.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

4.3.1 Differential Item Functioning with Mantel-Haenszel Chi-Square

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Table 4.7
List of Differential Item Functioning (DIF) with Mantel-Haenszel Chi-Square

	Mantel-Haenszel Chi-Square		
Item No	Effect Size	Sig	DIF Category
1	-0.21	.000	A
3	-0.30	.000	A
4	0.13	.000	A
5	-0.04	.000	A
6	-0.07	.001	A
13	-0.23	.001	A
14	0.07	.000	A
23	-0.24	.000	A
24	0.15	.000	A
25	-0.37	.000	A
28	-0.16	.001	A
29	0.16	.000	A
31	-0.01	.001	A
32	0.08	.005	A
33	-0.01	.001	A
39	0.14	.006	A
43	-0.27	.006	A
44	0.15	.004	A
47	-0.14	.002	A
49	0.15	.004	A
50	-0.33	.000	A

Note: sig= prob<.01

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

4.3.2 Differential Item Functioning using cumulative Log Odd Ratio

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The differential item functioning results from Table 4.8 reported only 6 items flagged as DIF out of the 50 items. He list of cumulative log odd ratio for 50 items is reported in Appendix E(i). Five items effect size was between $0.53 \leq \left| \hat{\alpha}_{LA} \right| < 0.74$ and one item $\hat{\alpha}_{LA} \left| \geq 0.74 \right.$. Five items, Item 4, 5, 23, 25 and 50 were reported to have a moderate magnitude of DIF and only item 14 shows large DIF magnitudes. Items 4 and 14 reported positive signs of effect size and the signs show it favours the honest group. On the other hand, items 5, 23, 25 and 50 reported negative sign which explained the items as in favour of the socially desirable groups. The item favouring between groups led to the examination of the item difficulty parameter of the six DIF flagged items across honest and socially desirable responses.

Table 4.9 illustrates the differences of difficulty parameter of the six items between honest and socially desirable groups. A complete list of 50 items difficulty parameter is attached in Appendix F. Thus, items 4 and 14 favour the honest group that has difficulty parameter value lower than the socially desirable groups. In contrast, items 5, 23, 25 and 50 that favour socially desirable groups has reported lower item difficulty parameter compared to the honest group.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.8

Differential item functioning from cumulative Log Odd Ratio

		DIF Tests of Significance				
Item	Effect size	Mantel	LOR Z	COX Z	DIF	Favouring
4	0.53	20.61	4.51	4.53	Moderate	HG
5	-0.56	21.99	-4.71	-4.69	Moderate	SDG
14	0.77	42.74	6.39	6.55	Large	HG
23	-0.53	19.17	-4.42	-4.37	Moderate	SDG
25	-0.58	23.39	-4.73	-4.84	Moderate	SDG
50	-0.56	20.70	-4.51	-4.55	Moderate	SDG

Note: HG=Honest Group, SDG=Socially Desirable Group

Table 4.9

Item difficulty parameter between honest and socially desirable groups

		Item Difficulty (logit)	
		Group	
Dimensions	Item	Honest	Socially Desirable
Neuroticism	4	0.13	0.38
Openness	5	0.48	-0.01
Neuroticism	14	-0.12	0.23
Conscientiousness	23	0.15	-0.13
Openness	25	-0.02	-0.21
Openness	50	0.14	-0.06

UNIVERSITI PENDIDIKAN SULTAN IDRIS

4.3.3 Differential Test Functioning (DTF)

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The differential test functioning results illustrates the impact of DIF items for the overall test and for each dimension in IPIP inventory. Results showed that DIF items have caused changes in variances of the overall IPIP measured between the honest and socially desirable groups. In addition, the Conscientiousness, Extraversion, Agreeableness and Neuroticism dimensions have reported small DIF effect variances for the honest and socially desirable groups. However, the Openness dimension has noted medium DIF effect variances in the measurement. Table 4.10 presents weighted estimates of DIF (nu-squared), standard error (SE) and the ratio of each DIF effect variance over respective standard error (Z).

Table 4.10

Results of Differential Test Functioning (DTF) for overall and all five dimensions of IPIP

Dimensions	Nu^2 Values	SE	Z	Interpretation
Openness	0.073	0.039	1.872	Medium effects
Conscientiousness	0.061	0.034	1.794	Small effects
Extraversion	0.051	0.029	1.759	Small effects
Agreeableness	0.037	0.023	1.609	Small effects
Neuroticism	0.041	0.039	1.64	Small effects
Overall Test	0.092	0.021	4.381	Medium effects

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

4.3.4 Differential Step Functioning

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Items that flagged DIF was further analyzed for Differential Step Functioning (DSF). The DSF analysis aids in examining the between group differences of the items at each step of a polytomous item. Table 4.11 reported the differential step functioning for flagged DIF items with DSF measures, magnitude and directions. Meanwhile, Table 4.12 summarizes the DSF magnitudes and its interpretations. Out of 6 items, items 4 and 14 demonstrated convergent pervasive DSF whereby all the steps in items 4 and 14 were affected by the differential responding instructions at the test administration. Thus, it can be concluded that item level effects cause items 4 and 14 function differently across honest and socially desirable manner. Meanwhile, items 5, 23, 25 and 50 were concluded as non-pervasive with affected step 3 and 4.

Therefore, this demonstrates the selection of higher response categories only and summarized as step level effects causing differential item functioning between the honest and socially desirable groups. The distribution of responses clearly exhibited the differences among the honest and socially desirable groups for higher category options. Table 4.13 reports the distribution of response selections for six DIF items across honest and socially desirable groups. The distribution of frequency shows that the socially desirable group has a higher frequency of "accurate" and "very accurate" category selections for items 5, 23, 25 and 50. This reflects to the DSF result that indicated effect of step 3 and 4 at the response categories for items 5, 23, 25 and 50. Whereby, the high response category causes the items to function differently between the honest and socially desirable groups. Items 4 and 14 showed differences in all categorical responses across honest and socially desirable effects. Hence, it can

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

be concluded that lower item difficulty and highly endorsed categories of items are prone to social desirability instructions.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Table 4.11

Differential Step Functioning Results for DIF items

Item	Step	CU-LOR	Z	Magnitude Of DSF Effect	Direction Of DSF Effect
4. Get stressed out easily	1	0.99	5.48	Large	+
	2	0.69	5.09	Large	+
	3	0.57	4.13	Moderate	+
	4	0.45	2.28	Moderate	+
5. Have a rich vocabulary	1	0.15	0.66	Small	+
	2	-0.19	-1.18	Small	-
	3	-0.54	-3.68	Moderate	-
	4	-2.18	-7.27	Large	-
14. Worry about things	1	1.43	6.27	Large	+
	2	1.05	6.57	Large	+
	3	0.94	6.69	Large	+
	4	0.56	3.15	Moderate	+
23. Get chores done right away	1	0.38	1.55	Small	+
	2	-0.29	-1.80	Small	-
	3	-0.73	-5.11	Large	-
	4	-1.07	-5.10	Large	-
25. Have excellent ideas	1	0.40	1.47	Small	+
	2	0.20	1.11	Small	+
	3	-0.89	-5.61	Large	-
	4	-1.47	-6.68	Large	-
50. Am full of ideas	1	0.14	0.54	Small	+
	2	-0.12	-0.67	Small	-
	3	-0.66	-4.29	Large	-
	4	-0.93	-5.19	Large	-

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 4.12

Summary of DIF items with DSF magnitudes and interpretations

Item	Step 1	Step 2	Step 3	Step 4	DSF Form	Interpretation
4	L+	L+	M+	M+	Convergent Pervasive	Item level effect
5	S	S	M-	L-	Non-pervasive convergent	Step level effect
14	L+	L+	L+	M+	Convergent Pervasive	Item level effect
23	S	S	L-	L-	Non-pervasive convergent	Step level effect
25	S	S	L-	L-	Non-pervasive convergent	Step level effect
50	S	S	L-	L-	Non-pervasive convergent	Step level Effect

Note: L+=Large positive, L-=Large negative, M+=Moderate positive, M-=Moderate negative, S=Small.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

Table 4.13

Frequency of response categories between honest and socially desirable groups

	Response Category									
	1		2		3		4		5	
Item	Honest	SD	Honest	SD	Honest	SD	Honest	SD	Honest	SD
4	60 (11.50%)	121 (23.20%)	117 (22.50%)	123 (23.60%)	136 (26.10%)	121 (23.20%)	134 (25.70%)	97 (18.60%)	74 (14.20%)	59 (11.30%)
5	56 (10.70%)	55 (10.60%)	99 (19%)	76 (14.60%)	260 (40.30%)	168 (32.2%)	138 (26.50%)	112 (21.50%)	18 (3.50%)	110 (21.10%)
14	37 (7.10%)	100 (19.20%)	73 (14.00%)	87 (16.70%)	132 (25.30%)	143 (27.40%)	172 (33.00%)	109 (20.90%)	107 (20.50%)	82 (15.70%)
23	35 (6.70%)	45 (8.60%)	111 (21.30%)	72 (13.80%)	199 (38.20%)	142 (27.30%)	134 (25.70%)	155 (29.80%)	42 (8.10%)	107 (20.50%)
25	36 (6.90%)	42 (8.10%)	81 (15.50%)	78 (15.50%)	222 (42.60%)	128 (24.60%)	141 (27.10%)	145 (27.80%)	41 (7.90%)	128 (24.60%)
50	38 (7.30%)	38 (7.30%)	78 (15.00%)	61 (11.70%)	184 (35.30%)	128 (24.60%)	142 (27.30%)	138 (26.50%)	79 (15.20%)	156 (29.90%)

Note: SD = Socially Desirable

UNIVERSITI PENDIDIKAN SULTAN IDRIS

CHAPTER FIVE

CONCLUSION, DISCUSSION, IMPLICATIONS AND RECOMMENDATIONS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

5.0 Introduction

This chapter comprises conclusion of this study's findings. Besides, this chapter also presents the discussion, implications and recommendations for future research.

5.1 Conclusion

The subtopics under conclusion explain the proposed research questions together with the obtained empirical results from the findings in answering the questions.

5.1.1 Research question 1

To what extent do the responses of group that distorted responses in socially desirable manner differ from the honest group?

Firstly, results have found differences in the mean score between the honest and socially desirable group. Overall mean score of social desirability group reported was slightly higher compared to the honest group. When examining the dimension level, Openness, Conscientiousness, Extraversion and Neuroticism dimensions revealed changes in the mean score between the honest and socially desirable groups. The high score obtainedby the honest group differentiate them from the socially desirable groups' responses.The score distribution between the honest and socially desirable

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

groups was found to be different. The difference in score between the groups was distinct especially at two standard deviations away from the mean. Thus, respondents of socially desirable group have higher scores with two standard deviation away from the mean. This is obvious as socially desirable responding individuals tend to inflate their scores that lead to the present of high score in the administered test. In addition, honest and socially desirable groups displayed differences in their observed and expected response patterns. The socially desirable group showed greater discrepancy in response patterns. This is supported by the results where the number of respondents in the socially desirable group with responses not-fitting the model for value greater than 2.0 logits was higher compared to the honest group. This indicates that socially desirable group responses exhibit greater inconsistency of responses from the expected response pattern. The cause of the non-fitting may be due to attempts made in score inflation and incoherence of response with the individual's ability. Therefore, high scores and non-fitting response are characteristics that can be useful informatin screening for the socially desirable responses.

5.1.2 Research Question 2

What are the characteristics of the items that are prone to social desirability?

In order to examine the items characteristics, items that were responded differently between the honest and socially desirable groups must be identified first. For the purpose of identifying the item, the Differential Item Functioning (DIF) methodologies were used to determine items functioning differently when the respondents were under honest and socially desirable instructions. Therefore, items

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

flagged as DIF in favor of the socially desirable group were said to be vulnerable to social desirability.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The findings showed that the DIF items were positively keyed items that have been functioning differently between the honest and socially desirable groups. Furthermore, the DIF items favoring socially desirable groups have lower difficulty parameter compared to the honest group. The items in Likert scale format seemed contributing for socially desirable respondents in the endorsement of response categories. Therefore, concludes that items with low difficulty and highly endorsed selective response categories as characteristics of the items that are vulnerable to social desirability.

5.1.3 Research Question 3

What are the personality dimensions that are prone to social desirability?

The instruction to respond in socially desirable manner affected the factor structure of the instrument. The same items, but for honest and socially desirable groups, were found to load on different factors and not as hypothesized in the International Personality Item Pool (IPIP) instrument. Although the instrument was tested with the same sample, yet when the same sample were told to respond in socially desirable manner, the item loadings changed and differ from when they were responding honestly. The result of the factor analysis revealed the influence of socially desirable responding resulted in changes in the factor structures.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Besides changes in factor structures, the information from all the personality dimensions was considerably more from the socially desirable groups than the honest group. This is especially the case for the Openness dimension. The presences of six DIF items in Openness dimension also suggested variances in the measurement between the honest and socially desirable groups. The DIF items from Openness dimension has directed to differences in the Openness dimension between the honest and socially desirable groups. Therefore, the number of DIF flagged items of Openness has caused differences in the dimension measured across the groups and was indicated as a dimension that is prone to social desirability.

Findings from the study reported the influence of socially desirable responding instructions to the individual's response styles in the personality inventory. The socially desirable responding instructions have found to raise the test score and distort item interpretations of personality items. Decisions made based on such test score will not be valid and accurate. Therefore, this study proposed high score and non-fitting responses pattern be used as evidence in detecting the socially desirable responses. Consequently, inflated response scores and the discrepancy between observed and expected responses as means of detecting socially desirable responding. The evaluation of items reported easy and highly endorsed extreme response category itemsas prone to social desirability. Besides that, study also reported selectiveness in the selection of items and dimensions for socially desirable responding in a provided situation. However, the conclusion made is to take into consideration of the fact that socially desirable responding is an integrated process of individual's motivation as well as the characteristics of the personality inventory.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

5.2 Discussion

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The main goal of this study was to detect individuals who responded to the personality inventory in a socially desirable manner. Thus, this study examined the characteristics of socially desirable group by comparing their scores to the honest group. This is determined by examining the dispersion of scores between the honest and socially desirable groups. Results showed the socially desirable group scores has deviated at least one standard deviation from the mean compared with the honest group. The two SD away from the mean score distinctively identifiedpotential individualswho inflate their scores to become higher than theiractual scores. This supports finding from previous research that found the socially desirable responding has resulted in score inflation and were reported as high scorers. Therefore, high scores in a personality test can becriteria to indicate socially desirable responding individuals. Similarly, other researchers showed faking individuals have elevated their scores on the tests (Morgeson et al., 2007). Identification of faking individuals is important as Hirsh and Peterson (2008) described such high scores on traits as over self-presentation and biased responses rather than providing accurate personality description.

In this study, a substantively high percentage of those with high scores with greater than two SD above the mean were the socially desirable responses.. As such, scores with deviation of greater than two SD above the mean is appropriately used to consider for initial screening of the socially desirable respondents. Similar results have been reported in a study on job applicants who obtained high percentage scores in a testing that dispersed at least two or three standard deviation above the mean (Rosse, Stecher, Miller & Levin, 1998).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Another piece of evidence in detecting the socially desirable responses was from the fit statistics. The fit statistics is an internal mechanism for identifying inappropriate response to the items (Green & Frantom, 2002). When individuals distort their responses in socially desirable manner, it was expected to be providing inconsistent response than their actual self. In motivating settings, the discrepancy between the true nature and observed scores are thought due to the intentional response distortions (Dilchert et al., 2006). Therefore, the inconsistency between individuals observed scores and expected scores of person were quantified and projected as the non-fitting responses. Findings from this study reported the majority of the non-fitting responses were from socially desirable groups. The most non-fitting responses from socially desirable groups showed that socially desirable instruction does affect the response patterns and resulted discrepancy between their observed and expected responses.

Two third of the responses with outfit values greater than 2.0 logits were from the socially desirable groups. The outfit values of more than 2.0 logits mean a greater inconsistency between observed and expected scores. Subsequently, individuals response pattern that do not conform to the model with outfit values greater than 2.0 logits are useful indicator to detect socially desirable responses in a personality inventory. However, one third of the non-fitting response patterns greater than 2.0 logits were from the honest group respondents. This method of identification may wrongly identify a socially responding response one third of the time. The results may also be attributed to the inability in controlling respondents to answer as they were told. This may be due to socially desirable responding is affected by the individual's ability and motivation in providing socially desirable responses (Magnus

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

et al., 2006). On the other hand, when respondents were instructed to respond in socially desirable manner, yet some continues to provide answer based on their actual characteristics and capability.

Differential Item Functioning (DIF) was applied in detecting items that are prone to socially desirability. Items are flagged as DIF when respondents from different group of the same ability have differences in their probabilistic endorsement of an item (Guilera, Benito & Hidalgo, 2009). In this study, DIF was used to examine items that have been responded differently by individual with same ability from honest and socially desirable groups. Thus, the DIF items were classified as items that are prone to social desirability. The results found all items that are prone to social desirability were positively keyed items. Literature has reported randomized distribution of items in a questionnaire so as to minimize the potential of response distortions. Although the items in IPIP was not arranged and presented according to their respective dimension, yet most of the flagged DIF items were from the Openness. This shows respondents ability in identifying evaluation at the test administration and the socially desirable way to respond it. Evidence showed applicants behavior adaptations accordingly as they were aware of the evaluation criteria (Konig, Melchers, Klieinmann, Richter & Klehe, 2006). This could be due to the non-contextualized item format. Whereby, non-contextualized opens room for varied interpretations of the items meaning.

The self-reported approach does allow individuals to interpret items in their own context and respond according to their needs. This is because, self-reported measures were given and ask to describe the extent to which items accurately describes their

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

personality that is assumed to be accurate responses (Lonnqvist, 2008). However, the mode of test administration influences the test presentation and the respondents' reactions (Houran & Lange, 2006). In order to obtain further information about the DIF items, examination of the differential step functioning (DSF) of the DIF items on the differences in the degree of response endorsement to provide more characteristics of the socially desirable groups. Individual usually distorts selected scores in order to promote a positive impression of themselves. This is supported by research findings from Martin, Bouen and Hunt (2002) that individuals distort responses toward their desired directions when using the Likert-scale format. The nature of the Likert-scale format provides opportunity to fake to the maximum positive or minimum negative behaviors (Graham et al., 2002). Thus, socially desirable group tend to endorse the high response categories compared to the honest group for the DIF items. They are capable of identifying items that promote positive traits and selectively endorsing responses leading to their purpose.

This study also examined personality constructs that are prone to social desirability. Findings from this study showed that socially desirable responding instructions had changed the factor structure of IPIP. The item loadings for the five dimensions were different between the honest and socially desirable groups. This could be due to the polarization of items that resulted from suppressing and enhancing of the desired personality items (Bradley & Hauenstein, 2006). Thus causes differences in the item loading and has directed to differential factor structures (Vanheule et al., 2007). In this study, the socially desirable group factor structure was different from the honest group. The results from confirmatory factor analysis (Cox, 2010) showed differences in the factor structures between the honest and faking groups. Moreover, study by

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Cox reported failure of faking group in extracting factors as hypothesized by the testing instrument. The difference in the factor structure reported the effects of social desirability in the item meaning and their interpretation. This supported that socially desirable responding has altered responses and resulting changes in the factor structures. This may be explained by responses in a demanding situations are projected from respondents operationalized interpretations of the situation and the desired responses (Jansen, Konig, Klienmann & Melchers (in press). Hence, lead to successful deliberation of responses that causes changes in the factor structures of the instrument which could be used as an indicator of socially desirable responding.

Literature has reported personality traits are differently biased with Conscientiousness and Emotional stability (Neuroticism) dimension more susceptible to socially desirable responding (Hirsh & Peterson, 2008). This study, however, did not show the Conscientiousness and Neuroticism dimensions as susceptible to response distortion by the Malaysian students. This is probably due to difference point of view and interpretation by the Malaysian children on the personality dimension based on the application settings. As mentioned by Thomas and Scroggins (2006), validity of personality tests depends on the inferences made about individual based on the test scores. This depends on the information available from the inventory to help in making the right decision. Henceforth, item information function and test information function were obtained in illustrating the information provided for each dimensions. Information obtained for Openness dimension was different between honest and socially desirable groups. In addition, the presence of DIF items in Openness dimension has also caused variances in the dimension measured between the groups and became vulnerable to social desirability. This

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

highlights the selectiveness of respondents in detecting promising items and distorting personality traits depending on the situational demands.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

5.3 Implications

The application of personality test becomes a concern when it is applied in a high-stake testing. This is due to the difficulty in differentiating respondent's true and socially desirable responses based on the score outputs. Findings from this study showed score inflation and high dispersion from the mean as an indication of socially desirable responses. Besides, the non-fitting response patterns to the Rasch model are found to be indicating the presences of socially desirable responses. This illustrates simple ways in detecting the socially desirable responses from a personality inventory. Therefore, the score dispersion and non-fitting response pattern can be implied in tests to screen out individuals who might respond in socially desirable manner. The detection of socially desirable responses would help in screening out applicants in the high-stake tests such as MEdSI and MUnSyI. Besides that, the changes in the factor structure, examination of items that are easy and the endorsement of extreme response categories help to identify items and personality dimensions that are prone to social desirability. The identification of easy items and items with highly endorsed extreme response categories indicates the items that are prone to social desirability. This becomes an advantage for the test users to exclude items that are prone to social desirability from the total test score. The application of the DIF and DTF carries in depth analysis at item and dimension levels. This extensive analysis scrutinizes the quality of items and dimensions measuring the personality attributes. Therefore, indirectly strengthens and improves the construct

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

validity of personality inventory. The elimination of socially desirable items from a personality test would provide more accurate description of the respondents. Literature reviews had mentioned person and personality inventory as the contributors of social desirability. Therefore, screening out the potential socially desirable responding individuals and the elimination of items and dimensions that are prone to social desirability would help to arrest the problem of social desirability in high-stake testing.

5.4 Recommendations

As this study was done for a Master level thesis, the sample size used was rather small and concentrated on two types of school. It is recommended that a similar study can be conducted with a much larger sample size to improve the generalization of the findings. The larger sample size will allow for the use of more sophistical statistical analyses. The large sample size will also enable establishing a norm that can be useful in detecting the socially desirable responses for small scale personality testing. As data from some of the personality testing in selection of candidates for specific purposes were not available for this study, the experimental design was adopted for this research. It will improve the validity of personality tests in selection purposes if real data under high-stake testing situations can be accessible to replicate this study. This can be helpful in accurate identification of the presence of social desirability effects in real life compared to the simulated settings. In addition to the methods used in this study, researchers interested in detecting socially desirable responses should compared the different ways of personality assessment instead of relying on self-reporting inventory to collect data. It is also recommended that,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

universities, teacher training colleges, employment and even school that use personality inventories should be more attentive to the results of high scoring and inconsistent response patterns. More steps should be taken to minimize the social desirability opportunities in the selection and administration of personality inventory. By doing so, candidate's scores and personality inventory are inspected and would apprehend the social desirability issues to enable more accurate predictions.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

REFERENCES

Abedi, J., Leon, S., & Kao, J. C. (2007). *Examining differential item functioning in reading assessments for students with disabilities*. Minneapolis: University of Minnesota Partnership for Accessible Reading Assessment.

Aiken, L.R. (1999). *Personality assessment methods and practices* (3rd ed.). United States of America: Hografe and Huber Publishers.

Aiken, L. R., & Marnat, G. G. (2006). *Psychological testing and assessment*. United States of America: Pearson.

Alex, C. (1975). *How to beat personality tests*. New York: Arco Publishing Company Inc.

Alliger, G. M., & Dwight, S. A. (2000). A Meta-analytic investigation of the susceptibility of integrity test to faking and coaching. *Educational and Psychological Measurement*, *60*(1), 59-72. doi: 10.1177/0013164004970367

Arifin Bin Hj Zainal, Asmawati Binti Desa, Hazalizah Binti Hamzah, & Nachiappan, S. (2009, November). *Hubungan antara personality dengan prestasi kerja guru dan implikasinya ke atas kesejahteraan warga sekolah*. Paper presented at the Persidangan Psikologi Malaysia, Selangor, Malaysia.

Arthur, W., Woehr, D. J., & Graziano, W. G. (2001). Personality testing in employment setting : Problems and issues in the application of typical selection practices. *Personnel Review*, *30*(6), 657-676. Retrieved from http://www.emerald-library.com/ft

Baker, F. (2001). *The basics of item response theory*. Retrieved from http://www.echo.edres.org.8080/irt/Baker/

Beck, M. S. L. (1994). *International handbook of quantitative applications in the social sciences* (Vol. 5). Singapore : Sage Publication

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Bell, R. C. (1982). Person fit and person reliability. Retrieved from http://www.rasch.org/erp8.htm

Beretvas, N. S., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scales. *Educational and Psychological Measurement*, *62*(4), 570-589

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317-335. Retrieved from http://www.olwnet.rice.edu/Birkeland%2oet%20al.%20(2006)% 20IJSA.pdf

Boer, D. P., Starkey, N. J., & Hodgetts, A. M. (2008). The California Psychological Inventory 434- and 260-item editions. In Boyle, G. J., Matthew, G., & Saklofske, D. H., *The sage handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2, pp. 97-112). Los Angeles: Sage.

Boyle, G. J., Matthews, G., Saklofske, D. H. (2008). *The sage handbook of personality testing and assessment, vol 2: personality measurement and testing*. Los Angeles: Sage.

Bradley, K. M., & Hauenstein, N. M. A. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science*, *3*, 313-335.

Brown, R. D., & Harvey, R. J. (2003). *Detecting personality test faking with appropriateness measurement: Fact/fantasy*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.

Brown, A. (2010). *Introducing concepts of measurement invariance: Investigating differential item functioning (DIF) using various approaches (Mantel-Haenszel and Confirmatory Factor Analysis (CFA) with covariates*). Retrieved from http://www.psychometrics.ppis.cam.../Summer%20School%202%20Day4.pdf

Burkevich, S. M., Jenkins, M., & Griffith, R. L. (2007). *Lying down on the job: Applicant faking and dependability*. Retrieved from http://www.cpla.fit.edu/io/documents/Burkevichetal2007.pdf

Carver, C. S., & Scheier, M. F. (2000). *Perspectives on personality* (4th ed.). United States of America: Allyn and Bacon.

Carrigan, M. (2007). Pre-employment testing-prediction of employee success and legal issues: A revisitation of Griggs v. Duke Power. *Journal of Business and Economic Research*, *5*(8), 35-44. Retrieved from http://www.cluteinstitute-onlinejournals.com/pdfs/172.pdf

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Cattell, H. E.P., & Mead, A. D. (2008). The sixteen Personality Factor Questionnaire (16PF). In Boyle, G. J., Matthew, G., & Saklofske, D. H. *The sage handbook of personality theory and assessment: Personality measurement and testing* (Vol.2, pp.135-159). Los Angeles: Sage.

Cervellione, K. L., Lee, Y. S., & Bonanno, G. A. (2009). Rasch modelling of the Self-Deception Scale of the Balanced Inventory of Desirable Responding (BIDR). *Educational and Psychological Measurement*, *69* (3), 438-458. doi: 10.1177/001316440

Charles, S. (2003). *Detecting intentional response distortions on measures of the 5 factor model of personality : An application of the differential person functioning*(Unpublished doctoral dissertation). University of Ohio, Ohio.

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). London: Continuum International Publishing Group

Comrey, A. L. (2008). The Comrey Personality Scales (CPS). In Boyle, G. J., Matthews, G., & Saklofske, D. H. *The sage handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2, pp. 113-134). Los Angels: Sage.

Cox, B. D. (2010). *Differential functioning by high and low impression management groups on a Big Five applicant and screening tool*(Unpublished doctoral dissertation). Auburn University, Alabama.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Citation classic*, *24*(18), 349-354.

Crowne, D. P. (1979). *The experimental study of personality*. New Jersey: Lawrance Erlbaum Associates Publishers.

Cruz, P., & Dipboye, R. L. (2003, April). *Justice in personality testing: Influence of outcome , modality and elaboration*. Poster presented at the 18th Annual Conference of the Society of Industrial and Organizational Psychology, Orlando, FL. Retrieved from http://rcoes.rice.edu/docs.cruz&dipoye2003.pdf

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal*, *5*(2), 125-144. Retrieved from http://www.iej.cjb.net

Day, N. T. (2008). *Item and person characteristics as predictors of faking*(Unpublished master thesis). Wright State University, Ohio. Retrieved from http://etd.ohiolink.edu/send-pdf.cgi/Day%20Nicholas.pdf?Wright1229968411

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Detrick, P., Chibnall, J. T., & Luebbert, M. C. (2004). The revised NEO-Personality Inventory as predictor of police academy performance. *Criminal Justice and Behavior*, *31*(6), 676-694. doi: 10.1177/0093854800268751

Detrick, P., & Chibnall, J. T. (2008). Positive response distortions by police officer applicants: Association of Paulhus Deception Scales with MMPI-2 and Inwald Personality Inventory Validity Scales. *Assessment*, *15*(1), 87-96. doi: 10.1177/1073191107306082

Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessment. *Psychology Science*, *48*(3), 209-225. Retrieved from http://www.moityca.com.br/pdfs/Response.pdf

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). *The validity of non-cognitive measures decays when applicants fake*. Retrieved from http://www.ipacweb.org/files/noncog.pdf

Drunmond, R. J., & Jones, K. D. (2006). *Assessment procedures for counsellors and helping professionals.* New Jersey: Pearson-Merrill Prentice Hall.

Dunn, D. S. (2009). *Research methods for social psychology*. United Kingdom: Wiley-Blackwell.

Edith, M. L. A., Stoop, V. K., & Meijer, R. R. (2000). *Computerized adaptive testing: Theory and practice*. Netherlands: Kluwer Academic Publishing.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability correction in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*(2), 155-166.

Elliot, S. (1997, November). *Some uses personality tests in military aviation training.* Paper presented at 45th Annual Conference of International Military Testing Association, Pensacola, FL.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. United States of America: Lawrence Erlbaum Associates, Inc.

Emotional signs. (2004, September 15). *Youth, Star Two*. Retrieved from http://www.leonard.com.my/publications.asp

Ferrando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person fit indices and detection scales. *Educational and Psychological Measurement*, *61*(6), 997-1012. Doi: 10.1177/00131640121971617

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Ferrando, P. J., Lorenzo, U., & Molina, G. (2001). An Item Response Theory: Analysis of response stability in personality measurement. *Applied Psychological Measurement*, *25*(1), 3-17.

Fitting the Bill. (2009, March 25). *Strait Times/ Asia News Network*. Retrieved from http://www.jobs.st701.com/career/.../indexx.php?c...Fitting –the-bill

Fox, J. P., & Meijer, R. R. (2008). Using item response theory to obtain individual information from randomized response data: An application using cheating data. *Applied Psychological Measurement*, *32*(8), 595-610. doi: 10.1177/0146621607312277

George, D., & Mallery, P. (1995). *SPSS/PC+ step by step: A simple guide and reference*. United States of America: Wedsworth Publishing Company.

Girard, M. L., & Scholtz, D. C. (2005). *Trial of psychological screening program to sniper selection in the Canadian Forces*. Retrieved from http:// www.internationalmta.org/Documents/2005/2005071p.pdf

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several Five-Factor Model. *Personality Psychology in Europe*, *7*, 7-28.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and scores inflation with faking: Do items attributes explain variances across items? *Journal of Business and Psychology*, *16*(4), 573-592. Retrieved from http://www.jstor.org/stable/250092794

Green, K., & Frantom, K. G. (2002, November). *Survey development and validation with the Rasch model*. Paper presented at International Conference on Questionnaire Development, Evaluation and Testing: Charleston, SC. Retrieved from http://www.jpsm.umd.edu/qdet/final_pdf_papers/green.pdf

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*(3), 341-355. doi: 10.1108/00483480710731310

Griffith, R. L., Peterson, M. H., Quist, J., Benda, A., & Evans, A. L. (2008). *Faking the personality profile: Easier said than done*. Paper presented at the 23rd Annual conference for the Society of Industrial and Organizational Psychology: San Francisco, California Retrieved from http://cpla.fit.edu/.../Griffth_Peterson_Quist_ Benda_Evan_2008_Final.pdf

Guilera, G., Benito, J. G., & Hidalgo, M. D. (2009). Scientific production of the Mantel-Haenszel Procedures as a way of detecting DIF. *Psicothema*, *21*(3), 492-498.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Guilford, J. P. (1954). *Psychometric methods*. (2nd ed.). United States of America: McGraw-Hill Book Company.

Hakstian, A. R., & Ng, E. L. (2005). Employment-related motivational distortions: Its nature, measurement and prediction. *Educational and Psychological Measurement*, *65*(3), 405–441. doi: 10.1177/0012164404267293

Harvey, R. J., Wilson, M. A., & Hansen, R. L. (2005). *Detecting CPI faking in police sample: A cautionary note*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology. Los Angles, CA. Retrieved from http://www.harvey.psyc.vt.edu/.../SIOP2005.handout.Harvey_ Wilson_Hansen.pdf

Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with a "Fake-Proof" measure of the Big Five. *Journal of Research in Personality*, *42*, 1323-1333. doi: 10.1016/j-jrp.2008.04.006

Holt, R. R. (1971). *Assessing personality*. United State of America: Harcout Brace Jovanovich, Inc.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, *30*(2), 161-172. doi: 10.1177/0146167203259930

Horst, P. (1968). *Psychological Measurement and Prediction*. United State of America: Brooks/Cole.

Houran, J. A., & Lange, R. (2006). *State of the art measurement in human resource assessment*. Retrieved from http://www.2020skills.com/asts/State _of_the_art_In_HR_Assessemr.pdf

Iddekinge, C. H. V., McFarland, L. A., & Raymark, P. H. (2007). Antecedents of impression management use and effectiveness in a structural interview. *Journal of Management*, *33*(5), 752-773. doi: 10.1177/0149206307305563

Issacson, J. A., Griffith, R. L., Kung, M. C., Lawrance, A., Wilson, K. A. (2008). *Liar, liar: Examining background checks and applicants who fail them*. Paper presented at the 23rd annual conference for the Society for Industrial and Organizational Psychology, San Francisco, CA.

Jackson, D. N., Wroblewski, V. R., & Asthon, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution. *Human Performance*, *13*(4), 371-388.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Jansen, A.,Konig, C. J., Kleinmann, M., & Melchers, K. (in press). The interactive effect of impression motivation and cognitive schema on self-presentation in a personality inventory. *Journal of Applied Social Psychology*. Retrieved from http://www.psychologie.uzh.ch/.../aopsy/.../jansenkonigkleinmanninpress.jasp.pdf

Kirkcaldy, B. (2001). Individuals ability to predict their own personality test scores. *Evaluation Journal of Australia*, *1*(2), 63-65.

Kline, P. (1976). *Psychological testing: The measurement of intelligence, ability and personality*. London: Malaby Press.

Konig, C. J., Melchers, K. G., Klieinmann, M., Richter, G. M,., & Klehe, U. C. (2006). The relationship between the ability to identify evaluation criteria and integrity test scores. *Psychology science*, *48*(3), 369-377.

Kroner, D. G., Mills. J. F., Yessine, A. K., & Hemmati, T. (2004). The generalized instructional set of Criminal Attribution Inventory (CRAI): Socially desirable responding and faking. *International Journal of Offender Therapy and Comparative Criminology*, *48*(3), 360-372. doi: 10.1177/0306624X04263452

Kumaresan, S., Aizat Mohd Nasurdin, & Ramayah, T. (2005). *Organizational and personalityinfluence on job stress: The case of Malaysian managers*. Retrieved from http://mgv.mim.edu.my/MMR/0412/041205.Htm

Lanyon, R. L., & Goodstein, L. D. (1982). *Personality Assessment* (2nd ed.). United States of America: Wiley-Interscience Publication.

Leite, W. L., & Beretvas, S. N.(2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement*, *65*(1), 140-154. doi: 10.1177/0013164404267285

Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR):A reliability generalization study. *Educational and Psychological Measurement*, *67*(3), 525-544. doi: 10.1177/0013164406292087

Lim, B. C., & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldberg's International Personality Item Pool: A multitrait-method examination. *Organizational Research Methods*, *9*(1), 29–54. doi: 10.1177/1094428105283193

Linacre, J. M. (2006). Winsteps (Version 3.65.0)[Computer Software], Chicago: Winsteps.com

Lonnqvist, J. E. (2008). *Issues in socially desirable responding and personality research*. Helsinki: Helsinki University Printing House.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
Magnus, J. M., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: the role of over-claiming, self-esteem and emotional intelligence. *Psychology Science*, 48(3), 336-356.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaire? *Personality and Individual Differences*, *32*, 247-256.

McCarthy, J. M., & Goffin, R. D. (2003). Is the test attitude survey psychometrically sound? *Education and Psychological Measurement*, *63*(3), 446-464. doi: 10.1177/0013164403063003007

Meade, A. W., Ellington, J. K., & Craig, S. B. (2004). *Exploratory measurement invariances: A new method based on Item Response Theory*. Paper presented at 19th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL. Retrieved from http://www.ncsu.edu/~awmeade/Links/Papers/ExploratoryDIF(SIOP).pdf

MEdSI. (2009). *Malaysian Educators Selection Inventory*. Retrieved from http://academik.upsi.edu.my/webportal.

Miller, T., Chahine, S., & Childs, R, A. (2010). Detecting differential item functioning and differential step functioning due to differences that should matter. *Practical Assessment, Research and Evaluation*, *15*(10), 1-13. Retrieved from http://pareonline.net/getvn.asp?v=15&n=10

Mitchelson, J. K., Wicher, E. W., LeBreton, K. M., & Craig, S. B. (2009). Gender and ethnicity differences on the Abridged Big Five Circumflex (AB5C) of personality traits. A differential item functioning analysis. *Educational and Psychological Measurement*, *69*(4), 613- 635. doi: 10.1177/001316440832325

Morgeson, F. R., Campion, M. A., Diboye, R. L., Holleabach, J. R., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection context. *Personnel Psychology*, 683-729.

Mueller, C. E., Bullock, E. E., & Leierer, S. J. (2010). *Examining psychometric and measurement properties of the Career Thought Inventory: Demonstration and use of the Rasch measurement model in career assessment research.* Retrieved from http://www.career.fsu/techcentre/TR51.pdf

Musson, D. M., Sandal, G. M., & Helmreich, R. L. (2004). Personality characteristics and trait clusters in final stage astronaut selection. *Aviation, Space and Environmental Medicine*, *75*(4), 342-349.

Nek Kamal Bin Yeop Yunus, (1997). *Polytechnic commerce students personality factors in relation to their preparedness towards workplace*(Unpublished Master thesis). University of Utara Malaysia, Malaysia.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Nunnally, J. C. (1975). *Introduction to statistics for psychology and education*. United States of America: McGraw-Hill series in Psychology.

Nye, C. D., Roberts, B. W., Saucier, G,. & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality*, *42*, 1524-1536. doi: 10.1016/j-jrp.2008.07.004

Oswald, F. L., Schmit, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as prediction of college students performance. *Journal of Applied Psychology*, *89*(2), 187-207. doi: 10.1037/0021-9010.89.2.187

Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York: Academic Press.

Penfield, R. D. (2007c). *Differential item functioning analysis system: User's manual*. Retrieved from http://www.education.miami.edu/facultysites/penfield/index.html

Penfield, R., Alvarez, K., & Lee, O. (in press). Using a taxonomy of differential step functioning form to improve the interpretation of DIF in polytomous items. *Applied Measurement in Education*. Retrieved from http://www.education.miami.edu/psell/.../_9_%20Penfield%20et%20al.pdf

Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics*, *6*(2). Retrieved from ttp://www.tc.columbia.edu/tesolalwebjournal

Pervin, L. A., & John, O. P. (2001). *Personality theory and research*. United States of America: John Wiley and Sons:

Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of future 1000 Firms. *North American Journal of Psychology*, *8*(3), 489-496.

Rennie, L. R. (1982). Detecting a response set to Likert-style attitude items with the rating model. *Educational Research and Perspectives*, *9*(1), 114-118.

Richman, W. L., Kiesler, S., Weisband, S., Drasgow, F. (1999). A Meta-analytic study of social desirability distortions in Computer-Administered Questionnaire, traditional questionnaire and interviews. *Journal of Applied Psychology*, *84*(5), 754-775.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. *Organizational Research Methods*, *3*(4), 348-365. doi: 10.1177/109442810034003

Robie, C., Tuzinski, K. A., & Bly, P. R. (2006). A survey of assessors beliefs and practices related to faking. *Journal of Managerial Psychology*, *21*(7), 669-681. doi: 10.1108/026839406110690204

Robie, C, Brown, D. J., & Beaty, J, C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, *21*(4), 489-509. doi: 10.1007/s10869-007-9038-9

Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (1991). *Measures of social psychological attitudes: Vol. 1*, San Diego, CA: Academic Press.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion in pre-employment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634-644.

Salgado, J. F. (2005). Personality and social desirability in organizational settings: Practical implications for work and organizational psychology. *Papeles del Psicologo*, *26*, 115-128. Retrieved from http://www.papelesdelpsicologo.es/emgish/1252.pdf

Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of person fit and effects of person fit on test validity. *Applied Psychological Measurement*, *23*(1), 41-53. doi:10.1177/01466219922031176

Sellbom, M., & Bagby, R.M. (2008). The validity and utility of the positive presentation management and negative presentation management scales for the Revised NEO Personality Inventory. *Assessment*, *15*(2), 165-176. doi:10.1177/1073191167310301

Sheppard, R., Han, K., Colarelli, S. M., Dai. G., & King, D. W. (2006). Differential item functioning by sex and race in the Hogan Personality Inventory. *Assessment*, *13*(4), 442-453. doi: 10.1177/1073191106289031

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). *Rasch fit statistics and sample size considerations for polytomous data*. Retrieved from http://www.biomedcentral.com/1471-2288/8/33/prepub.

Stabile, S. J. (1997). The use of personality tests as a hiring tool: Is the benefit worth the cost? *Journal of Labor and Employment Law*, *4*(2), 279-313.

Steffan, J. S., Kroner, D. G., & Morgan, R. D. (2007). Effect of symptom information and intelligence in dissimulation. *Assessment*, *14*(1), 22-34. doi: 10.1177/1073191106295404

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Stober, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity and relationship with age. *European Journal of Psychology Assessment*, *17*, 222-232.

Stocke, V., & Hunkler, C. (2007). Measures of desirability beliefs and their validity as indicators for socially desirable responding. *Field Methods*, *19*, 313-336. doi: 10.1177/1525822X07302102

Surat Pekeliling Perkhidmatan Awam. (2009). *Jabatan Perkhidmatan Awam*. Unpublished intranet document. Retrieved from http://www.jpa.gov.my

Suruhanjaya Perkhidmatan Awam. (2005). *Peperiksaan Khas Memasuki Perkhidmatan Awam Gred 41*. Retrieved from http://www.spa.gov.my

Syarat Kemasukan ke MRSM. (2010). Retrieved from http://www.mara.gov.my

Tan, L., & Grace, R. C. (2008). Social desirability and sexual offenders. A Review. *Journal of Research and Treatment*, *20*(1), 61-87. doi: 10.1177/1079063208314820

Thomas, S. L., Scroggins, W. A. (2006). Psychological testing in personnel selections: Contemporary issues in cognitive ability and personality testing. *Journal of Business Inquiry*, 28-38. Retrieved from http://www.uvu.edu.woodbury/jbi/vol5 /journals/jbi_5_thomas_pdf

Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). Testing for measurement invariances in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indication Research*, *78*, 341-360. Retrieved from http://www.faculty.ucr.edu/~sonja/papers/TOLB2006.pdf

Ujian Penetapan Khas. (2008). *Malaysian University Selection Inventory (MUnSYI)*. Retrieved from http://www.pohon.usm.my

Underhill, C. M., & Lords, A. O. (2002). *Fake resistance of forced-choice paired comparison personality measure*. Retrieved from http://www.internationalmta.org

Vanheule, S., Rosseel, Y., & Vlerick, P. (2007). The factorial validity and measurement invariances of the Maslach Burnout Inventory for human services. *Stress and Health*, *23*, 87-91. doi: 10.1002/smi.1124

Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis on fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*(2), 197-210.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Waller, N. G., Thompson, J. S., & Wenk, E. (2000). Using IRT to separate measurement bias from true scores group differences on homogenous and heterogeneous Scales: An illustration with MMPI. *Psychological Methods*, *5*(1), 125-146. doi: 10.1037/1082-989X.5.1.125

Welfel, E. R. (2002). *Ethics in counselling and psychotherapy: Standards, research and emerging issues* (2nd ed.). United States of America: Brooks/Cole.

Wolford, K. A. (2009). *Effects of items randomization and applicant instructions and distortion on personality measures*(Unpublished Master thesis). Bowling Green State University, Ohio. Retrieved from http://etd.ohiolink.edu/send-pdf.cgi/ wolford%20Katherine.pdf

Wrobel, N. H., & Lachar, D. (1998). Validity of self and parent report scales in screening students for behavioral and emotional problems in elementary school. *Psychology in Schools*, *35*(1), 17-27.

Yang, J., Bagby, R. M., & Ryder, A. G. (2000). Response style and the Revised NEO Personality Inventory: Validity scales and spousal ratings in a Chinese psychiatric sample. *Assessment*, *7*(4), 389-402. doi: 10.1177/107319110000700407

Younts, M. (2003). *The benefits of psychological testing for pre-employment of prospective firefighters in the High Point Fire Department*. Retrieved from http://www.usfa.dhs.gov/pdf/efop/efo35882.pdf

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent and experimental data sets: An application of mix model Item Response Theory. *Organizational Research Methods*, *7*(2), 168-190. doi: 10.1177.1094428104263674

Zickar, M. J., & Ury, K. L. (2002). Developing an interpretation of item parameters for personality items: Content correlates of parameter estimates. *Educational and Psychological Measurement*, *62*(1), 19-31. doi: 10.1177/0013164402062001002

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, *69*(4), 548-565. doi: 10.1177/0013164408324469

Ziegler, M., Toomela, A., & Buhner, M. (2009). A Reanalysis of Toomela: Spurious measurement error as cause for common variance between personality factor. *Psychology Science Quarterly*, *51*(1), 65-75.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

No: 59, Janal 1B/9
Antara Gapi, Serendah,
48200, Selangor.

Appendix A
Permission letter of IPIP: Malay version

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Pn Hazalizah Hamzah
Pensyarah
Fakulti Sains Kognitif dan Pembangunan Manusia
Universiti Pendidikan Sultan Idris (UPSI), Tanjong Malim,
35900, Perak.

12 Januari 2010

KEBENARAN MENGGUNAKAN *INTERNATIONAL PERSONALITY ITEM POOL* TERJEMAHAN VERSI BAHASA MELAYU

Puan,

Merujuk kepada perkara diatas, saya Priyalatha a/p Govindasamy, no k/p: 821002-01-5424 ingin memperoleh kebenaran meggunakan *International Personality Item Pool* (IPIP) yang diterjemahkan kepada Bahasa Melayu oleh puan.

2. Untuk pengetahuan puan, berikut adalah butir-butir diri saya.

Ijazah	: Sarjana Sastera (Pendidikan)
Bidang	: Psikometrik dan Penilaian Pendidikan (Penyelidikan)
Universiti	: Universiti Sains Malaysia
No Kad Matrik	: PPM 0005/09 (R)
Penyelia	: Dr Ong Saw Lan

3. Bagi tujuan penyelidikan, saya berhasrat menggunakan soal-selidik IPIP versi Bahasa Inggeris bersama-sama terjemahan Bahasa Melayu untuk mempermudahkan pemahaman responden terhadap item-item dalam borang soal-selidik IPIP.

4. Jika kebenaran diberikan, saya akan memastikan IPIP versi Bahasa Melayu hanya digunakan untuk tujuan penyelidikan sahaja. Malah saya juga berjanji membuat catatan bibliografi dengan sewajarnya.

5. Berikut disertakan ruangan untuk respon puan keatas permintaan kebenaran.Segala pertimbangan dan jasa baik puan amat dihargai serta didahului dengan ribuan terima kasih.

Sekian.

Yang benar,

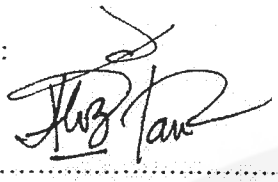
.....

(PRIYALATHA a/p GOVINDASAMY)

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Saya , dengan ini memberikan kebenaran kepada penulis menggunakan penterjemahan *International Personality Item Pool* (IPIP) Bahasa Melayu dengan syarat catatan bibliografi dilaksanakan dengan wajar.

Disahkan oleh:



.....

Nama : ..... Tarikh : 22/03/2010

HAZALIZAH BINTI HAMZAH
Pensyarah
Fakulti Sains Kognitif dan Pembangunan Manusia
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

IBU PEJABAT MARA
21, JALAN RAJA LAUT,
50609 KUALA LUMPUR.

Telefon : 03-2691 5377
03-2692 6011
03-2691 5111

Appendix B
Permission letter: MRSM

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Bil. MARA: BPM/G/22
Bil. Tuan:
Tarikh : 8 April 2010

Puan Priyalatha a/p Govindasamy
No. 59, Jalan 1B / 9
Antara Gapi, Serendah
48200, Selangor

KEBENARAN MENJALANKAN KAJIAN DI MAKTAB RENDAH SAINS MARA

Surat puan bertarikh 30 Mac 2010 tentang perkara di atas dirujuk.

2. Dimaklumkan bahawa permohonan puan untuk menjalankan kajian bertajuk *"Detecting Socialally Desirable Responses in Personality Questionnaire"* di MRSM Kuala Kubu Bharu diluluskan dengan syarat :

i. Memajukan Surat Akuan Mematuhi Syarat Kebenaran Menjalankan Kajian Di MRSM yang telah ditandatangani oleh penyelia / majikan kepada Pengarah BPL(M) sebelum puan memulakan kajian.
ii. Kerja-kerja berkaitan tidak akan menjejaskan kepentingan pelajar / staf MRSM.
iii. Satu salinan dapatan kajian hendaklah dihantar kepada Pengarah Bahagian Pendidikan & Latihan (Menengah).
iv. Pihak MRSM / MARA tidak bertanggungjawab atas apa-apa kos yang timbul / terbabit.
v. Kebenaran perlu diperoleh daripada Bahagian ini sekiranya dapatan kajian tersebut hendak dibentangkan di mana-mana forum atau seminar atau untuk diumumkan kepada media massa.

Sekian.

'BERKHIDMAT UNTUK NEGARA'
'CEMERLANG, GEMILANG, TERBILANG'
Yang benar,

SHAARI BIN HUSAIN
Pengarah
Bahagian Pendidikan & Latihan (Menengah)
b.p. Ketua Pengarah
MARA

s.k. Pengetua
Maktab Rendah Sains MARA Kuala Kubu Bharu

'KEUSAHAWAN KERJAYA PILIHAN'

MEMO_KAJIAN PRIYALATHA(SUPIAH 2010)F

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix C
List of modified Malay IPIP items

List of items modified from Hazalizah Hamzah's Malay version of IPIP

No	Original version	Hazalizah's version	Modified version
3.	Am always prepared	Saya mementingka nkesempurnaan dalam melakukan kerja	Saya sentiasa bersedia
6.	Don't talk a lot	Seorang yang pemalu	Tidak banyak bercakap
7.	Am interested in people	Suka mengutamakan orang lain	Saya berminat tentang orang
12.	Insult people	Suka mempertahankan pendapat saya	Mencemuh orang
13.	Pay attention to details	Memberi perhatian pada perkara remeh-temeh	Memberi perhatian terhadap sesuatu secara terperinci
18.	Make a mess of things	Kurang teliti dalam melakukan kerja	Jadikan benda-benda bersepah
19.	Seldom feel blue	Boleh mengawal perasaan	Jarang berasa muram
20.	Am not interested in other people's problems	Tidak suka bahan bacaan yang mencabar pemikiran	Saya tidak berminat dalam idea-idea abstrak
23.	Get chores done right away	Sukakan perkara yang teratur	Melaksanakan tugasan dengan serta-merta
26.	Have little to say	Seorang yang peramah	Tidak banyak untuk diperkatakan
29.	Get upset easily	Tidak cepat panic	Mudah tersinggung
30.	Do not have a good imagination	Suka melakuka nkerja-kerja yang rutin	Tiada daya imaginasi yang baik
31.	Talk to a lot of different people at parties	Sentiasa aktif	Bercakap pada ramai orang yang berlainan dalam majlis
32.	Am not really interested in others	Suka orang ikut caka psaya	Saya tidak sebenarnya berminat dengan orang
33.	Like order	Suka berkerja keras	Suka keadaan tersusun
34	Change	Mudah merasa sedih	Suasana hati saya kerap berubah

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Appendix C
List of modified Malay IPIP items

.	my mood a lot		
36.	Don't like to draw attention to myself	Suka berahsia	Tida kgemar dijadikan tumpuan orang ramai
38.	Shirk my duties	Seorang yang cuai dalam melakukan kerja	Elak dari tugas-tugas saya
39.	Have frequent mood swings	Seorang yang yakin pada diri sendiri	Kerap menghadapi perubahan suasana hati
42.	Feel other's emotions	Suka mempercayai orang lain	Memahami perasaan orang lain
44.	Get irritated easily	Cepat menggelabah	Mudah merasa keki
47.	Make people feel at ease	Boleh memahami perasaan orang lain	Membuat orang lain berasa selesa
48.	Am exacting in my work	Selalu cuba untuk mengurangkan tugas saya	Saya meminta-minta dalam kerjasaya
49.	Often feel blue	Boleh mengawal tekanan	Kerap berasa muram
50.	Am full of ideas	Kurang daya imaginasi	Saya penuh dengan idea

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
ORIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix D

Questionnaire: Cover page of honest group

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS



UNIVERSITI SAINS MALAYSIA



PUSAT PENGAJIAN ILMU PENDIDIKAN
SCHOOL OF EDUCATIONAL STUDIES

ID :_____

DETECTING SOCIALLY DESIRABLE RESPONSES IN PERSONALITY QUESTIONNAIRE

RESEARCHER:
PRIYALATHA D/O GOVINDASAMY
P-PM0005/09 (R)

SUPERVISOR:
DR. ONG SAW LAN

Dear Participant,

Thank you for participating in this research. You are requested to complete this personality inventory as honestly as you can. Please respond to the inventory by describing yourself as you are not how others see you or how you want others to see you. The results will be processed only for the research purposes and information gathered will be kept confidential.

KepadaPeserta,

Terima kasih atas kesudian anda menyertai kajian ini. Anda dikehendaki melengkapkan inventori personalit iini secara jujur. Sila jawab inventori ini mengikut pemahaman diri sendiri dan bukannya bagaimana anda dikenali atau ingin dikenali oleh orang lain. Data terkumpul hanya digunakan bagi tujuan kajian dengan kerahsiaan terjamin.

This questionnaire encompasses 2 sections. Section I on personal particulars of participant and Section II on personality items. Your cooperation is highly appreciated and thanked in advance. *Soal-selidik ini mengandungi 2 bahagian. Bahagian I merangkumi maklumat peribadi peserta dan bahagian II merangkumi item personaliti. Segala kerjasama anda amat dihargai dan didahului dengan ucapan terim kasih.*

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Appendix D(i)
Questionnaire: Socially desirable group

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F



UNIVERSITI SAINS MALAYSIA



PUSAT PENGAJIAN ILMU PENDIDIKAN
SCHOOL OF EDUCATIONAL STUDIES

ID :_____

DETECTING SOCIALLY DESIRABLE RESPONSES IN PERSONALITY QUESTIONNAIRE

RESEARCHER:
PRIYALATHA D/O GOVINDASAMY
P-PM0005/09 (R)

SUPERVISOR:
DR. ONG SAW LAN

Dear Participant

Thank you for participating in this research.
You are requested to complete this personality inventory **dishonestly** as you can. Please respond to the inventory by describing yourself as how others see you or how you want others to see you. Respond to the inventory in a way that you can portray the best personality characteristics. The results will be processed only for the research purposes and information gathered will be kept confidential.

KepadaPeserta,

Terima kasih atas kesudian anda menyertai kajian ini. Anda dikehendaki melengkapkan inventori personality ini secara ***tidak jujur****. Sila jawab inventori ini sepertimana and aingin dikenal ioleh orang lain. Anda dikehendaki menjawab dengan menggambarkan ciri-ciri personaliti terbaik. Data terkumpul hanya digunakan bagi tujuan kajian dengan kerahsiaan terjamin.*

This questionnaire encompasses 2 sections. Section I on personal particulars of participant and Section II on personality items. Your cooperation is highly appreciated and thanked in advance. *Soal-selidik ini mengandungi 2 bahagian. Bahagian I merangkumi maklumat peribadi peserta dan bahagian II merangkumi item personaliti. Segala kerjasama anda amat dihargai dan didahului dengan ucapan terima kasih.*

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Appendix D(i)

Questionnaire: Socially desirable group

Instruction : Please read the items carefully. Mark (√) at the respective boxes and fill in at the appropriate space given.

Arahan : Sila baca item dengan teliti. Tandakan (√) pada kotak yang sesuai dan isikan Maklumat di ruang yang disediakan.

1. Age / *Umur* : _____

2. Gender / *Jantina*
   : ☐ Male / Lelaki
   : ☐ Female / Perempuan

3. Ethnicity / *Bangsa*
   : ☐ Malay / Melayu
   : ☐ Chinese / Cina
   : ☐ Indian / India
   : ☐ Others / Lain-lain
   please state / silanyatakan _____

4. Year of study / *Tahunpengajian* : _____

5. Stream of study / *Jurusanpengajian* : _____

6. CGPA / *CGPA* : _____

Section II : Personality Inventory

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

BahagianII :Inventori Personaliti

Instruction : The inventory below lists fifty statements of personality. Please use the following scale to respond to each statement.

Arahan : Berikut merupakan senarai lima puluh pernyataan personaliti. Sila beri maklumbalas kepada setiap pernyataan di bawah mengikut skala yang diberi.

5 = Very Accurate	*Sangat setuju*
4 = Moderately Accurate	*Setuju*
3 = Neither Accurate Nor Inaccurate	*TidakPasti*
2 = Moderately Inaccurate	*Tidak Setuju*
1 = Very Inaccurate	*Sangat Tidak Setuju*

Please circle your choice. There is no right or wrong answers.
Sila bulatkan pilihan anda.Tidak ada jawapan yang betul ata pun salah.

No.	Items / *Item*	Scale / *Skala*				
1.	Am the life of the party. *Saya yang memeriahkan suasana sesuatu majlis.*	1	2	3	4	5
2.	Feel little concern for others. *Rasa kurang prihatin terhadap orang lain.*	1	2	3	4	5
3.	Am always prepared *Saya sentiasa bersedia.*	1	2	3	4	5
4.	Get stressed out easily. *Mudah tertekan.*	1	2	3	4	5
5.	Have a rich vocabulary. *Kaya dengan perbendaharaan kata.*	1	2	3	4	5
6.	Don't talk a lot. *Tidak banyak bercakap.*	1	2	3	4	5
7.	Am interested in people. *Saya bermina ttentang orang.*	1	2	3	4	5
8.	Leave my belongings around. *Meninggalkan barang-barang kepunyaan saya merata-rata.*	1	2	3	4	5
9.	Am relaxed most of the time. *Saya bertenang pada kebanyakan masa.*	1	2	3	4	5
10.	Have difficulty understanding abstract ideas. *Terdapat kesukaran memahami idea-idea abstrak.*	1	2	3	4	5
11.	Feel comfortable around people. *Berasa selesa berada dalam khalayak ramai.*	1	2	3	4	5
12.	Insult people. *Mencemuh orang.*	1	2	3	4	5
13.	Pay attention to details *Memberi perhatian terdahap sesuatu secara terperinci.*	1	2	3	4	5

Appendix D(i)
Questionnaire: Socially desirable group

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

5 = Very Accurate *Sangat setuju*
4 = Moderately Accurate *Setuju*
3 = Neither Accurate Nor Inaccurate *Tidak Pasti*
2 = Moderately Inaccurate *Tidak Setuju*
1 = Very Inaccurate *Sangat Tidak Setuju*

14.	Worry about things. *Risautentangsesuatu.*	1	2	3	4	5
15.	Have a vivid imagination *Mempunyai daya imaginasi yang tinggi.*	1	2	3	4	5
16.	Keep in the background. *Sukamengasingkandiri.*	1	2	3	4	5
17.	Sympathize with others' feelings. *Merasa simpati terhadap perasaan orang lain.*	1	2	3	4	5
18.	Make a mess of things. *Jadikanbenda-bendabersepah.*	1	2	3	4	5
19.	Seldom feel blue. *Jarang berasa muram.*	1	2	3	4	5
20.	Am not interested in abstract ideas. *Saya tidak berminat dalam idea-idea abstrak.*	1	2	3	4	5
21.	Start conversations. *Memulakan sesuatu perbualan.*	1	2	3	4	5
22.	Am not interested in other people's problems. *Saya tidak berminat dengan masalah orang lain.*	1	2	3	4	5
23.	Get chores done right away. *Melaksanakan tugasan dengan serta-merta.*	1	2	3	4	5
24.	Am easily disturbed. *Saya mudah merasa terganggu.*	1	2	3	4	5
25.	Have excellent ideas. *Mempunyai idea yang cemerlang.*	1	2	3	4	5
26.	Have little to say. *Tidak banyak untu kdiperkatakan.*	1	2	3	4	5
27.	Have a soft heart. *Seorang yang lembuthati.*	1	2	3	4	5
28.	Often forget to put things back in their proper place. *Sering terlupa meletakan kembali barang-barang di tempat sebenar.*	1	2	3	4	5
29.	Get upset easily. *Mudah tersinggung.*	1	2	3	4	5
30.	Do not have a good imagination. *Tiada daya imaginasi yang baik.*	1	2	3	4	5
31.	Talk to a lot of different people at parties. *Bercakap pada ramai orang yang berlainan dalam majlis.*	1	2	3	4	5

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix D(i)
Questionnaire: Socially desirable group

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

5 = Very Accurate	*Sangat setuju*
4 = Moderately Accurate	*Setuju*
3 = Neither Accurate Nor Inaccurate	*Tidak Pasti*
2 = Moderately Inaccurate	*Tidak Setuju*
1 = Very Inaccurate	*Sangat Tidak Setuju*

32.	Am not really interested in others. *Saya tidak sebenarnya berminat dengan orang.*	1	2	3	4	5
33.	Like order. *Suka keadaan tersusun.*	1	2	3	4	5
34.	Change my mood a lot. *Suasana hati saya kerap berubah.*	1	2	3	4	5
35.	Am quick to understand things. *Saya cepat memahami sesuatu perkara.*	1	2	3	4	5
36.	Don't like to draw attention to myself. *Tidak gemar dijadikan tumpuan orang ramai.*	1	2	3	4	5
37.	Take time out for others. *Meluangkan masa untuk orang lain.*	1	2	3	4	5
38.	Shirk my duties. *Elak dari tugas-tugas saya.*	1	2	3	4	5
39.	Have frequent mood swings. *Kerap menghadapi perubahan suasana hati.*	1	2	3	4	5
40.	Use difficult words. *Menggunakan perkataan yang sukar.*	1	2	3	4	5
41.	Don't mind being the center of attention. *Tidak kisah menjadi perhatian orang ramai.*	1	2	3	4	5
42.	Feel others' emotions. *Memahami perasaan orang lain.*	1	2	3	4	5
43.	Follow a schedule. *Mematuhi sesuatu jadual.*	1	2	3	4	5
44.	Get irritated easily. *Mudah merasa keki.*	1	2	3	4	5
45.	Spend time reflecting on things. *Meluangkan masa untuk mengamati pelbagai hal.*	1	2	3	4	5
46.	Am quiet around strangers *Saya berdiam ketika bersama orang yang tidak dikenali.*	1	2	3	4	5
47.	Make people feel at ease. *Membuat orang lain berasa selesa.*	1	2	3	4	5
48.	Am exacting in my work. *Saya meminta-minta dalam kerja saya.*	1	2	3	4	5

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

5 = Very Accurate *Sangat setuju*
4 = Moderately Accurate *Setuju*
3 = Neither Accurate Nor Inaccurate *Tidak Pasti*
2 = Moderately Inaccurate *Tidak Setuju*
1 = Very Inaccurate *Sangat Tidak Setuju*

49.	Often feel blue. *Kerap berasa muram.*	1	2	3	4	5
50.	Am full of ideas. *Saya penuh dengan idea.*	1	2	3	4	5

END / *TAMAT*

THANK YOU / *TERIMA KASIH*

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix E
DIF output: Mantel-Haenszel approach

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Item	Mantel-Haenszel Effect Size	Sig	Item	Mantel-Haenszel Effect Size	Sig
1	-0.21	.000	26	-0.01	.648
2	-0.17	.948	27	-0.10	.683
3	-0.30	.000	28	-0.16	.001
4	0.13	.000	29	0.16	.000
5	-0.04	.000	30	-0.01	.078
6	-0.07	.001	31	-0.01	.001
7	0.21	.121	32	0.08	.005
8	-0.09	.523	33	-0.01	.001
9	0.27	.016	34	-0.01	.010
10	-0.01	.038	35	-0.30	.016
11	-0.11	.083	36	0.01	.202
12	-0.19	.290	37	-0.14	.208
13	-0.23	.001	38	0.03	.359
14	0.07	.000	39	0.14	.006
15	-0.01	.844	40	0.14	.169
16	0.03	.028	41	0.03	.471
17	-0.02	.108	42	-0.24	.300
18	-0.07	.381	43	-0.27	.006
19	-0.07	.025	44	0.15	.004
20	0.03	.400	45	-0.02	.296
21	-0.11	.137	46	-0.02	.152
22	0.12	.021	47	-0.14	.002
23	-0.24	.000	48	-0.05	.301
24	0.15	.000	49	0.15	.004
25	-0.37	.000	50	-0.33	.000

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix E(i)

DIF output: Cumulative log odd ratio

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

		DIF Tests of Significance				
Item	Effect size	Mantel	LOR Z	COX Z	DIF Magnitude	Action
1	-0.40	11.86	-3.45	-3.44		
2	-0.01	0.00	-0.07	-0.07		
3	-0.50	17.35	-4.23	-4.16		
4	0.53	20.61	4.51	4.53	moderate	Inquiry
5	-0.56	21.99	-4.71	-4.69	moderate	Inquiry
6	0.38	10.38	3.22	3.22		
7	0.18	2.14	1.48	1.47		
8	-0.07	0.38	-0.62	-0.61		
9	0.28	5.65	2.33	2.37		
10	-0.25	4.32	-2.08	-2.08		
11	-0.19	2.68	-1.64	-1.63		
12	-0.13	1.17	-1.09	-1.08		
13	-0.41	10.83	-3.29	-3.29		
14	0.77	42.74	6.39	6.55	large	Inquiry
15	0.04	0.09	0.29	0.30		
16	0.25	4.61	2.13	2.14		
17	0.20	2.60	1.60	1.61		
18	-0.10	0.68	-0.83	-0.82		
19	-0.27	5.44	-2.32	-2.33		
20	0.11	0.73	0.85	0.85		
21	-0.17	1.96	-1.41	-1.40		
22	0.28	5.57	2.30	2.36		
23	-0.53	19.17	-4.42	-4.37	moderate	Inquiry
24	0.42	13.36	3.66	3.66		
25	-0.58	23.39	-4.73	-4.84	moderate	Inquiry

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix E(i)
DIF output: Cumulative log odd ratio

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

		DIF Tests of Significance				
Item	Effect size	Mantel	LOR Z	COX Z	DIF Magnitude	Direction of DIF
26	0.06	0.23	0.47	0.47		
27	0.06	0.21	0.46	0.46		
28	-0.39	11.17	-3.34	-3.35		
29	0.49	17.11	4.14	4.13		
30	0.22	3.31	1.82	1.82		
31	-0.38	10.35	-3.43	-3.23		
32	0.34	8.17	2.82	2.86		
33	0.31	6.15	2.45	2.48		
34	0.29	6.30	2.51	2.51		
35	-0.28	5.66	-2.37	-2.38		
36	-0.15	1.54	-1.24	-1.24		
37	-0.16	1.66	-1.31	-1.30		
38	0.12	1.05	1.03	1.03		
39	0.33	7.56	2.77	2.75		
40	-0.17	2.15	-1.47	-1.47		
41	-0.09	0.56	-0.75	-0.74		
42	0.12	0.89	0.95	0.94		
43	-0.33	7.47	-2.75	-2.73		
44	0.34	8.30	2.85	2.89		
45	0.13	1.09	1.04	1.05		
46	-0.16	1.78	-1.34	-1.34		
47	-0.39	9.73	-3.15	-3.12		
48	-0.13	1.14	-1.07	-1.06		
49	0.34	8.36	2.88	2.90		
50	-0.56	20.70	-4.51	-4.55	Moderate	Inquiry

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix F
List of item difficulties

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

ITEM	Group (logits)	
	Honest	Socially Desirable
1	0.23	-0.01
2	-0.28	-0.08
3	-0.06	-0.18
4	0.13	0.38
5	0.48	-0.01
6	-0.23	0.03
7	-0.15	-0.1
8	-0.11	-0.06
9	0.47	0.47
10	0.25	-0.03
11	0	-0.04
12	-0.23	-0.2
13	-0.16	-0.24
14	-0.12	0.23
15	-0.26	-0.3
16	-0.13	0.01
17	-0.56	-0.3
18	-0.15	-0.19
19	0.4	0.18
20	0.01	0.01
21	-0.05	-0.1
22	-0.19	-0.04
23	0.14	-0.06
24	0.03	0.28
25	0.15	-0.13

ITEM	Group (logits)	
	Honest	Socially Desirable
26	-0.08	0.02
27	-0.23	-0.15
28	0.19	-0.04
29	0.18	0.39
30	-0.3	-0.21
31	0.25	-0.03
32	-0.31	-0.09
33	-0.36	-0.24
34	0.05	0.17
35	-0.1	-0.19
36	0.36	0.18
37	-0.05	-0.16
38	-0.23	-0.08
39	0.11	0.21
40	0.44	0.31
41	0.26	0.19
42	-0.41	-0.2
43	0.14	-0.13
44	0.11	0.28
45	-0.17	-0.1
46	0.52	0.35
47	-0.18	-0.37
48	-0.2	-0.2
49	0.43	0.56
50	-0.02	-0.21

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN