

# THE UTILITY AND ACCURACY OF POST-CONVICTION POLYGRAPH TESTING WITH SEX OFFENDERS

By

Lars Bang Madsen

Master of Clinical Psychology, University of Queensland, 2000

Bachelor of Behavioural Science with Honours, Griffith University, 1995

Bachelor of Behavioural Science, Griffith University, 1994

Dissertation submitted in total fulfilment of the  
requirements of the degree Doctor of Philosophy

NEWCASTLE UNIVERSITY LIBRARY

-----  
204 26858 3  
-----

MED Thesis L8231  
-----

Department of Neurosciences & Psychiatry

University of Newcastle-upon-Tyne

May 2006

## ABSTRACT

The aims of the present research were two-fold: firstly, to investigate the utility of post-conviction polygraphy with community-based sex offenders; and secondly, to examine the accuracy of the polygraph in this context. The initial study examined whether periodic polygraph testing acted as a deterrent for engaging in risk behaviour. Fifty adult male sex offenders taking part in community treatment programs were allocated into 2 groups: “Polygraph Aware” subjects were told they would receive a polygraph examination in 3 months regarding their high-risk behaviours, while “Polygraph Unaware” subjects were told their behaviour would be reviewed in 3 months. Relevant behaviours for each subject were established at baseline interviews, following which both groups were polygraphed at 3 months. All subjects were polygraphed again at 6 months. Thirty-two subjects (64%) attended the first polygraph examination, with 31 (97%) disclosing an average of 2.45 high-risk behaviours each previously unknown to supervising probation officers. There was no significant difference between the two groups. Twenty-one subjects (42%) completed the second polygraph test, with 71% disclosing an average of 1.57 behaviours, a significant decrease compared with the first test. Disclosures to treatment providers and probation officers also increased. Polygraph testing resulted in offenders engaging in less high-risk behaviour, although the possibility that offenders fabricated reports of high-risk behaviours to satisfy examiners is also considered; similarly offenders seemed to be more honest with their supervisors, but this only occurred after the experience of the test itself. The second study examined the accuracy of the polygraph as used in a post-conviction context with sex offenders. One hundred and seventy-six sex offenders engaged in treatment and required to complete biannual polygraph tests focussed

upon offending and other risk behaviours. The participant's regular polygraph maintenance test was used for the study, however, in addition to the regular issues covered in this test the examiner included 'drug use' over the preceding three months as a relevant question. Immediately after the polygraph test a hair specimen was collected and subsequently analysed for drugs. The polygraph was reasonably accurate with identifying truth telling (79%), while 21% were wrongly accused of drug use. Only a small number of offenders ( $n = 5$ ) were found to be taking drugs and lying about having done so. The blind scorers correctly identified all of these individuals (100%). The Area under the curve index was .88. The inter-rater reliability between the blind scorers and the original examiners was poor. The original examiners were less accurate than the blind scorers (Area under the curve index = .68) and only correctly identified two of the five liars (40%). False positives were associated with lower intelligence and having experienced a sanction due to a polygraph result. False negatives were not associated with demographic characteristics, personality variables or intelligence. The majority of offenders found the polygraph to be helpful in both treatment and supervision. Nine per cent of offenders claimed to have made false disclosures; these individuals had higher scores on ratings of Neuroticism and lower scores on ratings of Conscientiousness. The implications of these results are discussed. Overall, the findings support the view that the polygraph is both useful and accurate in the treatment and supervision of sex offenders.



## ACKNOWLEDGEMENTS

I owe Don Grubin a great debt of gratitude for supervising this project, and also with having provided constant support, guidance and inspiration throughout its duration. A thank you is extended to Shaun Parsons for also supervising this thesis, and whose support and assistance with the initial study was helpful. I also wish to thank Brent Warberg, Dan Sosnowski, Lou Criscella and Shay Addison for their assistance with the polygraph testing and scoring.

I owe a debt of gratitude to Ian Stracken, Mark Farmer and Donald Finlader for their efforts in helping me implement the data collection process in the first study; and to Marilyn Roberts, Candice Osbourne, Gene Abel, Randy Fanning, Glenn Fraser, and Ron Hugley for assisting with the data collection for the second study. It is important to acknowledge the support from the NHS National Programme on Forensic Mental Health R&D.

A special thank you to Jane Eberhardt whose suggestions, corrections and general doggedness was very useful.

Finally, I wish to extend a special thank you to my wife, Shay Addison, without whose patience, love and wisdom this project would have taken much longer.



TABLE OF CONTENTS

List of Tables	1
List of Figures	3
CHAPTER 1	
Introduction	4
Study 1	9
Study 2	10
CHAPTER 2	
Lies and Deception	11
Lying	11
The Human Lie detector ?	16
Conclusions	20
Polygraph: History and Current Status	21
A brief history of lie detection	21
A brief history of the polygraph	24
Polygraph in the United Kingdom	41
The current status of Polygraphy	42
CHAPTER 3	
Polygraph Technique	45
Instrumentation	45
The pre-test interview	47
The examination process	49
Question Procedures	49
The Relevant Irrelevant Test	49
Comparison Question Tests	51
Computer Scoring	54

Concealed Information Tests	56
The Post-test interview	58
Conclusions	59

## CHAPTER 4

Sex offending & Relapse Prevention	60
Post-conviction polygraph testing	62
Pre versus Post Conviction Polygraphy	66
Post Conviction Sex Offender Testing Protocols	67
Research Evidence	73
Assessment of Sex Offenders	73
Monitoring of Sex Offenders	78
Factors affecting the polygraph's utility	79
Deterrence	79
The bogus pipeline effect	81
Confessions	81
False confessions	81
Personality	83
Conclusions	87

## CHAPTER 5

Science and the Polygraph	90
Theories of Lie Detection	95
Theories of the Comparison Question Techniques	97
Theories of the Concealed Information Test	102
Conclusions	103
Accuracy of the Polygraph	105
Research evidence from field studies	106
Research evidence from laboratory studies	111
A good accuracy study	115
Surveys of opinion	117
Unpublished evidence	118

Conclusions	120
Factors affecting polygraph accuracy	122
Countermeasures	122
Drugs & Alcohol	124
Expectations	126
Personality	129
Intelligence, Age & other Demographic Variables	131
Reliability	133
What needs to be done?	134
Study 1	134
Study 2	136
CHAPTER 6	
Study 1	
Method	
Participants	139
Polygraph Examination	139
Ethical Approval	141
Materials	
Personality	141
Personality disorder	144
Psychological distress	145
Risk	146
Polygraph Experience	148
Procedure	148
Session 1	149
Session 2 (Time 1)	152
Session 3 (Time 2)	154
Results	
Participants	156
Personality characteristics of sample	156
Personality disorder	160
Relationships between personality disorder and the FFM	160



Comparison of the FFM between PD and non-PD sex offenders	161
Psychological distress & Risk	162
Time 1(First polygraph)	
High risk behaviours reported	163
Factors associated with risk behaviours	166
Time of disclosure	168
Seriousness of behaviours	171
Personality, personality disorder and risk	172
Time 2 (Second polygraph)	
High-risk behaviours reported	174
Time of disclosure	176
Seriousness of behaviours	176
Self-report	179
Non-attendees	179
Main findings	180
 Study 2	
Method	
Overview of Study	181
Participants	181
The polygraph examination	183
Ethical approval	185
Materials	186
Risk and Personality	186
Intelligence	186
Psychological distress	187
Previous experiences with polygraph	188
Hair analysis	189
Procedure	190
Results	192
Characteristics of sample	193
Reliability	196

Numerical scores	196
Classifications	196
Accuracy	198
Receiver Operating Characteristic (ROC)	200
False positives	200
False positives (blind examiners)	202
False positives (original examiners)	206
Self Report	
Accuracy	209
False Positives (self-report)	210
False Negatives (self-report)	211
Utility	212
Sanctions	216
Countermeasures and false admissions	217
Main findings	218
CHAPTER 7	
Discussion	
The issue of utility	220
Risk & personality	224
Reasons for additional admissions	226
False confessions	227
Limitations & future directions	230
The issue of accuracy	234
Blind examiners	234
False positives & false negatives (blind examiners)	236
Base rates	240
Original examiners	244
Self-report	245
Reliability	246
Limitations of this research	248
Research recommendations	250
Conclusions & implications	254

Post-conviction polygraphy in the United Kingdom?	255
The issue of theory	255
Standardisation & monitoring	256
Inducing stress	261
The concern of countermeasures	261
Some recommendations for use of polygraph with sex offenders	262
Summary	266

CHAPTER 8

References	268
------------	-----

APPENDICES

A	Comparison Questions Techniques	322
B	Study 1 Information Sheets and Consent Form	327
C	Study 1 Sample Polygraph Report	331
D	Study 1 Questionnaire	333
E	Study 2 Information Sheet and Consent Form	336
F	Study 2 Questionnaire	341
G	The relationship between the five-factor model and personality disorder	347



## LIST OF TABLES

Table 1.	An example of the RIT question sequence	50
Table 2.	An example of DoDPI Zone test question sequence as used in a Sexual History Test	70
Table 3.	An example of an MGQT question sequence as used in a Maintenance Test	72
Table 4.	Diagnostic classification statistics	92
Table 5.	Behaviours rated as 'low', 'medium' and 'high' serious	151
Table 6.	Means, Standard deviation and ranges for the NEO-PI-R domain and facet scores	159
Table 7.	Pearson correlations between the FFM domains and the DSM-IV Personality Disorders	161
Table 8.	Number (%) of men reporting different high risk behaviours at Time 1 (n=32)	165
Table 9.	Categorical variables (high versus low) and risk behaviours	167
Table 10.	Comparison between conditions and seriousness of behaviour disclosed	171
Table 11.	Comparison between conditions and seriousness of disclosed behaviours excluding failed non-disclosing participants	173
Table 12.	Number (%) of men reporting different high risk behaviours at Time 2 (n=21 subjects)	175
Table 13.	Sample characteristics: intelligence, previous polygraph tests & DASS-42.	194
Table 14.	NEO mean scores, standard deviations and qualitative descriptions for overall sample study 2 (n=142)	195
Table 15.	Original Examiner Classifications Versus Blind Examiners Classifications	197
Table 16.	Accuracy rates for Original & Blind examiners	199

Table 17.	Sensitivity, Specificity, False negative & positive probability, Positive & Negative Predictive values for Original & blind examiners.	199
Table 18.	Area Under the ROC Curve and a 95% Confidence Interval Computed for the Original examiners and Blind re-scorers.	201
Table 19.	MANOVA means, standard deviations and number of variables used to examine differences between false positives and accurates for blind scorers	205
Table 20.	Logistic regression analysis of polygraph accuracy as a function of intelligence and having previously experienced a sanction (blind examiners)	205
Table 21.	MANOVA means, standard deviations and number of variables used to examine differences between false positives and accurates for the original examiners	208
Table 22.	Logistic regression analysis of polygraph accuracy as a function of extraversion (original examiners)	208
Table 23.	Self-reported accuracy rates for post-conviction tests	209
Table 24.	Means, standard deviations and number of variables used to examine differences between self-reported false and true positives for original examiners	210
Table 25.	Univariate analyses for false negatives and true positives	211
Table 26.	Univariate analyses for false negatives and true negatives	212
Table 27.	Behaviours less likely to be engaged in due to polygraph (n = 168)	212
Table 28.	The impact of base rates on the Positive & Negative Predictive Values at 90% & 95% accuracy	242
Table 29.	An example of the Reid CQT question sequence	325

LIST OF FIGURES

Figure 1	Example of polygraph charts	47
Figure 2	Study 1 design	155
Figure 3	Time of disclosure and polygraph results at Time 1	169
Figure 4	Time of disclosure and polygraph results at Time 2	178
Figure 5	Helpfulness of polygraph with avoiding risk behaviours and re-offences and with overall treatment? (%)	214



## CHAPTER 1

### INTRODUCTION

The polygraph, often referred to as a ‘lie detector’, has been widely used in the United States over the last 70 years, where it has typically been utilised in criminal investigations and for pre-employment and employee screening. More recently the polygraph has also been increasingly used in a post-conviction capacity in the treatment and supervision of sex offenders. Proponents claim that, within this context polygraph examinations assist with gaining additional information about an individual’s background and current behaviour. Numerous clinicians have endorsed the use of the instrument in this manner (Salter, 1997; Wilcox, 2000). The international Association for the Treatment of Sexual Abusers (ATSA, 1997) also recommends the use of the polygraph for validating an offender’s self-report. While many of these claims appear to have face validity, remarkably little research has actually been undertaken in the area and, the polygraph continues to be subject to considerable scientific and public controversy.

The polygraph – literally meaning “many writings” - was originally developed in the early 20<sup>th</sup> century. Whilst typically portrayed by popular media as a ‘magic’ mind reading machine, the device is actually an instrument that measures changes in physiological phenomena associated with arousal, generally sweating, cardiovascular and breathing responses. Practitioners typically do not claim that the polygraph measures deception directly, but assert that when an examinee is lying, the fear or stress of being detected and the related consequences will produce physiological changes associated with deception (Raskin & Honts,

2002). Proponents maintain that, when used properly, the polygraph is highly accurate. They emphasize that polygraphy has made valuable contributions to the wider community by resolving ‘countless’ criminal investigations, by uncovering spies, and by saving vast sums of money for businesses (Harrelson, Gerow & Gerow, 1998; Matte, 1996). Critics, however, vigorously assert that polygraphy is no more than an elaborate gimmick, claiming that it to be unreliable, invalid and prone to brand innocent persons as guilty (e.g. Cross & Saxe, 2001; Furedy, 1996a, 1996b; Lykken, 1998).

The research evidence for many of the claims made by both proponents and critics is generally limited at best, giving much of the debate a theoretical, and occasional vitriolic quality to it. Indeed, this lack of evidence appears at times, to have been confused with negative evidence against polygraphy.

It is also noteworthy that much of the on-going controversy has focused on applications of the polygraph, such as its use in criminal investigations, in employee screening and in security vetting. It is argued that these uses are distinct from the post-conviction polygraph testing with sex offenders (Holden, 2000). In the former, the issue of accuracy is particularly important, whilst in the latter, the focus is on its utility in eliciting disclosure of information (English, Jones, Patrick, Pasini-Hill, & Gonzalez, 2000). In this context, the polygraph is not considered a ‘test’ per se but rather a treatment tool that is used in conjunction with other tools (Chambers, 1994; Williams, 1995).

Advocates have argued in the therapeutic context that the polygraph enables clinicians to obtain more reliable sexual histories and more accurate offence descriptions from sex offenders. This assists the offender in overcoming denial and improves the clinician's assessment of treatment need and of the risk of re-offending (e.g. Abrams & Simmons, 2000; English, Jones, Pasini-Hill & Cooley-Towell, 2000). In terms of supervision, it assists not only with identifying breaches in supervision conditions but also with outright offences. In addition, it is claimed that it acts as a type of 'artificial conscience' by deterring offenders from engaging in problematic behaviour in the first place. Again, however, there is little empirical research with which to evaluate these claims, with much of the published literature being either theoretical or anecdotal in nature (Blasingham, 1998; Wilcox, 2000). What research that has been done has, however, tended to be generally supportive, although methodological problems, such as small numbers, retrospective methodologies and a lack of control groups, makes it difficult to disentangle the effects of therapy or supervision from those of the polygraph.

Ahlmeyer, Heil, Mc Kee and English (2000), for example, investigated the impact of polygraphy on admissions of offences and victims in adult sex offenders. Consistent with the claims made by proponents of polygraphy, the data showed increases in the number of reported offences and victims after a polygraph test for incarcerated and paroled sex offenders. Other studies have reported similar findings (e.g. Chambers, 1994; Emerick & Dutton, 1993; Hindman & Peters, 2001).



Abrams and Ogard (1986) studied the use of the polygraph in a supervision capacity, and compared the recidivism rates of probationers required to take periodic polygraph tests to those for supervision with probationers with no polygraph requirement. Although not specifically looking at sex offenders, they reported that over a two-year period, the majority of men who received periodic polygraph examinations remained offence free, whereas only a minority of those who did not receive polygraphs successfully completed their supervision. Whilst such a finding is promising, probationers were not randomly allocated into conditions and general criminality was not controlled for, and it is therefore unclear as to how much the polygraph actually contributed to a reduction of recidivism. The difference between the groups may simply be due to one group being less criminal than the other.

The proponents of post-conviction polygraphy have, nevertheless, made a persuasive clinical case for its introduction in the management of sex offenders. The research evidence supporting their claims is, however, limited. An additional concern regarding the argument presented by many of these advocates has been their tendency to ignore or dismiss unease within the wider scientific community regarding the polygraph's accuracy, while focusing instead on its potential utility. The utility argument becomes seriously compromised if the polygraph is not accurate in this context.

The accuracy of the polygraph is highly contentious within the literature. While a variety of studies have reported high accuracy levels (Honts, 1996; Patrick & Iacono, 1991a, 1991b), others have dismissed their findings by highlighting

methodological weaknesses in this research (Cross & Saxe, 2001; Furedy, 1996b; Lykken, 1998). At the core of much of this debate is the value of the different research methodologies used to investigate accuracy, these being field versus laboratory studies.

Field studies are commonly ‘real-life’ situations, such as a criminal investigation. In these studies, ‘accuracy’ is usually based on the outcome of an investigation or on the recording of a conviction. Critics of such research have asserted that the inability to assure criterion validity or ‘ground truth’ (i.e. knowing who is really lying) limit the value of the findings. In addition the retrospective nature of the methodology is likely to bias the sample used in the research. Laboratory studies are investigations in which field methods of polygraph examinations are used in simulated criminal situations. Such studies investigate either mock crimes set up by an experimenter with knowledge and collaboration of some subjects or actual small crimes induced by the experimenter (Kircher, Horowitz & Raskin, 1988; Patrick & Iacono, 1989). Criterion validity is assured because the researcher sets up the crime. Such studies are not, however, actual criminal investigations and subjects are usually aware that they are participants in research. They are often students who have little to lose by ‘failing’ a polygraph test. It is arguable, therefore, whether the results from these studies can be extrapolated to polygraph tests in real-life circumstances.

A recently published review of polygraphy by the National Academies of Sciences evaluated both of these types of studies. They concluded that under certain circumstances, the polygraph is likely to be accurate at levels greater than

chance (National Research Council, 2002), although this is likely to vary across different applications, and possibly also, with different populations. It is notable that all the research evaluated in the National Academies review focussed on the polygraph's use in pre-conviction contexts (i.e. employment screening and criminal investigations), the extent to which the results from these studies are applicable to post-conviction settings is also unclear.

The research reported in this thesis is designed to investigate the utility and the accuracy of post-conviction polygraph testing with sex offenders in the community.

The present investigation will comprise of two studies;

### *Study 1*

In the initial study polygraph testing incorporated in a community sex offender treatment program is examined to assess its contribution to treatment and supervision. This study is unique to others in the area because it incorporates a prospective design with comparison groups.

### *Study 2*

The challenge for researchers investigating the accuracy of the polygraph has been to develop a study where criterion validity is guaranteed, but where emotional environment is not compromised. The second study will utilise a unique

methodology to investigate the accuracy of post-conviction polygraph maintenance tests with adult sex offenders. By incorporating a question on drug use in the participant's regular maintenance test, and immediately afterwards by following-up with a drug-test, the objective drug test result will be compared to the result of the polygraph test. Finally, the study will examine the impact that personality and demographic factors may have on polygraph outcome.



## CHAPTER 2

### Lies and deception

#### *Lying*

Lying and deception has been a persistent feature of human behaviour from the time when Adam lied about a rather important apple to when President Clinton redefined 'sexual relations'. Numerous fairy tales, fables and stories, from Scandinavia to Africa, either celebrate the cunning liar (Homer's Odysseus) or warn of the impending doom for a lying scoundrel (Dante's Inferno). The occasional 'tall tale' has also been commonly used for purely entertainment purposes. Baron Munchhausen's documented exploits in combating the Turks in the mid 18<sup>th</sup> century provides an example, and undoubtedly, provided considerable amusement for its audiences at the time, as the earnest Baron recounted his experiences of flying on geese and leaping over seven-foot hedges (not an easy thing he asserted). Lying has, however, been generally viewed as an unacceptable and problematic behaviour in most circumstances.

The philosopher Emmanuel Kant considered all forms of lying to be inexcusable. Plato, on the other hand, thought that the occasional 'noble lie', which he defined as a deception that benefits the collective was justifiable. In contrast, the famous Chinese general-philosopher Sun Tzu in 600 BC considered deception to be a necessity to attain success in both diplomacy and war.

Humans do not have exclusive claim to deviousness. Animals can be just as crafty and sly. From lowly insects to the more evolved primates, animals have developed a variety of methods to deceive each other. Some of these are structural in nature, such as fauna masquerading as flora. Other creatures behave in threatening (though deceptive) ways to warn off potential predators. An example is provided by harmless sea snakes that impersonate the behaviour of their more dangerous cousins, the coral snakes. Still other animals practice more complicated forms of deceit by hiding, by pretending to be dead or by simulating injury to lure predators away from their nesting place (Ford, 1996). Attributing conscious intention to any of these behaviours would, of course, mean making a questionable anthropomorphic inference that this is comparable with lying in humans, which very much involves a specific intention to mislead another (Ekman, 1992).

Ekman (1992) defines lying as a deliberate choice to mislead a person or persons. This definition excludes self-deception, accidental misrepresentations, and psychosis, as the individual in such circumstances is not deliberately attempting to mislead another. Ekman (1992) emphasizes that for 'lying' to occur, the target person or persons must be unaware of the attempt to mislead, and also, must not have consented to be deceived. The audience in a theatre play or at a magician's show provides such an example (although Uri Geller is a liar because he claims that his tricks are real). Poker is another situation in which the rules notify the players that deception will occur, and bluffing cannot, therefore, be considered to be lying.

Broadly-speaking there are two primary ways of lying. These are ‘concealment’ and ‘falsification’. In concealment, the liar withholds information without actually saying anything untrue, while in falsification, the liar takes an additional step and presents false information as if it were true (McCann, 1998). An individual can falsify information by mixing truths with fiction, or by telling the truth with insincerity as in, “Yes! Of course I am having an affair with the neighbour”. Regardless of how a liar falsifies or presents information, the purpose of such behaviour is always to mislead (Wiley, 1998).

Lying is ubiquitous within the wider community. A survey study by Patterson and Kim (1991) found that 90% of people polled in North America admitted to having lied at some time in their life. This is perhaps not a surprising finding considering the social desirability problems often associated with self-report surveys. It may even be safe to assume that the actual percentage is higher (or as the joke goes, the other 10% are lying). The most commonly reported deceptions included lying about one’s feelings, income, accomplishments, sex life and age. Knox, Schact and Holt (1993) found in a sample of university students that deception within intimate relationships was also prevalent with 92% of subjects reporting that they had recently lied to a partner, or to a potential partner. In an interesting study, Kashy and Paulo (1996) asked students and non-students to keep a diary concerning their everyday lying behaviour. The results showed that, on average, people lied between three to four times per day. Students were the more frequent liars. In fact, in one-third of student interactions, some form of deceit occurred. In another survey study that specifically investigated adultery, Lawson (1988) estimated that between two-thirds to three-quarters of married



persons in North American and British samples had had an extramarital affair or 'one-night stand' and had lied to their partner about the liaison. Other studies have demonstrated that people typically lie when applying for work or promotions (Underwood, 1993). People also commonly use deceit as a strategy for enhancing collaboration, and for resolving conflict within a work place (Culbert & McDonough, 1992).

Deception also occurs on a larger and maybe less subtle scale in the advertising industry. A notable example was the "Vitamin O" scam exposed in 1999 in the US. This advertising campaign promoted the profound health benefits of "stabilised oxygen molecules in a solution of distilled water and sodium chloride" or more commonly known as salt water (Sullivan, 2001). Recently in the UK, the fast-food chain restaurant, Mc Donald's, was required to withdraw advertisements that implied that only fried potatoes were used in their French fries. The potatoes are actually fried in beef tallow and another secret 'natural ingredient' ([www.independent.co.uk](http://www.independent.co.uk)).

As these examples demonstrate, the motivation for lying encompasses a broad range of human aims. People lie for the obvious self-interest reasons, such as, to avoid punishment, or gain physical rewards (Bond & Atoum, 2000; Lanyon, 1997). Lying can also serve to enhance social relationships (Kashy & DePaulo, 1996). Consider, for example, the child who is told, "Make sure that you tell your grandmother how much you like the knitwear that she has made for you, so she'll feel good about having made it," when the garment has actually been the object of ridicule within the family. People lie to avoid embarrassment for themselves or



others. Lies can also take the form of pretending not to see, hear or know something that would be embarrassing. One example is pretending not to 'know' about the neighbour's problems with alcohol. Depending on the social circumstances that people find themselves in, the feigning unawareness of someone passing wind might be another example of this. Another type of deception generally deemed acceptable is described in German as 'nötluge'. This is a term meaning 'a lie of necessity'. This form of deceit 'takes care of the authoritarian thugs at your door' (Sullivan, 2001: p. 70) and refers to the lies told by individuals to protect others.

Some have claimed that the ability to lie serves an important developmental function by being a vehicle for separation and individuation in adolescents (Ford, King & Hollander, 1988). Others have argued that the ability to lie and be deceptive carries important evolutionary advantages with it. Whiten and Byrne (1988) highlight that human survival has depended upon the formation of shifting alliances and coalitions with others for the purposes of cooperating for meagre resources. It is suggested that the ability to successfully manipulate and deceive those seeking out the same resources, assists with maximizing an individual's chances of survival. In a related vein, Lykken (1998) suggests that the abundance of disparaging words and terms used to describe the naïve individual or least skilful 'lie-detector' (e.g. dupe, sucker, green-horn, mug, sitting duck, easy target) further proves our inherently predatory and mendacious nature. He reasons that if language can be seen as reflecting societies concerns, in the same way as the number of Eskimo words for snow reflects their preoccupation, then the number of synonyms for 'sucker' similarly confirms our devious character. Indeed, the

plethora of terms for ‘liar’ (e.g. fraudster, con artist, impostor, hoaxer, charlatan, swindler, etc.) perhaps provides further support for this theory. This type of social intelligence has sometimes been referred to as ‘Machiavellian intelligence’, after the 16th-century Italian writer who advised rulers to use deceit and cruelty to stay in power (Whiten & Bryne, 1988).

### *The human lie detector*

Curiously, there is no word or term that describes the unusually skilful human lie detector. The term ‘sceptic’, for instance, refers to the ‘inclination to doubt’ rather than an ability to detect deceit, whilst a ‘cynic’ is someone who doubts human merit and goodness (Oxford University Press, 1996). The lack of a specific term to describe the capable human lie detector possibly reflects the reality that humans are not particularly skilled ‘lie catchers’. This is, indeed, a conclusion that is generally borne out within the research literature.

In experimental settings, the ability of the average person to catch a liar is rarely above 60 percent (DePaulo, & Rosenthal, 1979; Vrij, 2000; Zuckerman, Spiegel, DePaulo, & Rosenthal, 1982). So-called professional ‘lie catchers’ tend to fare no better. Kraut and Poe (1980) found that customs officials were no more accurate than university students in detecting deceit in mock customs examinations. DePaulo and Pfeifer (1986) found no difference between federal law enforcement officers, regardless of experience, and students. Similarly, Kohnken (1987) showed police officers did no better than chance when they judged videotapes of college students who had lied or have been truthful in an

experiment. Though thought provoking, much of this research has questionable ecological validity. The typical experimental design, for example, involves recruiting students to lie or tell the truth about some relatively trivial issue. Some studies attempt to motivate the participants by emphasising the importance of lying, or by implying that only intelligent people succeed in the task. The participants are filmed, and this film is then shown to subjects who are asked to identify who is lying, and who is telling the truth. The observers for the most part have no vital interest at stake in achieving accuracy, nor are they offered any particular reward for being accurate.

A study by Ekman and O'Sullivan (1991) attempted to overcome these limitations by recruiting trainee nurses. In this research, some of the participants viewed a graphic film involving amputee and burn victims being admitted in an accident and emergency centre, whilst others watched a pleasant film about nature. Afterwards the participants were interviewed and asked to describe their feelings associated with the film they had just watched. All the participants were, however, told to pretend that they had watched a pleasant nature film and to conceal any negative emotions. Because the ability to control emotions in response to such graphic stimuli is an important and valued skill in nursing, Ekman and O'Sullivan reasoned that the trainee nurses would be highly motivated to succeed at the task. These interviews were filmed and then shown to individuals regularly involved in the detection of deceit. These included customs officials, police officers, court judges, secret service agents, polygraph examiners and other non-specified professionals. Once again, the results showed that these professionals, with the exception of secret service agents, were successful only slightly better than chance



at detecting deceit and no better than that achieved by college students. The findings did suggest, however, that the more accurate 'lie catchers' relied on non-verbal cues to identify deceit.

It is interesting that secret service agents were shown to be better than other professionals at detecting deceit. Ekman and O'Sullivan speculated that because such individuals were typically involved with scanning large groups of people for potential threat, they might be more attuned and skilled at evaluating non-verbal cues. It remains arguable, however, that lying about having passively watched an unpleasant film is comparable to lying about a specific behaviour. It would be valuable to utilise the same methodology but have one group actually attend an 'accident and emergency' ward, and then be required to lie about having done so.

Numerous studies have attempted to identify the specific verbal and non-verbal behaviours associated with lying. The behaviours traditionally thought to be associated with deception include gaze aversion, increased movement, longer and more frequent pauses, a slower speech rate and other more general disturbances (i.e. hesitations and speech errors) (Ekman, 1992; DePaulo, Lindsay, Malone, Muhlenbruck, Charlton & Cooper, 2003). Typically these behaviours are thought to be indicative of anxiety, and consistent with this belief, research has shown that liars tend to speak in a higher-pitched voice and display longer pauses during speaking. Somewhat counter-intuitively, however, liars move their arms, hands, fingers and legs less than truth-tellers (Vrij, 2000). Other behaviours such as gaze aversion, smiling, eye blinks or shifting position have not been associated with lying. As for verbal differences, liars tend to tell less plausible stories,



include more negative statements, and give more indirect and less detailed answers (DePaulo, et al., 2003). Typically, however, 'lie catchers' tend to use cues indicative of anxiety when evaluating truthfulness. This bias has been described as the *representativeness heuristic* phenomenon (Vrij, 2004). In other words, because people tend to believe that liars are more nervous than truth tellers, they infer deception from signs of nervousness in their subjects. Taken together, the research on objective and subjective cues to deception suggests that there is a mismatch between cues actually associated with deception, and cues that people associate with deception (e.g. nervousness). This difference may partly account for people's generally poor ability for detecting deceit.

In regards to this research, however, it is important to bear in mind that the utilised methodologies suffer from the same limitations as the previously discussed 'lie catching' studies. It is thus unclear whether these findings can be generalised to real-life contexts where the stakes are much higher for the liar.

It is interesting to speculate on the reasons why people are generally poor lie detectors. Again taking an evolutionary perspective one possibility could be that the consequences for group survival in the event of exposing of liar might be worse than the consequences of their lying. In other words, it is perhaps better to have someone contribute *something* to the group, rather than casting them out and thus be denied their contribution. Similarly, accusations of lying can have a significant effect on the functioning of a group, possibly leading to in-fighting and wider rifts among group members. Such damage could compromise the survival chances of the human group, particularly in harsh conditions.

## *Conclusions*

Lying and deception is prevalent within interpersonal relationships and the wider community. People lie for numerous reasons, not just for personal gain, but also to enhance relationships, and to protect others from harm and resolve conflict. Research has demonstrated that people are relatively poor at detecting deceit. This is also the case for professional lie catchers. Studies have also tended to show that there are no universal behavioural cues of deception.

## Polygraph: History and Current Status

### *A brief history of Lie detection*

Given that lying has been a constant feature of human behaviour since earliest times, it should not be surprising that detecting lies has also been a persistent interest for communities over the same period. Indeed, early societies developed a number of elaborate and creative procedures for identifying a liar. The most primitive lie detecting techniques typically involved torture or some form of ‘trial by ordeal’.

Versions of the trial by ordeal appear in ancient Greece, pre-Christian Scandinavia, Iceland, Polynesia, Japan and Africa (Segrave, 2004). The theory was that an omniscient higher power would rescue the innocent, while leaving the guilty to suffer the ordeal. For example, in the Middle Ages in Europe, an honest man was expected to be protected by God and so be able to hold his arm in boiling water for longer than a liar. The Bedouins of Arabia required conflicting witnesses to lick a hot iron, believing that the one whose tongue burned was lying (Larson, 1932). Similarly, in Scandinavia, if a woman was accused of adultery she was expected to ‘clear her self with the iron’, that is, hold a red-hot iron for a short time: if her hands burnt she was guilty of adultery. This hot iron test appears to have been a perennial favourite in charges of sexual misconduct in Scandinavia. Men were also subjected to it, although in their case, the charge brought against them had to be ‘carnal dealings with cattle of any sort’ (Sullivan, 2001).



Perhaps a more exotic ordeal was conducted in northwest Africa, where a fang from a snake was inserted under a suspect's eye-lid. If the person was truthful, he was expected to be able to eject it by rolling the eye. Another method of lie detection not based on the ordeal but on a type of 'sleight of hand' that is worthy of mention, was used in Israel in biblical times. In this procedure a donkey's tail was blackened with ink and placed in a darkened room. The suspect was then instructed to enter the room and pull the 'magic donkey's' tail, as it would identify the liar by braying. When the suspect returned, however, his or her hands were checked and the one whose hands were clean was identified as the 'liar' (Larson, 1932).

Another method commonly employed to determine truthfulness involved some form of combat (Lea, 1878/Larson, 1932). A particularly unusual example of such a procedure was used in Germany in the Middle Ages to settle allegations of infidelity. In this type of medieval marriage counselling, the husband was placed waist deep in the ground with a club in his right hand. The wife, who was not similarly restrained, wore a long-sleeved shirt with a rock sewn into the sleeve. The procedure involved the wife attempting to hit the restrained husband, whilst avoiding his attempts to hit her (Lea, 1878/Larson, 1932). Presumably the 'last one standing' was the faithful party. Perhaps the procedure served more to discourage such allegations from being made in the first place.

In other ordeals, the stress imposed on the individual was more psychological than physical in nature. In China, for instance, suspects were



required to chew rice powder and then spit it out. If the powder was dry, the suspect was considered to be guilty (Sullivan, 2001). A variation of this test was used during the Spanish Inquisition. The suspect had to swallow a slice of bread and cheese instead of rice. If it stuck in the suspect's palate, he was deemed to be lying (Lea, 1878/Larson, 1932). These latter lie detecting techniques implicitly reflect the belief that emotionality is associated with deception, as the outcome of both 'rice' and 'bread' tests are probably due to changes in the individual's saliva flow, which tends to decrease in response to stress. If the volume of saliva decreases then the suspect is likely to be frightened and, therefore, (in theory) lying. This explains why the rice appears dry when spat out, and why the bread sticks to the palate if the suspect is lying.

The belief that emotionality is associated with deception has been around for some time, as is illustrated by Daniel De foe's essay on the prevention of street crime in 1730 where he postulated, "guilt carries fear always about with it; there is a tremor in the blood of the thief". Indeed he suggested, "take hold of his wrists and feel his pulse... a fluttering heart, an unequal pulse, a sudden palpitation shall evidently confess he is the man, in spite of his bold countenance or false tongue " (De Foe, 1730/Larson, 1932).

In an early account of the use of pulse to determine truthfulness Trovillo (1939) described a story of a nobleman in the Middle Ages who was concerned about his wife's fidelity. The plans that the nobleman devised to uncover his wife's affair involved having an advisor sit next to her during the evening meal. When the name of the suspected paramour was mentioned he was quickly to take

her hand and measure the pulse. Perhaps unsurprisingly, there was a quickening of her pulse, though condemningly there was no similar response when her husband's name was mentioned. When confronted with this 'evidence' the wife allegedly confessed all. Modern lie detector theory and practice is based on similar reasoning and physiological notions.

More contemporary approaches make use of other physiological indicators of arousal in addition to pulse. These include breathing rate to determine truthfulness. In the late 19<sup>th</sup> century, Lombroso, an Italian criminologist probably better known for his theories on the atavistic criminal, was among the first to adapt the plethysmograph to monitor changes in blood volume during interrogation to infer a suspect's veracity (Larson, 1932). It is reported that, using this device, he accurately identified a suspect as innocent of stealing 20,000 francs, but guilty of stealing documents. He did this by observing a drop in blood pressure when the suspect was questioned about these documents (Trovillo, 1939). An Austrian, Benussi, was the first to investigate the relationship between lying and multiple physiological measures, such as blood pressure, pulse and breathing rate. From his research Benussi concluded that lying was accompanied with a change in the ratio of expiration to inspiration, the so-called 'Benussi Ratio' (Larson, 1932).

### *A brief history of the polygraph*

The American pioneer of modern lie detection was Hugo Munsterberg at Harvard University. Munsterberg published a book in 1908 (in which he pointedly avoided crediting any European scientists), proposing that research should be done

to investigate the physiological correlates of deception. A student of his, William Marston, devised the 'systolic blood pressure deception test' in 1915. He reportedly got the idea after his wife said that her blood pressure rose when she got mad or excited (Lamb, 2001). In Marston's lie test, an individual's blood pressure was measured intermittently during questioning using a standard blood pressure cuff and stethoscope. In 1917, Marston reported high positive correlations between lying and changes in systolic blood pressure. He boldly claimed at this time to have discovered the specific lie response, much like Pinocchio's nose, and predicted "the end of man's long, futile striving for a means of distinguishing truth-telling from deception" (Marston, 1938: p. 45). Marston is credited with coining the term 'lie detector' (Lykken, 1998). He also gained fame, under the pseudonym Charles Moulton, for being the creator of the first female comic strip hero, 'Wonder Woman' who was suppose to be a displaced Amazonian princess fighting the forces of evil and whose "magic lasso" forced all who were corralled within it to tell the truth ([www.wonderwoman-online.com/fc-marston.html](http://www.wonderwoman-online.com/fc-marston.html)).

Marston was an avid publicist of the lie detector. He famously appeared in a Gillette razor-blade advertisement in which his lie detector uncovered men's "true" feelings about various shaving aids. In the advertisement, Marston is seen analysing a polygraph tracing while a man is shaving, where it is explained that;

'Strapped to Lie Detectors, the same scientific instruments used by G-men and police throughout the country, hundreds of men take the part in an astounding series of tests that blast false claims and reveal the naked truth about razor blades. These men, shaving under the piercing eye of Dr.



William Moulton Marston, eminent psychologist and originator of the famous Lie Detector test, come from all walks of life, represent all types of beards and every kind of shaving problem. Knowing that the Lie Detector tells all... these men shave one side of the face with Gillette Blade, the other side with substitute brands.' (Saturday Evening Post, October 8, 1938, retrieved from [www.antipolygraph.org](http://www.antipolygraph.org)).

This 'Lie test' overwhelmingly discovered that Gillette blades were the preferred product. Marston also appeared in Look magazine highlighting the practical value of the lie detector in marriage guidance counselling. Presumably as a method of determining a wife's trustworthiness, he compared a wife's reaction to a kiss from her husband to that of an attractive stranger (Lykken, 1998; National Research Council, 2002).

Marston enthusiastically advocated for the use of his lie detector in criminal investigations. After the kidnapping of the aviator Charles Lindberg's baby, he contacted Colonel Lindberg to offer his services. Lindberg apparently did not reply, but undeterred, Marston approached the defence counsel of the man convicted of the kidnapping, Bruno Hauptmann, and offered to test him on death row. They too, however, declined, prompting Marston to lament, "the secret knowledge of the crime that Hauptmann had locked in his brain died with him" (Marston, 1938: p. 80).

In 1923, Marston was the first to attempt to submit the results of his lie detector test as evidence in a court case (Frye v. United States, 293 F.1013 [1924]).



According to Marston's (1938) account, James Frye, a 19-year-old man, had been accused of robbery and murder, but after initially denying the offences he confessed and provided the police with accurate details of the crimes. Frye withdrew the confession a few days later, claiming that he admitted to the crimes because he had been promised a share of the reward for his own conviction. Marston administered his lie test and concluded that Frye was telling the truth. The trial judge, however, refused to permit Marston to either testify about the examination or to conduct a re-examination in court. Frye was convicted, but the case was appealed on the grounds that the trial judge had erroneously excluded Marston's testimony. The initial decision was upheld on appeal on the grounds that Marston's lie detector test had not gained sufficient acceptance within the scientific community to be considered valid evidence. Ironically, Frye having been sentenced to life imprisonment was later exonerated and set free - Marston had been right all along (Office of Technology Assessment, 1983). This legal precedent became known as the 'Frye test' and remained a significant barrier to the inclusion of polygraph evidence in American courtrooms for the next 70 years. In 1979, the Kansas Supreme Court declared "The Frye test has been accepted as the standard in practically all of the courts of this country which have considered the question of admissibility of new scientific evidence" (Segrave, 2004: p. 17).

The Frye test was eventually displaced in 1993 by the U.S. Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals* (*Daubert v. Merrell Dow, Inc.*, 509 US. 579, [1993]). The Daubert ruling did not specifically address the polygraph. It did, however, provide general guidelines for determining the admissibility of contentious scientific evidence. This included considering the

known error rates, the reliability, and the general acceptance of the technique within the scientific community. These guidelines allowed courts to make decisions on admissibility of polygraph evidence on a case-by-case basis. As of 2003, polygraph evidence can be admissible in 19 states in the US.

Larson, a forensic psychiatrist, was influenced by Marston's claims and antics and commenced research into blood pressure and respiratory changes during police questioning. Larson is recognised as having created the first modern polygraph instrument in 1921. This was a device that made simultaneous recordings on paper of three physiological processes: blood pressure; pulse rate; and respiration. He reported a number of successes with this device, the most spectacular of which involved identifying a thief among 38 college girls living in the same house (Larson, 1932). Using a question technique that later became known as the Relevant-Irrelevant Test (RIT) he tested all the women in one day. The subjects were presented with a series of yes or no questions of which only some were related to the topic of investigation. The underlying reasoning was that the guilty person would experience an increased physiological response to the questions related to the crime, whilst in innocent examinees there would be little difference in the strength of reactions to either type of question. For example, a series of questions might be "Did you steal twenty dollars on Tuesday from the register?" (relevant) compared to "Is your name Lucy?" (irrelevant). One woman displayed this deceptive profile, and when confronted with the results promptly confessed to the theft (Larson, 1932; Matte, 1996).

Larson became sceptical of the accuracy and value of polygraph despite this and other successes. In one of the first published studies on polygraph chart evaluation, Larson (1938) had nine psychologists independently evaluate 62 polygraph charts, of which 61 were truthful. The number of charts classified as deceptive, however, ranged from 5 to 33. Larson concluded that due to such a high disagreement in the interpretation of the charts “the deception test alone should never be used as court evidence” (Larson, 1938; p. 896).

Leonarde Keeler, who was a protégé of Larson, developed the first portable polygraph instrument. This device could simultaneously record pulse rate, blood volume change and breathing. In 1939, he added the galvanic skin response channel to his lie detector and patented it. Keeler reasoned that a lying person would sweat more than a truthful person. This would then decrease skin resistance due to a higher concentration of negatively charged chloride ions on the surface of the skin. Keeler’s device became the prototype for the modern polygraph, and it was the first instrument purchased by the Federal Bureau of Investigation (FBI). Keeler is recognised by many as being the ‘father of modern polygraphy’ (Harrelson, Gerow & Gerow, 1998). In contrast to Larson who tried to develop standardised approaches to the polygraph interview, Keeler stressed the role of the polygraph as an interrogative device and advocated enhanced examiner discretion. Keeler is credited with developing the original polygraph testing formats: the previously mentioned Relevant-Irrelevant Test (even though Larson and Marston had used similar testing procedures); and the Peak of Tension Test (POT) or Concealed Information Test (CIT) (Matte, 1996). The POT is based on a different premise to the RIT. Instead of detecting deception about having committed a



crime, the POT's aim is to detect whether a suspect has information about the crime that only the guilty subject would have. During the test, subjects are asked a series of multiple-choice questions, each question having only one correct alternative. In a car theft case, for example, a suspect might be asked: "Regarding the car that was stolen do you know that it was: Yellow? Red? Blue? Black? White? Brown?" The theory is that the individual who committed the crime would know the correct alternative and display an augmented response to this critical item. An innocent suspect would, however, be unaware of the critical details and would show similar response patterns to all stimuli (Matte, 1996).

Keeler established the first polygraph unit within a police department in Chicago. This occurred primarily in response to the spiralling gang war between the gangs lead by Al Capone and George 'Bugs' Moran. At that time, public outrage caused by the execution style murder of seven men in an incident that became known as the St Valentine's Day Massacre, forced officials to employ new methods of 'fighting crime'. This included the use of the polygraph (Segrave, 2004).

In 1938, Keeler and another polygraph enthusiast, lawyer John Reid founded competing polygraph schools. Polygraph training proved a lucrative business for these initial schools. In 1942, the standard polygraph course available to police and military was two weeks long. Six years later the course had expanded to six weeks (Sullivan, 2001). Today the American Polygraph Association's (APA) accredited course is eight weeks in length.



Throughout the 1940's, polygraphy was increasingly becoming used by law enforcement agencies for criminal investigations and it was also taken up by various government agencies (Matte, 1996). In 1945, for instance, the US government experimented with using the polygraph to screen German POWs for the purpose of identifying good candidates to train for a police force to be installed in the post-war Germany. That effort, with Keeler as the chief polygraph examiner, entailed testing 274 POWs, with the aim of identifying undesirable affiliations, sympathies or intentions. Many Nazi members and sympathisers were apparently uncovered. It was also reported that much criminal activity was discovered, including plans to commit sabotage. For unreported reasons, however, this polygraph program was abandoned (Sullivan, 2001).

Keeler initiated another screening project at the Oak Ridge facility, where the atomic bomb was created (Linehan, 1990). This project reportedly identified a number of serious security breaches, including concealing from authorities the spillage of dangerous product material (Matte, 1996). There were also disclosures of tool thefts and unauthorised disclosures of classified information. The 'success' of polygraph testing resulted in a large number of employees being sacked. This program was, however, eventually discontinued in 1953 amidst accusations of coercion (Sullivan, 2001).

During the 1940's and early 50's the RIT technique was the standard polygraph questioning procedure. There was, however, an emerging awareness of the inadequacies of this question protocol. It was clear, for example, that some relevant questions such as "Did you stab and kill Janice (suspects mother) last

night?” could be just as distressing for innocent as for guilty subjects. Indeed, as in this example demonstrates, it is likely in some circumstances, to be even more stressful for innocent subjects. Reid (1947) therefore developed an alternative format that incorporated so-called ‘comparison questions’ that were not directly concerned with the crime under investigation but, unlike the irrelevant questions, were, nonetheless, calculated to induce an emotional reaction.

In Reid’s new procedure the comparison question was presented in a way that led the examinee to believe that it was relevant to the case at hand. For example, he or she may be told: “I need to ask you some other questions to determine whether you would be the ‘type’ of person that would lie about this (issue under investigation)”. This question was, however, general, non-specific and difficult to answer confidently with an unequivocal ‘no’. In regards to a specific theft, for instance, the examinee would be told that the comparison question “Have you ever stolen from someone who trusted you?” was relevant, when in fact it was not. Reid argued that such an awkward question would cause the examinee to feel uncomfortable, and lead him to withhold information about past thefts, even extremely minor ones, because of his fear that disclosing them would make him appear like the ‘type’ of person that would steal. Reid’s idea was to use the examinee’s physiological responses to the comparison questions as a standard against which to measure his reactions to the relevant ones which were specifically about the issue under investigation. The theory was that an innocent person would be more concerned with the comparison questions and less concerned with the relevant questions, because he would know that he was innocent and thus be more confident in answering these latter questions. It was

argued that an innocent 'non-deceptive' person would emit a larger physiological reaction to the comparison question when compared to the relevant, while it was expected that the converse would be the case for the guilty examinee (Reid, 1947). This procedure was named the Control Question Test, although its name was later changed to the Comparison Question Test (CQT) because it was argued that the 'control' question was not a control in the true scientific sense (Honts & Perry, 1992; Lykken, 1998).

Reid advocated that behavioural features and other information in addition to the polygraph charts should be used when determining the truthfulness of an examinee. Reid and Arther (1953) proposed that there were specific behaviours that were exhibited by liars within a polygraph examination. Some of these 'deceptive' behaviours included appearing nervous, reluctant, angry, having a 'gurgling' stomach or avoiding eye contact. Indeed describing oneself as religious, being over-friendly with the examiner or wishing to leave promptly after the exam were also identified as tell-tale signs of the liar by them. Both Reid and Keeler emphasised the clinical interpretation of the examinee's behaviour, and argued that the 'real lie detector' was the examiner guided by his or her training and background (Reid & Inbau, 1977). This form of test evaluation was, however, heavily criticised at the time for its lack of standardisation, and also because of the implication that there were specific behaviours or characteristics uniquely associated with deception (Backster, 1963a). In response to these problems, Cleve Backster (1963b), a student of both Keeler and Reid, developed the first numerical scoring system for evaluating polygraph charts.



Backster's (1963b) scoring system involved the systematic comparison of the reactions to each relevant question with the surrounding comparison questions. This process involved the assignment of numerical values on a 7-point scale ranging from -3 when the reaction to the relevant question was dramatically stronger, through to 0 when there was no difference, and through to +3 when the reaction to the comparison question was stronger.

Summing these scores over all components and all charts gives a total score (Backster, 1963b). If the global score exceeds +5 then the chart is typically considered to be 'no deception indicated' (NDI). If the score is less than -5 then the chart is classified as 'deception indicated' (DI). If the score is between -5 and +5 the chart is considered to be inconclusive. No other information is used when scoring the charts using Backster's system. Backster's innovation was important because it was the first time that a polygraph examination could lay claim to being a 'scientific' test, although the scientific basis of the scoring parameters themselves had still to be demonstrated.

By the beginning of the 1960's, polygraph use was widespread in the US while numerous other countries had started their own polygraph programmes, including Japan, China, Israel and Korea (Matte, 1996). At this time, US federal agencies were conducting approximately 19,000 polygraph examinations per year. The polygraph had also started to be used in the private sector (Office of Technology Assessment, 1983). Businesses, banks, fast-food chains and the like utilised the polygraph to vet job applicants not only to determine whether they were telling the truth on their applications, but also to determine whether they were



the type of person worth employing. For example, the polygraph was used to examine whether applicants had ever stolen from previous employers, had used drugs, or engaged in other types of criminal or immoral behaviour (Segrave, 2004).

During this time the polygraph was also introduced to periodically screen employees to determine whether they were stealing from the company or engaging in any other form of criminal behaviour (Lykken, 1998). In 1965 a proposal to use polygraphy in a similar manner with Federal employees prompted the Committee on Government Operation to conduct the first evaluation of polygraphy. Their report concluded that there was no scientific evidence to support this type of application, and that the research evidence for its accuracy was inadequate. The report also raised concerns about the ethics of using the polygraph for screening. It concluded, “there is no lie detector, neither man nor machine. People have been deceived by a myth that a metal box in the hands of an investigator can detect truth or falsehood” (US Congress, 1965: p. 1). In response to these conclusions and concerns President Kennedy indicated, “it was a mistake to suggest the polygraph” (Brooks, 1985: p. 348). In succeeding years a number of bills were proposed to restrict or ban polygraph use by federal agencies, although none were enacted (Office of Technology Assessment, 1983).

By the 1970’s employee screening had become a multi-million dollar business, with approximately two million people each year in the US being required to take a polygraph to either apply for, or to keep a job in the private sector (Lykken, 1979). The pre-employment screening program used by the Coors brewery in Colorado provides an example of how the polygraph could be misused

in such contexts. Company director William Coors is quoted as saying that the polygraph helped ensure “that the applicant does not want the job for some subversive reason such as sabotaging our operation” (Segrave, 2004: p. 122). But some of the actual questions used on the tests included: “What are your sexual preferences?”, “How often do you change your underwear?”, “Have you ever done anything with your wife that could be considered immoral?” and “Are you a homosexual?” (Segrave, 2004). A contrasting example of its successful use in an investigative capacity occurred in 1976, and has been described by Lykken (1998). A Californian bakery company were facing ruin because their products was being returned with glass and other materials in it. After an exhaustive investigation into the manufacturing process, it was concluded that an employee within the factory was deliberately contaminating its products. The police were unable to identify the culprit, and so the company decided to polygraph its employees. Apparently the eighth person tested confessed to the sabotage, having been angered after being overlooked for promotion (Lykken, 1998).

During the 1970’s, the majority of law enforcement agencies established a staff of polygraph examiners, although federal government use of the polygraph had decreased dramatically to approximately 7000 tests per year (Office of Technology Assessment, 1983). Also at this time, two Judges, unbeknownst to each other, informally initiated post-conviction polygraph testing. It was believed that aside from the polygraph’s ability to identify deception, it could also have considerable value in the supervision and management of probationers (Holden, 2000). It was argued that periodic polygraph tests would act as a deterrent for

engaging in criminal activity, in addition to also being an effective method of identifying when re-offences occur.

Teuscher (1978), in an evaluation of these early initiatives, found that in a sample of 117 ‘career criminals’ 60 had successfully completed probation over a four-year period. Despite the lack of comparison groups, the results led him to concur with initial predictions. He concluded that the polygraph had been successful in reducing recidivism rates. He also reported that its use had led to the recovery of “thousands of dollars worth of stolen property and illegal substances” (Teuscher, 1978: p. 3). At this time, the polygraph was hailed as a valuable supervision tool, not only for ensuring community safety by preventing crime and identifying criminals, but also, potentially as a viable method of reducing prison populations (Consigli, 2002).

Yet despite its extensive use, the polygraph was not universally accepted. The technique was vigorously criticised by some within the scientific community, notably Lykken (1974, 1979) who claimed that polygraph testing was no more than an elaborate gimmick, devoid of any theoretical or empirical foundation. Much of this initial controversy focused on a particular question protocol, the Comparison Question Test (CQT) (Ben-Shakhar, 2002; Furedy, 1996a, 1996b). This debate, which continues today, will be explored in greater depth when evaluating the various theories of polygraphy in chapter 5.

By the early 1980’s, it was estimated that in excess of 23,000 polygraph tests were being administered by various Federal agencies, including the US Postal



Service, while in the private sector approximately one million tests were being conducted in the private sector each year. Three hundred thousand of these tests were for employment purposes (Kleinmuntz & Szucko, 1984; Office of Technology Assessment, 1983). In 1983 President Reagan, in the wake of a damaging leak regarding the funding of defence plans, famously professed to being ‘up to my kiester in leaks’, and issued National Security Decision Directive 84 (Brooks, 1985). This authorised all Federal agencies to polygraph their employees to identify any disclosures of classified information (US Congress, 1983). This directive affected more than 2.5 million government employees in addition to 1.5 million civilian employees at companies doing business with the US government (Office of Technology Assessment, 1983).

The Committee of Government Operations in the US House of Representatives responded, however, by formally requesting the Office of Technology Assessment (OTA) to conduct a thorough review and evaluation of the scientific evidence for polygraph tests. This report concluded, amongst other things, that the accuracy of the polygraph was still undetermined. It specifically highlighted the fact that there was no scientific evidence to support the polygraph’s utility in a screening capacity. The report also raised concerns about the polygraph’s vulnerability to countermeasures (i.e. strategies used to pass a polygraph test), and the rates of false positives (i.e. individuals erroneously identified as deceptive). The report additionally highlighted a number of methodological problems with much of the previous research into polygraphy, and recommended that additional research be undertaken in the area (Kleinmuntz &



Szucko, 1984). President Reagan's directive was rescinded within three months of the publication of this report.

In the 1980's, there was also a growing concern within the wider community about the use of the polygraph. The case of Floyd Buzz Fay provides a compelling example of what can happen when the polygraph 'gets it wrong'. Fay was arrested after a botched armed robbery, in which a clerk was fatally wounded. The dying store clerk picked Fay as the man who had shot him. Fay was arrested but the prosecution, lacking any corroborative evidence, offered him a 'deal'. The charges would be dismissed if he could pass a polygraph exam. If he failed, however, another test would be administered and if he failed that one, Fay was to plead guilty to the lesser charge of murder. If he refused to do so, he would be charged with aggravated murder and the polygraph results would be submitted as evidence. Fay failed both and was eventually convicted of aggravated murder and given a life-sentence. After serving two years in prison, Fay was released when another man confessed to the robbery and identified two others involved (Sullivan, 2001). In a similar vein the website, [www.antipolygraph.org/statements.html](http://www.antipolygraph.org/statements.html), provides numerous testimonies of individuals claiming that the polygraph 'got it wrong' in pre-employment contexts.

The outcome of the OTA report, in conjunction with growing concern about the abuses associated with polygraph testing from both the general public and scientific community, led President Reagan to enact the 1988 Employee Polygraph Protection Act. This legislation restricted the polygraph's use within

the private sector, although it specifically does not apply to Federal agencies and public service employees such as the police (National Research Council, 2002).

Despite these apparent setbacks to the polygraph's credibility, its use by the US Federal government has continued to increase dramatically. In 1986, a congressional mandate called for the expansion of the Department of Defence's polygraph program by allowing for the screening of federal employees in order to 'effect a significant deterrent for spying' (National Research Council, 2002: p.12). The observable fruits of this legislation came in the case of Harold Nicholson, the most highly placed Russian spy in the CIA. In 1996, Nicholson was found to be deceptive in response to the question "Are you working for a foreign intelligence service?" He subsequently failed two additional polygraph tests and was uncovered as a spy (Segrave, 2004). It is also worth noting that Aldrich Ames, another notorious although less illustrious CIA renegade, is said to have passed a number of polygraph tests (Lykken, 1998). This claim is disputed by the American Polygraph Association (D. Sosnowski, personal communication, January 20, 2004). In prison, Ames reportedly said that his Russian handlers had laughed at his worries about taking a polygraph test, telling him to relax because lie detectors did not work (Sullivan, 2001). It is of course possible that he was lying about this, and was not prepared to reveal the type of training he may have received to 'beat' the polygraph.

More recently in 1999, after the discovery of a significant security breach at the Los Alamos National Laboratory, allegedly by a spy from the Chinese government, Energy Secretary Bill Richardson ordered polygraph tests for all of



the laboratory's nuclear weapons scientists. Richardson's swift resort to polygraph testing was meant to plug the holes in federal security and quell controversy (Segrave, 2004). Instead, it launched a debate about the accuracy of the tests and about whether federal agencies should use them on their employees, even though government officials claimed that the polygraph assisted with identifying Wen Ho Lee as the spy (Sullivan, 2001). Concerns about the polygraph's use in this screening and security vetting capacity led the Department of Energy to request that the National Academies of Science (NAS) conduct a scientific review of the research on the polygraph (National Research Council, 2002).

### *Polygraphy in the United Kingdom*

To date the polygraph has not been used in an official capacity in the United Kingdom (UK), although it was subject to considerable scrutiny in the mid-1980's (British Psychological Society, 1986). This was prompted by the Geoffrey Prime spy scandal. Prime, a translator in the Government Communications Headquarters (GCHQ), had been supplying secrets to the Soviet Union for more than a decade (Segrave, 2004). In 1982, Prime was arrested for attempting to molest female children, and it was after the investigation of these sex crimes that his extensive career as a spy was uncovered. The British government, therefore, became interested in the potential of using the polygraph for security vetting purposes.

A working group was assembled by the British Psychological Society (1986) to investigate its reliability, validity and associated ethical concerns. The



report concluded that polygraph procedures at this time were not sufficiently standardised to be acceptable as a scientific test. It stressed the limited amount of empirical evidence to demonstrate the polygraph's accuracy and reliability. It further expressed particular concern about inducing anxiety in subjects, as well as about misleading them about the efficacy of the procedure, suggesting that such methods were contrary to the spirit of the British Psychological Society's code of conduct, and also likely to contravene British law thus rendering information gained from such procedures to be inadmissible in the court systems. The report raised also concerns about the regulation of polygraphy within the UK. Partly as a consequence of this report, the UK government decided not to adopt the polygraph for personnel screening and other investigative uses.

### *The current status of Polygraphy*

Today the polygraph continues to be widely used in the US and in numerous other countries. According to Barland (1999), sixty-nine countries have polygraph capability, whilst commercial manufacturers of polygraph instruments have recently appeared in Russia and China. Practically all federal and many local law enforcement agencies in the US use the polygraph in criminal investigations (Beardsley, 1999; Honts & Perry, 1992; Lehrer, 1998). Polygraph testing also continues to find application in the workplace. Although many screening uses of the polygraph in the private sector in the US are prohibited, employers can still request a polygraph examination to investigate specific losses or security concerns (Segrave, 2004). In the public sector, federal, state, and local governments utilise polygraphy in personnel selection, and also, for security clearance processes in

national security agencies (Honts, 1994). The polygraph over the last decade has also been increasingly used in treatment and supervision of community-based sex offenders (Holden, 2000; Wilcox, 2001).

The admissibility of polygraph evidence in US courts varies between the states. Polygraph evidence (i.e. test results or disclosures whilst testing) can be presently admissible in nine of the twelve federal circuits and, as referred to earlier, in nineteen states (Daniels, 2002). Usually such evidence can either be accepted at the discretion of the trial judge or when both parties agree to admissibility prior to the polygraph test being conducted. Yet despite this apparent increased preparedness to consider polygraph evidence in the courts, there are only a few examples in which requests for polygraphy evidence has been accepted by trial judges (Cross & Saxe, 2001). Other jurisdictions have absolute bans on polygraph evidence and even the suggestion that a polygraph examination is involved can lead to a retrial (e.g. Oklahoma, the 5<sup>th</sup> circuit) (Daniels, 2002).

The American Polygraph Association (APA), which was established in 1966, is the largest professional association for polygraphy in the world, and has around 2,500 members in 33 countries (D. Sosnowski, personal communication, January 20, 2004). The APA also has a code of ethics, standards of practice and grievance procedures (see also [www.polygraph.org](http://www.polygraph.org)). In the US, the APA accredits fifteen polygraph schools, and a further four schools internationally, although scores of other 'non-APA accredited' schools exist offering training to be a 'polygraph examiner'. Twenty states in the US require polygraph examiners to be registered with a state board or the APA to practice. Apart from this, there is little

regulation of polygraphy and individuals do not have to be licensed or have undergone any particular training in order to purchase a polygraph instrument and to call themselves a ‘polygraph examiner’, or to practice privately. The APA has no power to sanction poor practice by examiners other than to end their membership.



## CHAPTER 3

### Polygraph Technique

In a polygraph test, physiological changes are assessed in reaction to a carefully structured set of questions. All polygraph techniques assume that the measured physiological reactions are not under voluntary control, and will occur with greater strength in response to questions that are most important to the individual. There are three general classes of question procedures:

- Relevant-Irrelevant Test (RIT);
- Comparison Question Tests (CQT);
- Concealed Information Tests (CIT).

All of these have been referred to in chapter 2, and are described in more detail here. These techniques differ in their theoretical rationale, and also in the choice and sequencing of the questions asked during the examination. All consist, however, of three basic phases. These are: a pre-test interview; an examination process or data collection phase; and a post-test interview (Raskin & Honts, 2002; National Research Council, 2002).

#### *Instrumentation*

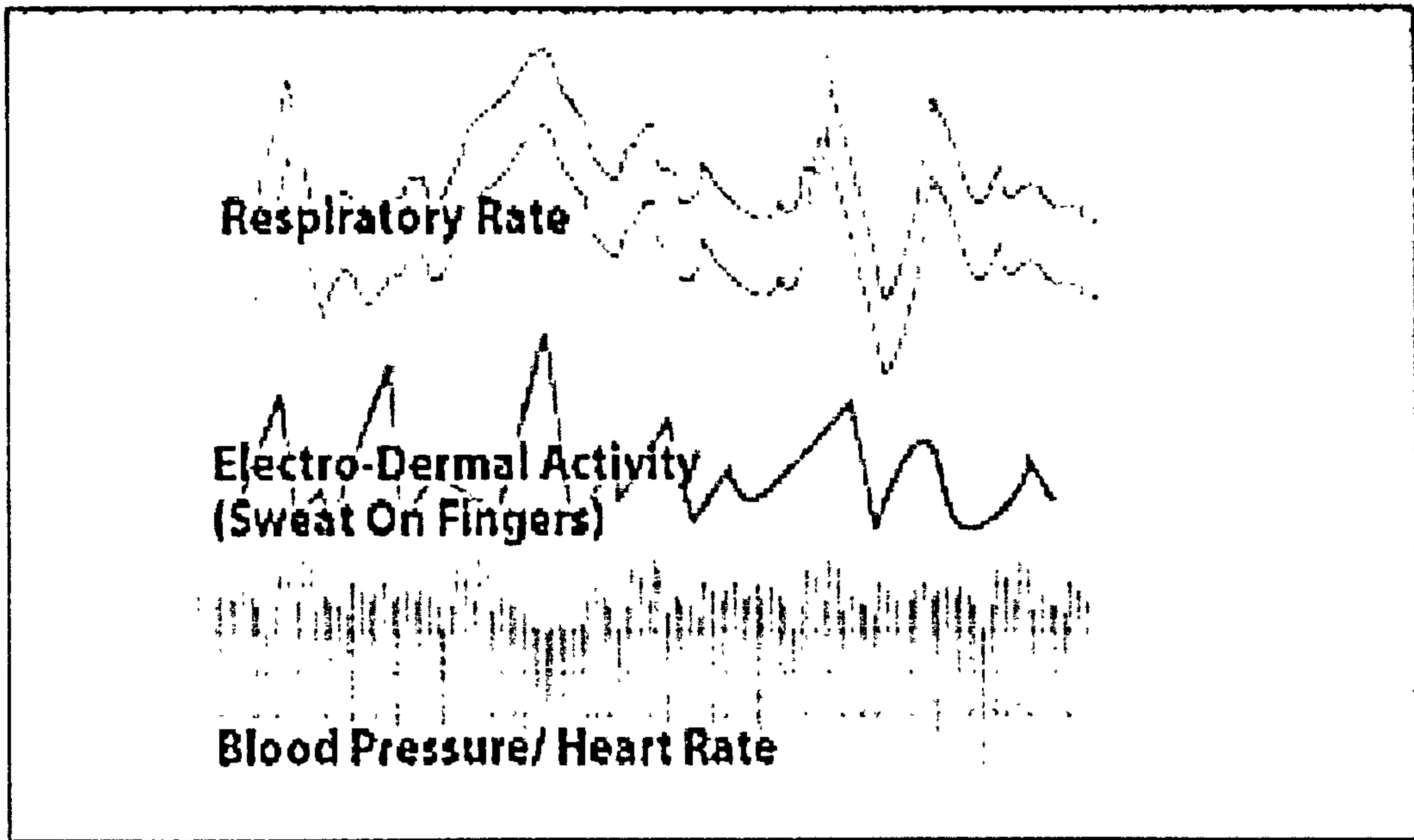
The typical polygraph instrument measures three physiological phenomena associated with arousal. These are cardiovascular activity, breathing and sweating. Cardiovascular data is collected via the use of a blood pressure cuff placed on the

upper arm. This cuff is filled with air and connected to the polygraph machine by air-filled tubes. Changes in blood pressure modulate the air pressure in the cuff, and are recorded by the polygraph instrument and displayed on a computer screen. Pneumotubes are attached across the chest and abdomen to record the rate and depth of respiration. The measurement of sweat, which is known as the galvanic skin resistance, is conducted by a two-piece galvanometer attached to two of the examinee's fingertips. The galvanometer works by sending a small electric current into the skin from one of the fingerplates and records how much current was allowed to pass through on the other fingerplate. Dry skin is not a very good conductor of electricity. If a subject perspires, however, the water and salt from the sweat reduces the resistance of the skin. This decrease in resistance allows a larger amount of electric current to travel along the surface of the skin. The amount of electric current recorded by the galvanometer reflects, therefore, the amount of sweat that was produced in the subject's fingertips (Matte, 1996).

While early polygraph instruments recorded this information with ink pens over a roll of paper on a moving cylinder, contemporary instruments digitise the information and present it on computer (see figure 1). The recorded data is called the 'polygraph chart' (Raskin & Honts, 2002).

Figure 1.

*Example of polygraph chart*



### *The Pre-test interview*

The first part of any polygraph examination consists of a pre-test interview. The initial portion of this process involves obtaining basic biographical information and a brief health history from the examinee. The purpose of the test and the specific issues that are going to be covered are then discussed. The examiner will explain the instrumentation and the theory of the polygraph to the subject. The examiner may then conduct a brief demonstration test (sometimes referred to as an acquaintance test) to accustom the examinee to the recording procedure, and also, to convince the examinee that the procedure is an accurate and reliable method for identifying deception and truthfulness (Kleinmuntz & Szucko, 1984). Several variations of the acquaintance test exists, a common example of which involves the examinee choosing a number between three and six. The



examinee is then asked to deny having chosen any of the numbers one through eight while recordings are made with the polygraph. After completing the acquaintance test, the examinee will be informed that a clear reaction occurred to the deception on the chosen number and that no change was observed in response to the truthful answers given to the other numbers. The subject is then shown the chart demonstrating this (Raskin & Honts, 2002). The examinee may also be told that any deception on the actual test would induce an even larger physiological reaction, because the deception would be more serious (Abrams, 1973; 1991).

Practitioners claim that the acquaintance test serves to reassure the innocent subject, whilst causing increased levels of apprehension in the deceptive subject. This is expected to lead to greater differential responsivity between deceptive and non-deceptive subjects to questions on the examination (Bradley & Janisse, 1981a). Research has tended to support this claim (Kircher, Packard, Bell & Bernhardt, 2001).

Finally, during the pre-test interview the examiner will formulate the questions that are to be asked during the examination phase. These questions are constructed so that the examinee can answer “Yes” or “No”, and are reviewed in detail with the examinee during the pre-test interview to ensure that there is no ambiguity in the meaning of the questions (Abrams, 1991). Depending on the complexity of the case, examiner-examinee interaction and the testing technique used, the pre-test interview can last between 30 minutes to over 2 hours (National Research Council, 2002).

## *The examination process*

The polygraph attachments are placed on the subject either during the pre-test interview or at its conclusion. A short period of about 10 to 15 seconds, is used to observe the examinee's initial respiratory cycles. This will establish a baseline and will allow any initial response to fade. The examiner then asks the first question. Between each question, the examiner waits for about 15 to 20 seconds until the physiological response to the preceding question has returned to baseline. The examiner notes on the chart when the exam begins, when questions are asked, and when it ends. Extraneous behaviour that affects the recordings may also be noted. The examiner then inspects the chart and may ask the examinee about his or her reactions to the questions. The examiner will then repeat the same procedure to obtain two more charts (Raskin & Honts, 2002).

## *Question Procedures*

### *The Relevant-Irrelevant Test*

The RIT was the first polygraph procedure developed. The typical RIT employs a series of 10 to 15 questions comprised of relevant questions and irrelevant (or neutral) questions. The relevant questions focus on the issue under investigation. These are worded in simple terms that allow for an unambiguous interpretation of their meaning so that examinees can answer in a 'yes' or 'no' manner. An example is 'Did you steal the £50 from the cashier till on Thursday?' Irrelevant questions have nothing to do with the issue under investigation, for

example, ‘Do people call you John?’ Irrelevant questions are typically placed first in the question sequence because the physiological responses that follow are presumed to have no diagnostic value. They are also placed at other points in the question sequence (Reid & Inbau, 1977). The rationale of the RIT predicts that the guilty subject will react stronger to the relevant items when compare to the Irrelevant items (Matte, 1996). An example of a typical test sequence is shown below:

Table 1.  
*An example of the RIT question sequence*

Question	Question Type
1. Is today Wednesday?	Irrelevant
2. Do you intend to lie to me on this test?	Relevant
3. Are you sitting down?	Irrelevant
4. Did you rob the supermarket last Tuesday?	Relevant
5. Did you use a gun to rob the supermarket last night?	Relevant
6. Do you sometimes listen to the radio?	Irrelevant
7. Is your name Barry?	Irrelevant
8. Did you rob the cash register at the supermarket?	Relevant
9. Did you take money from the supermarket last night?	Relevant
10. Do you live in the United States?	Irrelevant
11. Have you lied to me today?	Relevant
12. Do you like alcohol?	Irrelevant

The RIT was mainly developed in a law enforcement context and tends to be orientated towards obtaining confessions. The pre-test interview is, therefore, typically confrontational and may lead to an interrogation before the polygraph test



is completed (Lykken, 1998). Examiners interpret the test results globally by inspecting the charts to see whether or not there is a pattern of stronger responses to the relevant questions (Raskin & Honts, 2002). Polygraph researchers have generally tended to view the RIT as outdated because of the questionable premise that only 'deceptive' individuals will physiological respond to the relevant questions. It does, however, continue to be used by some practitioners especially for employment screening tests (D. Sosnowski, personal communication, January 20<sup>th</sup>, 2004).

### *Comparison Question tests*

Reid (1947) developed the original Comparison Question Test (CQT). Various versions have emerged since this time, although the basic structure of the test and the premise on which it is based remains the same. Comparison Question Tests can be used to investigate a single or multiple issues, and are the question technique used in post-conviction sex offender testing (Holden, 2000; Lundell, 2000; Matte, 1996). Like the RIT, the CQT asks relevant questions and also assumes that the deceptive examinee will display greater physiological reactions to these questions when compared to others. The CQT, however, incorporates 'comparison questions'. Comparison questions ask about general undesirable acts and are designed to elicit a lie (Matte, 1996). A typical introduction to the comparison questions may be as follows:

“Since this is a matter of theft, I need to ask you some general questions about yourself in order to assess your basic character with regard to

honesty and trustworthiness. I need to make sure that you have never done anything of a similar nature in the past and that you are not the type of person who would do something like robbing that supermarket and then lie about it. Therefore, I need to ask you some questions for that purpose. So, if I ask you, “Before the age 27, did you ever do anything that was dishonest or illegal?” you could answer that “No,” couldn’t you?” (Raskin & Honts, 2002: p. 15)

These types of questions are referred to as probable-lie comparison questions, that is, they are designed to induce subjects to lie or at least experience some doubt about the veracity of their answer (Raskin & Honts, 2002). If an examinee answers “Yes” to this question, the examiner will typically ask for an explanation, attempt to minimize the examinee’s explanation, and then discourage further disclosures until a “No” is obtained. During this process, the examiner is essentially attempting to manoeuvre examinees into a situation where they feel they have no choice but to be deceptive, or at the very least, to be very unsure of their response (Matte, 1996). An examiner, for example, may respond to disclosures in regards to the comparison issue by saying, “that occurred a long time ago, but surely you won’t do that type of thing now, would you?”

Instead of a probable-lie comparison question, some CQT’s incorporate directed-lie comparison questions (Raskin & Honts, 2002). The examinee is specifically instructed to lie in these formats. In such tests, the purpose of the directed lie questions may be explained to the subject in the following manner:

“On this test I need to ask you some questions to which I want you to lie. Just as on the number test, I need to have questions to which you and I both know you are lying and some that you and I know you are answering truthfully. That way, I can see the difference in your reactions when you lie and when you tell the truth, and I will be able to see if your reactions on the questions about the robbery are the same or different compared to the questions I know you answered with a lie. Therefore, I am going to ask you, “During the first 27 years of your life, did you ever tell even one lie?” I want you to lie to that question. Also, I want you to think of a particular time when you did lie in the past, and I want you to have that in mind when you answer this question on the test.” (Raskin & Honts, 2002: p. 23)

Regardless of the type of comparison question used, the expectation is always that the innocent examinee will react more strongly to the comparison questions, whilst the deceptive examinee will react more strongly to the relevant questions. The CQT typically employs from two to four relevant questions in a test sequence of 10 to 12 questions, the number and type of comparison questions used varying with different CQT protocols (Matte, 1996). Some of these different variations of the CQT are discussed in Appendix A.



## *Computer Scoring*

In addition to numerical scoring procedures computer algorithms have also been developed to score polygraph charts. Computer scoring represents a potentially important development for polygraph, as such methods can overcome examiner biases and eliminate problems of inter-rater variability. The two most commonly used computer-scoring systems at present are the Computerised Polygraph System (CPS) and PolyScore.

The CPS was developed using the physiological data from mock polygraph examinations. This program replicates what human scorers do by utilising information from three physiological sources (i.e. skin conductance, blood pressure change, and respiration) (Kircher & Raskin, 2002). The CPS algorithm uses a multivariate discriminant function analysis, and Bayes' Theorem to calculate the probability of deception. By contrast, the PolyScore algorithm was developed using logistic regression on the results of 'real-life' polygraph test charts. The developers report that the algorithm transforms signals on galvanic skin response, blood pressure, and upper respiration into what they described as 'more fundamental' signals of deception. These signals are then used in the determination of probability of deception (National Research Council, 2002).

Kircher and Raskin (2002) reported on eight studies that compared the CPS algorithm results with those obtained from human scorers. These results indicated that the CPS was no different than human scorers in discriminating truth telling and deceit in mock polygraph examinations (i.e. tests conducted with volunteers).

In another study, Dollins, Krapohl and Dutton (2000) compared a number of algorithms, including the CPS and PolyScore. The study reported that there was no statistical difference in the classification powers of the different algorithms.

There are a number of difficulties that need to be borne in mind when considering computer-scoring algorithms. The first difficulty concerns the problem of developing algorithms that perfectly separate truthful and deceptive individuals using a variety of multivariate methods and a large set of data is relatively easy. Such a process often leads, however, to the ‘over-fitting’ of data (National Research Council, 2002), but on a new set of data these complex algorithms often perform less accurately. The second difficulty is that the algorithms are only as good as the data from which they were developed. The CPS was developed using mock polygraph tests. Whilst data gained from such sources are statistically ideal because of standardised procedures, these examinations lack realism so it is arguable that they replicate the stimulus intensity of real situations. Using field data (i.e. real life polygraph tests) poses other difficulties. Real life polygraph cases exhibit considerable variability in the format and administration of the tests. The National Research Council (2002) conducted a brief evaluation of 149 criminal cases that were used in the development of PolyScore method and found that there were considerable differences in the type of crime investigated, the testing format, and in the number and questions asked. Finally, it is worth emphasising that it would be incorrect to assume that research comparing computer-scoring algorithms and human scores reveal anything about the overall accuracy of the polygraph.

## *Concealed Information Tests*

The Concealed Information tests works on an entirely different premise from either the CQT or the RIT. Instead of detecting deception, concealed information tests aim to detect whether a suspect has information about a crime that only a guilty subject would have, or in some cases, to detect the information itself (Office of Technology Assessment, 1983). Such information might include details about the site of the crime, or about the means used to commit it (e.g., the type of murder weapon used). Concealed information tests take two forms: the peak of tension (POT) test and the guilty knowledge test (GKT). The discriminating feature between these procedures is whether the examiner knows the information that is being sought.

Criminal investigators use the POT technique to discover and to develop additional information about a case. In such tests, the examiner asks the examinee about a series of details but does not know which is actually relevant to the crime. The detail that provokes an exceptional physiological response is used as a clue in the investigation (Matte, 1996). For example, an examiner might use POT to determine the exact location where stolen goods were hidden. This kind of examination is called a searching peak of tension test. The searching POT technique has been used, for example, in cases in which employees are suspected of having stolen money, but where there is no evidence about the extent of the theft. The examiner asks the employee if he has stolen money ranging from a small amount to the entire amount taken. The amount that provokes the largest response



is assumed to be the amount of the total that the employee stole. The examinee does not even need to response “Yes” or “No” to the questions.

The GKT, described initially by Lykken (1959/1998) works in much the same way as POT but includes a larger set of questions, and these questions may be of the multiple-choice type. A typical GKT may include five to nine nearly identical “Yes or No” questions, specifically focused upon details related to a crime. The detail may be a type of object used, or the colour of an item. One question actually includes the relevant detail, while the others include plausible but false details of a parallel nature. The questions and the sequence in which they are asked are reviewed with the subject in the pre-test interview. The subject is instructed to answer “No” to each question. The question with the true detail is usually presented in the middle of the sequence, so that the subject’s physiological reactions will increase up to the critical question, where they will reach a peak and fall back down again (Nakayama, 2002). Nakayama (2002) provides an example of a typical GKT’s question sequence as follows:

Question 1. When was he killed? Was he killed in:

1. January?
2. February?
3. March?\*
4. April?
5. May?

In this case, alternative 3 is the critical question and the others are non-critical questions. Concealed information tests are applicable only under restricted conditions. These are when there is a specific incident or activity known only to the investigators and the offender that can be the subject of questioning. The testing format becomes invalid if such information is either revealed through the mass media or during interviews associated with the investigation (Ben-Shakhar, Bar-Hillel & Lieblich, 1986). In evaluating a GKT produced chart the examiner considers the examinees reactions to each question separately to identify the strongest response. If the examinee has the strongest response to the known information, it is assumed that the examinee has ‘guilty knowledge’ about the issue under investigation (Ben-Shakhar & Dolev, 1996).

### *The Post-test interview*

The final stage of the assessment is the post-test interview. Different test techniques and circumstances can lead to somewhat different examiner behaviour in this phase. In many cases, the examiner will indicate a ‘diagnosis’, and if an examinee is thought to be deceptive, the examiner typically pursues this with further questioning (Lykken, 1998). This may or may not be done directly and may be couched in terms of providing the examinee with an opportunity to clarify his or her responses. For example, the examiner may say “You seem to be having a problem in the area of X (the relevant item)” and ask the examinee whether he or she can think of a reason for having a strong reaction to that question (Raskin & Honts, 2002). If the examinee is identified as ‘truthful’, the examiner will ask them about reactions on the comparison questions.

## *Conclusions*

What is known as a ‘polygraph exam’ is actually a relatively complex set of procedures for asking questions, and then, measuring and interpreting physiological responses in order to identify deception in the examinee. Broadly speaking, the various protocols all assume that individuals will respond with involuntarily physiological reactions to the questions that are the most relevant for them.



## CHAPTER 4

### Sex Offending & Relapse Prevention

Sexual offending is a serious social problem. Home office research has estimated that in 1993 at least 260,000 men had been convicted of some kind of sexual offence, of whom 210,000 had convictions with a victim (Home Office, 1995). In the United Kingdom the criminal justice system manages most sex offenders with a combination of incarceration, specialized treatment and community supervision (Browne & Lynch, 1998). Many exceptional criminal justice policies, such as the Sex Offender and Criminal Justices Act 1997, target those offenders returning to the community likely to re-offend.

Currently the most common treatment approach for sexual offenders is the Relapse Prevention Model (RP; Launay, 2001; Laws, 1999; Marshall, Anderson & Fernandez, 1999). The central tenet of this approach is that a sexual re-offence is not an impulsive act, but rather a sequence of behavioural and cognitive steps that result in sexual offending if no interventions are taken. A relapse is defined as a return to sexual offending, whilst a lapse is seen as an emotion, fantasy or behaviour that is part of the individuals deviant cycle of sexual offending behaviour. These 'high-risk' features can be idiosyncratic for the individual offender and can include behaviours such as, masturbating to deviant (offence-related) fantasies, using pornography, or having unsupervised contact with children or other potential victims (Beech & Fisher, 2002; Cumming & Buell, 1996).

Whilst such behaviour may appear harmless to others, each of these mini-decisions moves the individual closer to an actual sexual offence (Laws, 1999). For example, an offender, previously convicted of voyeurism may choose to walk his dog late at night, convincing himself that he is only doing so 'because the dog needs the exercise'. Such decisions may seem relatively benign, however, contributes to an increase in the probability of exposure to high-risk situations (e.g. opportunities to observe windows late at night) or high-risk factors (e.g. erosion of the offender's sense of self-control). At such a point, an adaptive response, such as walking the dog during the day, can result in a reduced risk of re-offence. A failure to cope, however, may lead to further lapses (e.g. lingering around windows) and eventually a re-offence.

During the initial stages of RP treatment considerable attention is paid to identifying high risk factors that may lead to a sexual offence. In addition deficits in the coping skills of the offender are identified (Mann, 2004). The therapy then shifts the focus to assist the offender develop skills and strategies to manage these high-risk factors. The various treatment modalities utilised include: cognitive-restructuring; covert sensitisation; and social skills and empathy training (Maletzky, 1998; Mann, 2004). At the completion of therapy a Relapse Prevention Plan will be developed by the offender that details their high-risk behaviours and situations, and also outlines strategies to manage such factors (Center for Sex Offender Management, 2001; Marshall, Anderson & Fernandez, 1999).

Community supervision dovetails with the treatment plan by monitoring and restricting the offender's identified risk behaviours, and thereby, ensuring that

their risk of a re-offence remains as low as possible (Center for Sex Offender Management, 2000a; Mandeville-Norden & Beech, 2004). Traditionally, however, it has proven difficult to ensure that offenders honestly engage in therapy (i.e. fully disclose the nature of their sexual behaviour problems) and also in complying with their set probation conditions, such as avoiding contact with children. The development of risk assessment tools, such as the Static 99, and the identification of dynamic risk factors associated with sexual recidivism, such as psychological distress, partly assist with overcoming this problem by differentiating the offenders most likely to re-offend (Hanson & Bussière, 1998; Hanson & Harris, 2000). Supervisors can therefore specifically target the more high-risk offenders with appropriate strategies. Nonetheless, a large portion of supervision and treatment remains reliant on the offenders self-report, and studies have generally found that deception occurs with the majority of sexual offenders.

Maletzky (1991) reported that 87% of these offenders denied all or parts of their crimes. Barbaree (1991) similarly observed that 54% of rapists and 66% of child molesters completely denied having offended, and indeed 98% of all the sexual offenders in his sample either completely denied or minimised their offences. In response to this shortcoming of community supervision a number of methods have been suggested for monitoring an offender's behaviour in the community; these include 'tagging' and 'intensive' supervision (Home Office, 1995). While such methods do possess some merit they remain problematic in areas. For example, 'tagging' provides information only about an offender's location and does not serve to gauge explicit behaviour, whereas twenty-four hour supervision for all community-based sex offenders is infeasible.



The inability to effectively monitor a sexual offender's behaviour, therefore, represents a concerning limitation of current supervision practices. The polygraph could be a method of assessing lapse behaviour and also adherence to probation conditions.

### *Post-conviction polygraph testing*

Much of the polygraph's applications described in the earlier chapters relate to investigative and pre-conviction law enforcement uses. As briefly mentioned earlier, however, the criminal justice system in North America has also witnessed a steady increase in the use of the polygraph in post-conviction circumstances. In 1973, Abrams, a psychologist and polygraph examiner in Jackson County Oregon, was the first to initiate the use of the polygraph specifically with sex offenders. He reasoned that the polygraph could improve supervision and treatment efficacy by counteracting the inherently secretive nature of sex offending (Abrams, 1991). Throughout the late 1970's and 1980's, the polygraph was used by a small number of treatment and probation programmes mainly on an ad-hoc basis (Abrams, 1989).

In 1991 the first formal polygraph-testing programme for sex offenders in the North West of the US was established (Abrams & Abrams, 1993). Shortly after, similar programs developed in Hawaii, Texas, and Tennessee. During this time English and colleagues in Colorado pioneered the 'Containment Approach' (English, 1998; Heil, Ahlmeyer, McCullar & McKee, 2000). This model

incorporates polygraphy as one component of an overall strategy that includes collaboration between treatment providers and probation officials, to ‘contain’ community-based sex offenders (English, Jones, Patrick, Pasini-Hill & Gonzalez, 2000).

A number of well-respected clinicians within the field of sex offender assessment and treatment have also come out in support of the polygraph’s use. For example, Anna Salter (1997) claims that they do not have ‘denier groups in the North West of the US anymore’ because of polygraph. More recently the Center for Sex Offender Management (2000a) in the US has described polygraph testing as “an important asset in treatment and supervision of sex offenders” (p.11). Indeed even David Lykken, one of the most vehement critics of polygraphy, concedes that “periodic testing might well serve both as a deterrent and also encourage more complete reporting”, although he adds “...it should not be interpreted as a valid test for truth” (D. Lykken, personal communication, October 29, 2003).

In support of these developments, the Association for the Treatment of Sexual Abusers (ATSA) have established detailed standards for polygraph use. In ATSA’s (1997) *Ethical Standards and Principles for the Management of Sexual Abusers* it states that the polygraph has “significant clinical value” (p.36), although it cautions professionals to avoid “... over reliance on the instrument” (p. 36). These guidelines further suggest that polygraph results “should always be used in conjunction with other sources of information” (p. 36) and not be considered the sole “determination of guilt or innocence” (p.36).

In 1994 a telephone survey of 732 probation and parole supervisors in the US showed that 9.8% reported that offenders in their programs were often or always required to take polygraph tests for treatment or supervision (English, Pullen & Jones, 1996). By 1996, this had increased to 16.3% (English, et al., 2000). Simmons (1998) reported in a recent survey that 33 states used polygraphy with sex offenders. When considering this number, however, it should be noted that this is not an official figure but the perception of the interviewees (i.e. probation officers working in the state). Bearing in mind the relatively small percentage of services reported in the English et al. (2000) study, Simmons (1998) data probably presents an exaggerated view of the use of polygraphy with sex offenders at that time.

Numerous states in the US have enacted legislation that requires sex offenders to take polygraph tests. The Supreme Court of the State of Washington, for instance, has ruled that Trial Courts have legal authority to subject sex offenders to polygraph tests (Spencer, 1998). In Texas, polygraph testing is mandatory in work with adolescents, and in an apparent rush of enthusiasm, the state also considered making monthly tests mandatory for all sex offenders. This proposal was abandoned, however, after the Texas Association of Polygraph Examiners advised against such a plan due the concern that offenders could habituate to the tests (McKay, 2000). In Tennessee, anyone convicted of two or more sexual offences are required to undergo periodic polygraph examinations. Colorado has also recently passed comprehensive legislation for the treatment of



sex offenders that includes mandatory polygraph testing (Colorado Sex Offender Management Board, 2000).

The American Polygraph Association (APA) has also been quick to respond to this growing interest and demand for post-conviction polygraph examinations. In 1998, a Post-Conviction Sex Offender Testing (PCSOT) subcommittee was established and standards of practice were developed. At present, examiners wishing to conduct post-conviction testing are required to complete a 40-hour specialised training programme ([www.polygraph.org](http://www.polygraph.org)). To date, 560 examiners have completed this course (D. Sosnowski, personal communication, January 20, 2004).

### *Pre versus Post-Conviction Polygraphy*

Much of the on-going controversy surrounding the polygraph focuses upon pre-conviction ‘investigative-type’ uses of the polygraph. It is argued that such uses are distinct from post-conviction polygraphy (Abrams & Simmons, 2000; Holden, 2000). In the former, issues of accuracy are particularly important, as someone may falsely be implicated in a crime or fail a pre-employment screen. In the latter, the focus is on its utility (Cooley-Towell, Pasini-Hill & Patrick, 2000; Lundell, 2000). Within a post-conviction context, it is claimed the polygraph is not considered a ‘test’ but a treatment tool, used in conjunction with other tools, specifically to elicit admissions regarding past and current behaviours (Williams, 1995; Wilcox, 2000). Another important distinction is in the role of the polygraph examiner. In a post-conviction context, the examiner is often considered part of a

treatment team that includes the offender's therapist and probation officer. These professionals collaborate to gather information regarding the offender, and institute supervision strategies and treatment interventions to ensure that the offender remains at low risk of re-offending (Dutton, 2000; English, et al., 2000).

There are three types of post-conviction polygraph examinations that are used at different times and stages of treatment and supervision (Lundell, 2000). Sexual history or disclosure examinations are usually utilised shortly after an offender has commenced treatment. Upon commencing treatment, offenders are typically required to complete a comprehensive Sexual History Questionnaire. This covers their previous deviant and non-deviant sexual interests and behaviours. The polygraph is used to encourage the offenders honesty when completing this survey (Abrams, 1991; Blasingame, 1998). Specific Issue or Denial tests are commonly used to verify the veracity of an offender's disclosures during treatment. Such tests are usually administered when the offender is in denial of his offence or if his version of the crime varies substantially from the victim's version. Specific-issue exams may also be used to address a single concern or suspicion that arises during an offender's probation, as for example, when there is suspected contact with children (Abrams, 1991; Cooley-Towell, et al., 2000). Finally, monitoring or maintenance exams focus upon the offender's current behaviour, and are used on a regular basis to monitor compliance with supervision and treatment conditions (Holden, 2000). All these exams require the polygraph examiner, treatment provider and supervising officer to work together to identify the most appropriate areas to target in the polygraph test (Consigli, 2002).

## *Post-Conviction Sex Offender Testing protocols*

The American Polygraph Association (APA) promotes two Comparison Question Test (CQT) formats to be used in sex offender testing. These are Department of Defense of Polygraph Institute Zone Comparison Test (DoDPI Zone) and the Modified General Question Test (MGQT), or sometimes called the Air force MGQT (AFMGQT) (Dutton, 2000). Like the previously described Reid and Backster versions these protocols incorporate comparison and relevant questions. A notable difference is that the pre-test in the Post Conviction Sex Offender tests (PCSOT) is conducted in a low-key, non-accusatory manner designed to obtain information from the subject without confrontation or pressure (Matte, 1996). The specific format and scoring procedure for each of these protocols are described below.

The DoDPI Zone is used to investigate single issues, such as may be required in specific issue or sexual history tests (Dutton, 2000; Matte, 1996). The test has ten questions (three relevant (R); three comparison (C); two symptomatic; one irrelevant, and one sacrifice relevant). An example of a question sequence is presented in Table 2. The first two questions (irrelevant and sacrificial relevant) are considered to be ‘buffers’ and are designed to absorb the examinees reaction to the initial question, and also to the first question asking about the relevant issue. The ‘symptomatic’ questions are identical, and are expected to serve the same purpose as Backster’s outside issue questions (Matte, 1996).



Like other CQT formats, this test is evaluated by comparing the relevant and comparison questions. For each relevant question, the examinees physiological reactions are compared to their reactions to an adjacent comparison question. The reactions are compared for each physiological channel, that is, the GSR on the R is compared to the GSR on the C, and so on. If the reaction on the R question is greater than the nearby C it is scored as  $-1$ , if the C is greater than the R then it is scored  $+1$ ; and if there is no difference it is given a 0. These scores are then combined for each specific R question to give a total score between  $-3$  and  $+3$ . This is done for each of the R questions, for all three charts, thus giving three scores for each R question and nine scores in total. These are then combined to give an overall chart score. If this score is greater than  $-6$ , the polygraph examination is considered to be deceptive. If the score is greater than  $+6$ , it is considered non-deceptive. The result of polygraph examination is considered to be inconclusive if the score lie between  $-6$  and  $+6$  (Matte, 1996).

Table 2.

*An example of DoDPI Zone test question sequence as used in a Sexual History Test*

	Question	Question Type
1.	Is this the month of December?	Irrelevant
2.	Regarding your sexual history, do you intend to answer truthfully each question about that?	Sacrifice Relevant
3.	Are you completely convinced that I will not ask you a question on this test that has not already been reviewed?	Symptomatic
4.	Since you were arrested, have you done anything you wouldn't want your family and friends to know about?	Comparison
5.	Are there any undisclosed victims you haven't told me about?	Relevant
6.	Since you were arrested have you kept secrets from anyone who loves and trusts you?	Comparison
7.	Have you falsified any information about your sexual history?	Relevant
8.	Is there something else you are afraid that I will ask you a question about, even though I have told you I would not?	Symptomatic
9.	Since you were arrested, have you lied to anyone to cover up your whereabouts?	Comparison
10.	Are you withholding any information about your sexual history?	Relevant

The MGQT is typically utilised to investigate multiple-issues, and may be used in a maintenance or monitoring exam (Dutton, 2000). The format includes ten questions (four relevant, four comparison, and one irrelevant and sacrifice relevant). An example of the MGQT is presented in Table 3. Because this is a multiple issue test the scoring is slightly different to the DoDPI Zone. Each of the R questions are scored in the same manner as for the DoDPI Zone, but these scores

are not combined to give an overall chart score. To be considered to have passed an MGQT, an examinee must score at least +3 on every R question. If an examinee has a score of –3 on any of the four R questions it is considered to be a ‘failed’ (deception indicated) test. If an examinee does not achieve +3 or higher on every R question, and does not score a –3 on any of the R questions, then the test is considered inconclusive (Matte, 1996).



Table 3.

*An example of an MGQT question sequence as used in a Maintenance Test*

Question	Question Type
1. Do some people call you John?	Irrelevant
2. Regarding your sexual behaviour do you intend to answer all questions about that truthfully?	Sacrifice Relevant
3. Other than what you told me about, since you've been on probation have you lied to keep yourself out of trouble?	Comparison
4. Since you've been on probation have you been alone and unsupervised with anyone under the age of 16 years?	Relevant
5. Other than what you have told me about, since you've been on probation have you lied to anyone who loves and trusts you?	Comparison
6. Since you've been on probation have you set up a situation to have sexual contact with anyone under the age of 16 years?	Relevant
7. Since you've been on probation have you deliberately mislead any member of your treatment group?	Comparison
8. Since you've been on probation have you gone to any places to specifically access anyone under the age of 16 years?	Relevant
9. Since you've been on probation have you done anything you wouldn't want your family and friends to know about?	Comparison
10. Since you've been on probation have you masturbated to images of anyone under the age of 16 years?	Relevant

## *Research Evidence*

Advocates argue that sexual history and specific-issue polygraph examinations enable clinicians to obtain more reliable sexual histories and more accurate offence behaviour descriptions. Both of these assist in overcoming denial in the offender and improve the therapist's assessment of treatment need and risk of re-offending (Lundell, 2000; Salter, 1995). It has been argued that in terms of supervision, maintenance polygraph exams have the potential to identify not only breaches in supervision and outright offences, but also act to deter offenders from engaging in problematic behaviour in the first place (Abrams, 1991). Remarkably little empirical research has evaluated these claims, with much of the published literature on post-conviction polygraphy being theoretical or anecdotal in nature (Abrams & Simmons, 2000; Blasingham, 1998; Cooley-Towell, et al., 2000; Kokish, 2003; Wilcox, 2000). What research that has been done has, however, generally tended to be supportive.

## *Assessment of Sex Offenders*

Emerick and Dutton (1993) compared assault history information for adolescent sex offenders at three times:

- initial file information;
- after an intake interview, and finally
- after a polygraph examination.

The results showed there was a significant increase in the mean number of victims (1.5 to 1.9 and then 2.8) and offences (20.6 to 27 and then 76.6) reported

after a polygraph examination. They further reported gaining additional information regarding the nature of force used during their offences, the use of pornography and the number of paraphilic interests.

Emerick and Dutton concluded that the polygraph assisted with gaining significant amounts of additional information. Because all the offenders were identified as ‘high-risk’ it is questionable, however, whether these findings can be generalised to other adolescent sex offenders. Also the lack of control or comparison groups makes it difficult to disentangle the effect of the polygraph with other factors that might affect disclosure, such as treatment effects. Finally, as none of the disclosures were verified, it is unclear just how many of these reported additional offences actually occurred. Offenders may have disclosed additional false information because they believed it might benefit them in some way.

The approach taken by Emerick and Dutton’s study represents the typical methodology utilised to investigate the impact of polygraphy on admissions in sex offenders. Other studies have reported similar findings, consistent with these results. In unpublished research, for instance, O’Connell (1998) reviewed 127 archival clinical files of adult sex offenders. In this study an offender’s admissions were compared on three different occasions:

- at the referral interview;
- after a clinical interview, and
- after polygraph testing.



Again the results indicate significant increases in the incidence of sexually abusive behaviours and numbers of victims reported after a polygraph test. In another study, Chambers (1994) found that the mean number of known victims and offences increased by 5.8 and 13.9 respectively after a polygraph test. In the only UK study, Wilcox and Sosnowski (2001) found that in a small sample ( $n = 14$ ) the polygraph, when compared to clinical interviews, assisted with gaining significant amounts of additional information. This included an increase in number of known contact (16 to 73) and non-contact (32 to 110) offences. Overall, while these findings are promising they suffer, however, from the same problems as the Emerick and Dutton's study. This is a lack of comparison groups and absence of verification.

In a study that does include a type of comparison group Ahlmeyer, Heil, Mc Kee and English (2000) compared the admissions of inmates and parolees on four different occasions during the course of treatment. Again consistent with the earlier studies, the results suggested that the polygraph assisted with gaining significant amounts of additional information in both groups. For example, the 35 inmates in this latter study initially reported a mean of 83 victims and 394 offences after completing a Sexual History questionnaire, these numbers increased to 183 and 528 respectively after polygraph testing. This effect was less apparent for parolees (whose pre-polygraph reporting consisted of 2 victims and 3 offences versus their post-polygraph reporting of 7 and 23 respectively). The difference in the degree of admissions between the groups is notable, especially considering that both groups had similar sexual offending histories according to their pre-sentence reports. It is conceivable, that the incarcerated participants were higher risk and

more deviant than the parolees, thus explaining why they were in prison, and also, why the larger amounts of additional information disclosed. If these groups, however, are not different from each other, or if the parolees are more deviant, then the results would appear to cast doubt on the accuracy and value of present risk assessment tools. It could alternatively suggest that there are other reasons why inmates would disclose substantially more information than parolees.

It is noteworthy that the parolees were mandated to engaged in therapy, whilst the inmates volunteered. Furthermore, the parolees had spent less time in what was described as less intensive therapy than had the inmate sample. It is possible, therefore, that the increased amount of information reported by the inmates may partly be due to a treatment effect. One must also bear in mind that making additional disclosures regarding offending behaviours is likely to be considered favourably at parole hearings, so inmates may have been further motivated to appear to be disclosing information. Parolees have conversely less to gain from making additional disclosures, and perhaps, even more to lose. Consider, for instance, the likely change in their circumstances should they disclose an extensive previously unknown history of sexual offending. In this regard, if parolees were withholding information it would be expected that they would be more likely to fail the polygraph than inmates. The study does not present data on failure rates for the two groups, although it is reported 21% of parolees did not disclose additional information after a failed polygraph test compared to only 5% of inmates, suggesting that this may indeed have been a factor.

In another study, English et al. (2000) compared sex offender treatment programs that were seen with and without polygraph testing. The results showed that polygraphed offenders were more likely to report having offended against both male and female victims, and also against both juveniles and adults. These offenders generally disclosed greater amounts of sexually deviant activity than the offenders not required to take a polygraph examination. This study also compared information known before and after a polygraph test. Again consistent with the research mentioned earlier, the findings suggest that significantly more information was gained from the polygraph test. In the sample of 180 sex offenders, for example, almost twice as many admitted to having male victims (20% to 36%) and three times as many admitted to perpetrating offences against both sexes (10% to 29%) after a polygraph test.

In a similar study, Hindman and Peters (2001) retrospectively compared the histories of sex offenders whose self-report were verified by polygraph with those who did not take a polygraph. Consistent with English, et al. (2000), polygraphed offenders reported on average more victims (13.6 compared to 2.5), less history of having been sexually victimised themselves (32% compared to 65%), and a higher incidence of having offended as juveniles (68% compared to 22%). A comparable pattern was also observed for juvenile sex offenders. Upon closer analysis, however, these findings are perhaps not unexpected, considering that some of the non-polygraphed offenders were seen only for an evaluation, whilst all the polygraphed offenders were seen for treatment and also offered immunity from prosecution for reporting prior unknown sex offences.



## *Monitoring of Sex Offenders*

Even less research has been conducted with regards to the use of the polygraph in the supervision of sex offenders. Abrams and Ogard (1986) compared the recidivism rates of probationers required to take periodic polygraph tests for supervision with probationers with no polygraph requirement. Whilst the study, did not specifically investigate sex offenders it was reported that over a two-year period 69% of men who received periodic polygraph examinations remained offence free, whereas only 26% who did not receive polygraphs successfully completed their supervision. Because participants were not randomly allocated and general criminality and previous treatment were not controlled for, it is unclear how much the polygraph actually contributed to a reduction of recidivism. The difference between the groups, for instance, may simply have been due to one being less criminal than the other.

In a study that specifically investigated sex offenders, Edson (1991) reported that 95% of 173 sex offenders on parole or probation, and who were required to undertake periodic polygraph testing did not re-offend over a nine-year period. Yet again the absence of a comparison group makes it difficult to conclude that the low recidivism rate is due to use of the polygraph for monitoring.

In a survey study, Harrison and Kirkpatrick (2000) asked a small sample of sex offenders ( $n = 28$ ) whether the polygraph assisted them in their treatment and with the problem of adhering to their probation conditions. A slight majority (57%) reported a general decrease in risk behaviours that they attributed to use of

the polygraph. This included having contact with potential victims and masturbating to deviant fantasies. Somewhat surprisingly, considering the small majority (in reality only 2 offenders), the researchers concluded that periodic polygraph testing had altered the offenders' attitude to treatment and supervision, which had consequently affected their behaviour in a positive direction (i.e. less high-risk and offending behaviour). In the larger survey study mentioned earlier, Kokish and Blasingham (2002) found that 72% of sex offenders reported that periodic polygraph testing was helpful with avoiding high-risk behaviours and re-offences. The results of these questionnaire studies provide some support for the view that the polygraph has therapeutic value in the treatment and management of sex offenders. As mentioned previously, however, self-report studies are complicated by numerous problems, including social desirability biases.

### *Factors affecting the polygraph's utility*

#### *Deterrence*

Research in the area of social influence through surveillance (Mc Guire, 1969) appears consistent with the reasoning that the threat of polygraph testing could act as a deterrent. If an individual believes, for example, that the likelihood of being caught by the polygraph is high, and that the consequences of being caught are undesirable, then an offender will be inclined to avoid the behaviours that are being monitored. This reasoning is similar to that which applies to the use of random drug tests as deterrents. The results of a study by the National Research

Council (1994) support the theory that random and frequent drug tests are associated with lower levels of drug use.

### *The bogus pipeline effect*

There is also indirect evidence within the realm of social psychology to provide some support for the value of post-conviction polygraphy. In a study by Jones and Sigall (1971), for example, students were asked various personal questions regarding racial attitudes while attached to a machine that they were told was a highly accurate ‘lie detector’ (it was not). Those questioned under this ‘bogus pipeline’ condition were more likely to disclose politically incorrect responses. It was concluded that because subjects believed that the machine was an accurate lie detector, and they felt it was in their interest to tell the truth (even though it reflected poorly upon them) rather than be caught in a lie. In a similar study, students were given information in advance on how to cheat on a classroom test. Sixty-five percent later admitted to receiving this information when connected to a bogus pipeline, compared to only 5% who did so without being connected (Quigley – Fernandez & Tedeschi, 1978). A meta-analysis review has recently supported the conclusion that subjects offer socially undesirable information because of their fear of being caught in a lie by the bogus pipeline and not just because of the expectations of the experimenters (Roese & Jamieson, 1993). Whilst the bogus pipeline research is not necessarily analogous with post-conviction polygraphy, it does demonstrate a likely placebo effect that the polygraph may have within a post-conviction context.



## *Confessions*

It is only in recent years that research has investigated the psychological aspects of confessions, examining this literature provides some insight into the reasons why sex offenders may disclose information during polygraph test rather than in other contexts. Gudjonsson (2003) reviewed five theories of what makes suspects confess to crimes they have committed during police questioning. These, combined with empirical evidence (Moston, Stephensen & Williamson, 1992), indicated that there are three main reasons why suspects confess to crimes they have committed (Sigurdsson & Gudjonsson, 1996a). These are suspects' perceptions of the strength of evidence against them, external pressure (e.g. fear of custody) and internal pressure (e.g. wanting to 'clear their conscience'). Usually more than one of these factors are present when a suspect confesses, but the empirical evidence suggests that it is the perception of the evidence against them that is the most powerful reason for confessing to the offence. In this regard, the polygraph in a post-conviction context may enhance the perception in the offender mind that the evidence is 'stacked against them', and thus, prompt them to 'confess'. This also raises the issue of false confessions.

## *False confessions*

False admissions during a polygraph test have not been addressed by any published literature. In an unpublished study, Kokish and Blasingham (2002) examined, amongst other things, 'defensive lying' (false admissions) in a sample of 95 sex offenders mandated to take periodic polygraph tests in California. In this

study participants were asked on an anonymous questionnaire whether they had ever made false admissions after they had failed a polygraph exam. Only six offenders (5 %) reported making false admissions at this time, suggesting that this issue is not a significant problem. This study did not provide details of what false admissions the offenders claimed to have made. If the offenders reported substantial numbers of additional victims and offences, it would challenge the findings from other studies demonstrating the polygraph's therapeutic value. All the same, questionnaire studies suffer from a variety of distortions, including social desirability bias. With regard to the study, therefore, despite the anonymous nature of the research, offenders may have had concerns that their responses could be traced back to themselves and consequently may have responded in an edited manner.

Within the wider empirical literature little research has been published on the issue of false confessions in criminal justice contexts. Gudjonsson and Sigurdsson (1994) and Sigurdsson and Gudjonsson (1996a, 1996b) carried out research into alleged false confessions among prison inmates. In both studies, 12% of inmates reported that they had made false confessions to the police sometime in their lives. Three main reasons were given for having made a false admission: 1) to protect someone else; 2) to avoid police pressure, and 3) to escape police detention. Comparison of alleged false confessors with other prison inmates showed that the 'false confessors' were significantly more personality disordered (Sigurdsson & Gudjonsson, 2001). Gudjonsson, Sigurdsson, Bragason, Einarsson and Valdimarsdottir (2004) reported similar findings in a more recent study with a large sample of students ( $n = 1050$ ). Only a small minority (3.7%), who had

previously been interrogated by police, claimed to have made false admissions to the police. Again the false confessors were significantly more personality disordered than the other volunteers.

Overall, whilst these findings suggest that false confessions are likely to be reasonably low in criminal justice contexts, individuals with personality disorder may be predisposed to making erroneous admissions. This is concerning, because sex offenders as an overall sample have generally been described as having a high prevalence and a broad range of personality disorders (PD) (Fraedrich & Pfaefflin, 2000; Fazel, Hope, O'Donnell & Jacoby, 2002). False admissions may be more of an issue, therefore, in sex offender populations, particularly when the polygraph is used in a therapeutic manner focussed upon an offender's disclosure of risk behaviour and offences.

### *Personality*

No research to date has explored the impact of personality on the utility (i.e. gaining additional information and acting as a deterrent) of the polygraph in post-conviction contexts with sex offenders. Numerous studies have shown that personality, particularly Antisocial Personality Disorder and Psychopathy, is strongly linked with both general and sexual recidivism (e.g. Gudjonsson & Sigurdsson, 2004; Hanson & Harris, 2000; McGuire, 2000). Thus, it could be speculated that these characteristics would be associated with reporting a greater number of risk behaviours, offences, and victims. It is also conceivable that more antisocial sex offenders would be less likely to view the polygraph as a deterrent.



It is valuable to consider the impact of nonpathological personality characteristics. A better understanding of these factors can assist with, not only, identifying *who* the polygraph works best with, but also *how* the instrument can best be used. Presently, the Five-factor model (FFM) of personality, or Big Five, is the predominant model in general personality research (Widiger, 2005). It is a particularly robust model and has succeeded in representing diverse collections of traits of alternative models within a single, integrative, hierarchical structure (O'Connor, 2002). The FFM has also received substantial interest as an alternative diagnostic system to current categorical approaches of personality disorder (Costa & Widiger, 1994; Widiger & Costa, 2002).

The FFM was developed through empirical studies of the trait terms within various languages (Widiger, 2005). This dimensional approach to personality description has emphasised five broad domains - Neuroticism; Extraversion; Openness to experience; Agreeableness; and Conscientiousness (Costa & McCrae, 1992a). Each of these domains has been differentiated into six more specific traits or facets. Neuroticism, for example, is composed of the facets: anxiety; angry hostility; depression; self-consciousness; impulsiveness; and vulnerability (Costa & McCrae, 1992b). Empirical support for the FFM is considerable, including convergent and discriminant validity across self, peer, and spouse ratings (Miller, Pilkonis & Morse, 2004), temporal stability over numerous years (McCrae & Costa, 2003) and cross-cultural replication (Ashton & Lee, 2001).

Proponents have also argued that the FFM represents a useful taxonomy for describing normal and abnormal personalities (Saulsman & Page, 2003, 2004). Numerous studies, using a variety of measures, have explored the relationship between the FFM and PD (Widiger & Costa, 2002). Generally the results from such research have supported the view that PD can be understood in terms of the FFM (e.g. Axelrod, Widiger, Trull & Corbitt, 1997; Blais, 1997; Miller, Reynolds & Pilkonis, 2004). Within the FFM, PD characterised by emotional distress show positive associations with Neuroticism (e.g. Borderline PD), while those characterised by shyness and reclusive qualities show negative associations with Extraversion (e.g. Schizoid PD). Disorders characterised by interpersonal difficulties show negative associations with Agreeableness (e.g. Antisocial PD), and those characterised by orderliness show positive associations with Conscientiousness (e.g. Obsessive-Compulsive PD).

Such findings demonstrate the empirical relationship between the FFM and PD, the research has, however, struggled to distinguish the specific PDs using the FFM. Morey, Gunderson, Quigley and Lyons (2002), for example, found it difficult to differentiate between Borderline, Schizotypal, Avoidant and Obsessive-Compulsive PDs, and suggested that these PDs are variants of the same FFM profile. Likewise, Saulsman and Page (2004) in a recent meta-analysis found that all PDs (except Dependent) displayed the characteristic FFM profile of high Neuroticism and low Agreeableness. They argued that these two domains could represent the core of personality pathology (Saulsman & Page, 2003).

With consideration of post-conviction polygraphy, it could be argued that because PD has been associated with sexual recidivism, neuroticism and agreeableness would also be associated with engaging in risk behaviours. The relationship between the FFM and PD has, however, not been replicated in sex offender samples using validated measures, only a single study has examined this relationship in a sample of adult sex offenders. Lehne (1994) compared the FFM domains with PD as measured by the Million Clinical Multiaxial Inventory (MCMI). The results were broadly consistent with other studies using non-forensic populations, showing strong associations between most of the PDs and the domains of neuroticism and agreeableness. The MCMI scales do not, however, directly translate to DSM diagnoses (Lehne, 1994).

If the relationship between PD and the FFM traits is replicated in sex offender samples, it could be expected that neuroticism and agreeableness would also be associated with engaging a larger number of risk behaviour. Even if this relationship were not apparent, however, these traits may still be associated with recidivism. The capacity to self-regulate and ability to follow through with risk reduction strategies are important for offenders wishing to change their behaviour (Marshall, Anderson & Fernandez, 1999). Thus, offenders with deficits in these areas, that is, are neurotic and antagonistic (i.e. low agreeableness), are unlikely to follow through with stated intentions, and are likely to fail to avoid high-risk situations and behaviours.

A single study has explored the relationship between FFM and general recidivism. Clower and Bothwell (2001), in a small sample of inmates, found that



low conscientiousness and openness to experience was associated with number of arrests. This suggests that people, who are careless, and engaging in criminal activity are more likely to be caught (i.e. low conscientiousness). They also argued that because openness to experience was associated with socioeconomic status, low scores may reflect a tendency to identify with countercultures that engage in criminal activity. Nonetheless, these results only accounted for 13% of the variance in number of arrests, suggesting that the effect of these traits were relatively minor. It is, of course, also difficult to generalise these findings due to the non-specific and small sample size ( $n = 51$ ).

## *Conclusions*

Proponents of post-conviction polygraphy make a persuasive case for its introduction in the management of sex offenders. When used in this context, the polygraph has the potential to overcome limitations in current supervision practices, as well as generally improving the assessment and treatment of sex offenders. Its use in such a capacity has increased markedly over the last decade in the US. This expansion is likely to continue as professional organisations and leading sex offender practitioners endorse its utility. Whilst enthusiasm to polygraph sex offenders is perhaps understandable, the research evidence demonstrating its value in these settings is, at best, limited. Much of this research is complicated by methodological problems, such as small sample sizes, retrospective methodologies, lack of comparison groups, social desirability biases and the use of non sex-offender samples. In spite of these issues, many of the claims made by proponents have face validity and appear consistent with theories

from social psychology. Yet, no matter how logically compelling a proposal may be, it is by data that we should be guided.

Another concerning feature of the literature on post-conviction testing is that very little of it has actually been published in peer reviewed scientific journals. Studies have typically appeared in probation journals such as *Interchange* and *Federal Probation*, or others, such as *Polygraph*, the journal for the American Polygraph Association (APA). Whilst these journals are reputable within their respective fields, they do not have a scientific orientation, and it is unclear whether contributions under-go the same level of scrutiny and independent review as would be expected in more traditional scientific journals. Perhaps for these reasons, the recently published review of polygraphy by the National Academy of the Sciences (2002) concluded that there exists no scientific evidence in any setting that demonstrates the value of polygraphy as a way of eliciting admissions or serving as a deterrent. The report does highlight, however, that the existing evidence supports the plausibility of such uses but surprisingly, it did not consider a single study involving post-conviction testing with sex offenders.

In addition to these issues, there is a concern over the emphasis placed on the utility, rather than the accuracy, of post-conviction polygraph testing. Whilst proponents rightly argue that these issues are theoretically distinct, they are very much related in practice. The utility of the polygraph as a deterrent depends inherently on the offender's belief that the procedure will reveal any deceptions and that the costs of being judged deceptive is undesirable. If the polygraph is not particularly accurate, then offenders' are not going to maintain a belief in its

accuracy over time. The continued value of this technique as a deterrent and as an aid to treatment would then greatly diminish. In such circumstances, overconfidence in the polygraph could create a false sense of security among supervisors, treatment providers, and the general public, that may in turn lead to inappropriate relaxation of other methods of supervision and treatment. Thus, if the polygraph is not highly accurate then the 'utility' argument becomes seriously compromised. The long-term usefulness of post-conviction polygraphy depends intrinsically, therefore, upon the procedure being an accurate method of detecting deception.



## CHAPTER 5

### Science and the Polygraph

To date only one study has attempted to investigate the accuracy of the polygraph as used in a post-conviction context with sex offenders. This was the previously mentioned survey by Kokish and Blasingham (2002). In this research the subjects were asked about the accuracy of the polygraph tests they had completed whilst on probation. In total 386 tests were identified and according to the offenders' self-report, in 343 of these they were 'truthful' and in 43 they were 'deceptive'. The polygraph accurately identified truth-telling 93% of the time and only wrongly accused offenders of lying on 7% of such occasions. Deception was accurately identified 74% of the time, and offenders claimed to have 'beaten' the polygraph 26% of the time. When considering these results it is unclear, however, whether the offenders were being deceptive to the comparison or relevant issues. The distinction between these questions are not made for the examinees it is possible that some may have believed that they had 'beaten' the polygraph when they were in reality 'lying' to the comparison questions.

In another study, Faller (1997) investigated the relationship between the polygraph results of individuals accused of sexual abuse (i.e. in a pre-conviction context) with other indications of abuse, such as victims interviews, psychological symptoms and medical evidence. The study found that while the polygraph result was not significantly related to any of these measures, passing the polygraph was, however, predictive of non-prosecution. Faller's study is not a validity study

because it lacks an objective measure of criterion validity (i.e. knowing who is really lying and truthful). It is also debatable whether such a relationship would exist. Because polygraph results can generally not be used as evidence, investigators are more likely to use the procedure when there is little other evidence to go by.

The wider scientific community remains deeply divided about the polygraph's accuracy. Some studies have reported high, near perfect, accuracy rates (e.g. Forensic Research Council, 1996). Other researchers have entirely dismissed these findings, claiming they utilised methodologies that are fundamentally flawed (e.g. Iacono & Lykken, 1997).

It is important to define a number of key psychometric concepts prior to evaluating the scientific evidence for the polygraph. The term 'validity' refers to whether an instrument measures the phenomenon that it is suppose to measure. In regards to the polygraph it refers, therefore, to whether the instrument is actually accurate in discriminating between truthfulness and deception. There are several aspects to validity that are important to distinguish. *Construct validity* refers to the theory of how and why a particular test works. A test is considered to have high construct validity when the theory underlying it provides a plausible explanation for the empirical findings (Grimm & Yarnold, 2000). *Criterion validity*, on the other hand, refers to the accuracy of the instrument. It is important to further differentiate two concepts related to accuracy. These are *sensitivity* and *specificity* (Swets, 1996). With regards to polygraph testing, *sensitivity* refers to how well the test actually identifies 'liars', while *specificity* refers to how well it

identifies ‘truth-tellers’. Examinees that are misdiagnosed are referred to as either *false positives* (examinees erroneously identified as deceptive) or *false negatives* (examinees erroneously identified as non-deceptive) (Swets, 1996). Inconclusive results are typically excluded when presenting the data. Table 5 presents the four possible combinations of polygraph test results.

Table 4.

*Diagnostic classification statistics*

	True Condition		Total
	Positive (Truly Deceptive)	Negative (Truly Truthful)	
Deception Indicated (Lying)	True Positive (TP)	False Positive (FP)	TP+FP
No Deceptive Indicated (Truthful)	False Negative (FN)	True Negative(TN)	FN + TN
Total (n)	TP / (TP + FN) <i>Sensitivity</i>	TN / (FP +TN) <i>Specificity</i>	TP + FN + FP +TN

The ‘percentage correct’ for *sensitivity* and *specificity* is the typical format for presenting data on the accuracy of the polygraph. Some early studies have utilised an overall percentage correct to describe the accuracy of the polygraph, but such a figure is essentially meaningless. Consider the example where base rates of deception are quite low, for example 10 out of 100. Simply by assuming that everyone is innocent and passing all the polygraphs would theoretically yield 90% ‘accuracy’. Even randomly failing 10 out of these 100 would still yield a respectable 82% ‘accuracy’.



Other related statistics that are of value when considering the accuracy of the polygraph are *positive* and *negative predictive power* (National Research Council, 2002). *Positive predictive power* refers to the percentage of individuals testing positive (deceptive) on the polygraph who are actually deceptive. This measure is obtained by dividing the true positive rate (total number of actual liars) by the total number of individuals who have tested deceptive. *Negative predictive power*, on the other hand, is the percentage of individuals scoring negative (non-deceptive) on the test who are truly negative. This statistic is calculated by dividing the true negative (total number of truth tellers) by the total number of individuals who have scored negative on the test (McCann, 1998). The value of these statistics is that it provides an indication of how confident the researcher can be that a 'failed' or 'passed' polygraph result is actually correct.

Some studies have recently applied scientific methods based on signal detection theory (National Research Council, 2002). A receiver operating characteristic (ROC) curve is used to assess the quality of the discriminatory power of a test using *sensitivity* and *specificity* data (Swets, Dawes & Monahan, 2000). The actual curve is a graphical display with the *false-positive* rate on the horizontal axis and the *true-positive* rate on the vertical axis. Such a curve allows one to identify how well a test discriminates between positive (deception indicated) and negative (no deception indicated) results and to identify the best trade-off between *sensitivity* and *specificity* when evaluating a test. A convenient overall quantitative index of accuracy is the proportion of correct identifications that is made by a polygraph examiner when confronted repeatedly by pairs of random test outcomes, one of which is deceptive and the other truthful. For

example, a polygraph examiner will make the correct choice eight out ten times when using a test with  $ROC = 0.8$ . Its possible range is from 0.5 at the ‘chance’ diagonal and 1 to ‘perfection’ (Swets, 1996).

Other important concepts related to validity to bear in mind when considering studies on the accuracy of the polygraph are those of *internal* and *external validity*. *Internal validity* refers to whether studies have been sufficiently controlled for extraneous variables that may affect the accuracy of the test, typically a prospective design and the presence of control groups enhances a study’s internal validity. The term *External validity* refers, on the other hand, to whether the findings from a study can actually be generalised to other contexts or circumstances (Grimm & Yarnold, 2000). *External validity* is enhanced when the testing conditions and the sample replicate the situations and subjects that the polygraph would be used with.

## Theories of Lie detection

The theoretical underpinnings of polygraph testing are important to consider for a number of reasons. Aside from explaining ‘how’ the polygraph works, a validated theory gives confidence in the robustness of the test results across examinees and settings. A theory can identify circumstances and situations where the accuracy of the polygraph is likely to be compromised, and thus can provide direction for improving the procedure.

Much of the early theory on polygraphy held the assumption that there existed a specific physiological profile for deception, in much the same way that Pinocchio’s nose grew in size when he was dishonest (e.g. Marston’s systolic blood pressure theory or Benussi’s breathing theory). While the polygraph instrument does measure physiological correlates of arousal, contemporary polygraph theory does not claim to measure ‘lying’ per se. The *CQT* was developed precisely because there is no such thing as a specific lie response (Lykken, 1998). Within some sections of the literature, an oft-levelled criticism of the polygraph is that it endeavours to measure such a thing. For example, in a relatively recent article by Bashore and Rapp (1993), the concern was raised about “the assumption that there is a physiologic response that is specific to lying” (p.5), while Steinbrook (1992) stated that “lie detection assumes that telling a lie causes a specific and reproducible physiologic responses” (p.122).

Perhaps a reason for the continued myth of the lie response is the wider media’s portrayal of the polygraph as a type of mind reading machine (or maybe a



‘magic lasso’). An example of this can be seen in the recent BBC 1 television program ‘Meet the Folks’. Here potential suitors to an individual are evaluated over a period of time by the parents. One of the methods used in this process is the polygraph. During this part of the show the suitor is seen attached to a polygraph instrument whilst the parents ask numerous questions (e.g. “Do you like our daughter?” or “would you be faithful to her?” or “have you ever cheated on anyone?”). The individual responds ‘yes’ or ‘no’, while the polygraph examiner (who is seated behind them) gives the thumbs up or down, presumably indicating the veracity of the individual’s responses. Such demonstrations imply that there is a specific lie response, or at least, suggest that stress is exclusively associated with deception.

Contemporary polygraph theories propose that individuals who are being deceptive or truthful show different patterns of physiological responses when their reactions to relevant and comparison questions are compared. An innocent examinee will display, for example, a greater reactivity to the comparison questions than to the relevant ones. By contrast, the guilty examinee has a larger physiological response to the relevant when compared to the comparison questions. In *CIT*’s the theory is slightly different. Here it is assumed that a subject will display an augmented physiological response to information that is familiar to them. These theories do not assume that there exists a specific lie response, be it behavioural or physiological or otherwise. Rather it suggests that there are autonomic correlates of concern (i.e. anxiety or arousal) and that guilty and innocent persons might be differentially anxious about relevant and other

questions. A number of theoretical accounts have been proposed to support the theory for the *CQT* and *CIT*.

### *Theories of Comparison Question Techniques*

The theory of ‘psychological set’ was initially developed by Backster (1963a) and has generally been put forward by many proponents as the basis of how the *CQT* works (D.Sosnowski, personal communication, January 20, 2004). It proposes that during a polygraph test, examinees will focus the greatest amount of attention towards the questions that reflect the most immediate threat to their well-being. It is argued that this increased attention leads to greater physiological responses. Backster claims that due to the pre-test interview the innocent examinee will be orientated, and thereby more physiologically aroused, to the comparison questions because they represent the most immediate threat to their well being. Deceptive individuals, on the other hand, will be most threatened by the relevant questions, even though they will be lying to the comparison questions, and thus, will display greater physiological responses to these. The ‘Easterbrook’ Hypothesis supports the underlying theory of psychological set. This theory has demonstrated that attention is correlated with increased physiological arousal, especially to threatening stimuli (Easterbrook, 1959).

A variation of this perspective is the ‘threat of punishment’ theory proposed by Davis (1961). This theory postulates that lying is an active attempt to avoid anticipated serious consequences. It is argued that this ‘fear of punishment’ causes the deceptive individual to display greater physiological reactions to the

relevant questions, whilst the converse is true for the innocent subject. Ekman and Sullivan (1991) described this phenomenon as ‘detection apprehension’, and proposed that it is greatest in circumstances where the stakes are high (i.e. being caught in a lie is very undesirable). Another variation is Gustafson and Orne’s (1963) ‘motivation theory’, which postulates that individuals whose highly motivated to lie will experience an elevated physiological response to the relevant questions because of their motivational state (to lie) at those times.

All these theories suggest that detecting deceit will be most robust in real-life contexts where the consequences are authentic and the individual cares about the outcome of the test (i.e. they don’t want to be caught in a lie). The problem, however, with these theories is that under certain circumstances they predict that some innocent subjects will have a greater response to the relevant questions than will deceptive examinees.

Because no specific lie response has been identified the validity of the *CQT* is largely dependent on the examiner during the pre-test interview. He manipulates the individual into a situation where he lies, or at least feels very threatened by the comparison questions. Vendemia (2003) describes this as the ‘Hobson’s Choice’ situation, after Thomas Hobson, a stable owner in the 16<sup>th</sup> century who offered customers the horse nearest to the door or none at all (in effect a ‘no-choice’ option). Similarly examinees are put in a situation where being truthful is not an option, because the examiner will constantly re-define the boundaries of the questions during the pre-test interview, until they are being deceptive or at least very unsure about their answers. Doing this, in theory, increases the probability



that innocent examinees will display an augmented physiological reaction to the comparison questions when compared to the relevant ones. An apparent problem with this reasoning is, however, that the examiner does not know who is telling the truth and cannot determine, therefore, whether the innocent are suitably threatened by, or indeed lying to, the comparison questions. Consider an innocent man suspected of sexually abusing his own child. In such a situation, it could be expected that he could react to the relevant questions for numerous reasons, such as being angry about the abuse, or being fearful that he is going to be accused of it. He could also react to the relevant questions because he is simply traumatised and upset that the abuse has occurred in the first place. Similar concerns have been raised elsewhere (Ben-Shakhar, 2002; Furedy, 1996a; Lykken, 1998; Patrick & Iacono, 1991).

There is also the concern that deceptive individuals could be potentially over-threatened by the comparison questions. An individual that has an extensive history of antisocial and criminal behaviour, for instance, could conceivably find comparison questions such as ‘have you ever lied to someone that trusted you?’ or ‘have you ever cheated someone that trusted you?’ more threatening than the relevant questions. Similarly some critics argue that the task of developing comparison questions that are differentially sensitive to the inferred discomfort of the innocent person, vis-à-vis the discomfort invoked by relevant questions, are unrealistic (e.g. Bashore & Rapp, 1993; Furedy, 1993). Abrams (1999: p. 24), a proponent for polygraph, concurs:

‘There is a delicate balance between the comparison and relevant questions and many variables can tip this balance in either of those two directions. Too much discussion of one or the other during the pre-test.’

The directed-lie comparison question goes some way to overcoming such difficulties. The directed-lie questions remains controversial, however, even with proponents of polygraph. Abrams (1999), for example, again argues that placing excessive emphasis on these questions increases the probability of false negative outcomes. Consistent with this reasoning, Iacono and Lykken (1999) suggest that explaining the purpose of the direct-lie questions during the pre-test clarifies to the examinee the importance of giving strong responses to these questions, and thus inadvertently makes the test easier to pass.

In response to these concerns, proponents have generally argued that the pre-test interview serves to allay many of the concerns and fears of the examinee, whilst also clarifying the importance and seriousness of the relevant questions. The American Polygraph Association’s standards of practice require examiners to spent ‘sufficient time to identify the issues and potential problems’ during the pre-test interview ([www.polygraph.org/standards](http://www.polygraph.org/standards)). Others have proposed that for many innocent examinees, the opportunity to ‘prove’ their innocence with the polygraph is often ‘enthusiastically’ taken up, rather than avoided (Harrelson, Gerow & Gerow, 1998). It is argued consequently, that truthful individuals will be primed to respond to the comparison rather than to the relevant questions, because these will be the most salient for them. For deceptive individuals, in other hand, the relevant issues will be most salient, and thus, will be more physiological



arousing. This seems, however, a somewhat tautological position, and while it may be plausible, it presents a simplistic view of human nature, and it is arguable whether or not such a response would be the norm for all individuals asked to take a polygraph test.

Another theory put forward to explain the underlying mechanisms of the *CQT* is the 'conditioned response' theory (Davis, 1961). This theory postulates that relevant questions represent conditioned stimuli that evoke physiological responses in deceptive individuals. There is considerable evidence demonstrating that autonomic responses can be classically conditioned (Diven, 1937; LeDoux, 1995). It is argued that the recollection of the issue under investigation, elicited by the relevant question, acts as a conditioned stimulus for deceptive individuals that elicits a minor autonomic response. Innocent individuals, because they never under-go this conditioning, do not display a similar physiological reaction to the relevant questions. The problem with this theory is that, again, there is the possibility for a polygraph examiner to misinterpret an examinee's truthfulness. Because lying is not the only possible elicitor of an autonomic response, innocent individuals may display physiological reactions due to some other feature of the relevant question. It is possible that innocent individuals who previously have been falsely accused of the issue under investigation, or believe that they are being or are about to be accused, could developed a conditioned response to the issues being investigated. In such a situation, the relevant question would act as a conditioned stimulus for anger or fear.



## *Theories of Concealed Information Test*

Orienting Theory has been used to explain the underlying mechanisms for *Concealed Information Tests (CIT)*. The 'orienting response' has been described as the 'what is it?' reaction that occurs in response to novel, or personally significant stimuli (Sokolov, 1963). This phenomenon can be illustrated by the example of an individual at a large social gathering overhearing his name mentioned in another conversation. Individuals will typically focus their attention towards the source of the stimulus, whilst ignoring other stimulus. Lynn (1966) identified the physiological profile of an orientating response as including decreased heart rate, an increased skin conductance, and a general muscle tone, and also, a vasoconstriction in the limbs. Orientating theory predicts, therefore, that in *CIT* formats, examinees will display involuntary physiological reactions to familiar or personally significant information.

This theory is not, however, without problems. The assumption that individuals will display a physiological response only to familiar and important stimuli appears improbable in some circumstances. For example, an orientating response could generalise to other *similar* categories of relevant stimuli. Consider an individual implicated in a gun crime, who also owns an illegal gun. Such an individual may display enhanced responses to a variety of questions related to guns, even though he has no information about the actual murder weapon. An examinee's response in those circumstances would indicate 'guilty knowledge', even though it was actually due to a stimulus generalisation response.

## *Conclusions*

All the theories described above are derived from the principles of Arousal Theory, which holds that the stronger the stimulus, the stronger the psychological reaction, and the more pronounced the physiological response. In polygraph testing because there is no characteristic physiological ‘lying response’, examiners make inferences regarding an individual’s psychological state (i.e. whether he is lying or truthful or concealing knowledge) based on his physiological reactions to various questions. The proposed theories go some way to providing plausible explanations for why a deceptive examinee may respond in a particular manner on a polygraph test. Orienting Theory which seeks to describe the underlying mechanisms for the *CIT*, is perhaps the most robust among these. All these theories have, however, worrying limitations and suggest that under certain circumstances truthful individuals will appear deceptive.

It is important to bear in mind that little research has actually been conducted that validates any of these theories. This was initially pointed out by Guertin and Wilhelm in 1954, who stated, “there is relatively little theoretical evaluation of the processes underlying the responses to the lie detector (p.158).” The Office of Technology Assessment (1983) arrived at an identical conclusion almost thirty years later stating that: “the basic theory of polygraph is only partially developed and researched” (p.6). This concern is again raised in the most recent review by the National Academy of Sciences (National Research Council, 2002: p.

2): “research has not developed and tested theories of the underlying factors that produce the observed responses in the polygraph”.

Much criticism of the polygraph, especially the *CQT*, has focussed upon the underlying theories. When considering such criticism it is important to account for the difference between negative evidence (i.e. research that disproves a theory) and a lack of evidence. Because the research has not been conducted one way or another, many of the criticisms made by opponents, while plausible, remain unsupported, as of course do, many of the claims made by proponents. Just because polygraph theory has not been adequately researched, however, it does not necessarily follow that the polygraph is an invalid method of lie detection. As Charles Honts declared in a presentation for the National Academy of Sciences panel: “I don’t know how it (the CQT) works, but I know it does.” ([www4.nationalacademies.org/weber.usf](http://www4.nationalacademies.org/weber.usf)).



## Accuracy of the Polygraph

In the wider scientific literature the accuracy or criterion validity of the polygraph remains a divisive issue. Suitably reflecting the polemic and polarised nature of the literature are two quotes from the same publication in a recently published book that arrive at vastly different conclusions regarding the accuracy of the *CQT*. In the initial review, Raskin and Honts (2002) deduced that, “the voluminous scientific literature indicates that they (polygraph examinations) can be highly accurate when properly employed in appropriate circumstances” (p.38). On the other hand, Ben-Shakhar (2002) concluded, “there is neither sound empirical nor research-based evidence that indicates any kind of validity for the *CQT*” (p.118). Much of the variation of these conclusions is based on differing opinions regarding the value of the two different methodologies used to evaluate the accuracy of polygraph, field and laboratory studies.

*Field studies* are retrospective evaluations of the accuracy of ‘real-life’ polygraph tests conducted on criminal suspects. In these studies, the result of a polygraph test is compared with some other indicators of the ‘truth’, such as judicial outcomes. The two main problems with this approach are identifying who is actually lying, and then the bias introduced by the retrospective nature of the evaluation. *Laboratory studies* overcome these limitations by assigning volunteers to deceptive or non-deceptive categories. Because these studies involve simulations of crimes with volunteers who know they are participating in ‘lie detection’ research, it is debateable, however, whether the findings can be generalised to real-life contexts (Ben-Shakhar, 2002; Honts & Perry, 1992).

### *Research evidence from field studies*

As mentioned previously, field studies are ‘naturally’ occurring polygraph-testing situations. In these studies, the researcher has no experimental control over the situation and subjects cannot be systematically assigned to conditions of deceptive or truthful. To determine, therefore, whether the polygraph is accurate, the test result is compared with retrospective evaluations of truthfulness. This can include actual judicial outcomes, decisions by a panel of legal ‘experts’ assembled specifically for the study or a confession by the presumably guilty party (Honts & Perry, 1992; Office of Technology Assessment, 1983). These methods for determining the truth, however, are problematic for a variety of reasons:

- Individuals can be wrongly convicted or acquitted. Also convicted offenders often plead guilty to lesser charges or are convicted of different charges;
- Panel decisions can suffer from a sampling bias, because cases cannot be randomly selected since there needs to be enough information available for the panel to make a decision (Saxe, Dougherty & Cross, 1985). Research has also cast doubt on the accuracy and value of panel decisions for establishing criterion validity. In a study by Dohm and Iacono (1993), panels consisting of police investigators, attorneys and laypersons, were provided with information from independently verified cases and were asked to make a determination of guilty or innocent. The results indicated

that the panel decisions were not accurate. The researchers concluded, more generally, that panel decisions were unlikely to establish criterion validity in field studies with any degree of certainty;

- Using polygraph elicited confessions can also be problematic. The Polygraph examiner's typically elicits confessions when they confront an individual in the post-test interview after a 'failed' test result. Subjects that 'pass' are unlikely to be confronted in a similar manner. Thus deceptive examinees that pass the polygraph (i.e. false negative's) are excluded from the research. Similarly, innocent individuals wrongly accused of lying (false positive's) are also excluded from the research, because there is no confession. Examinees that confess in the pre-test interview also cannot be included in the research because they did not complete the test. Reliance on an examinees confession as a criterion for accuracy is, therefore, likely to artificially inflate overall accuracy rate.

In a determined attempt to overcome these problems Patrick and Iacono (1991a) conducted a field study with the Royal Mounted Police in Canada. In this research, 402 polygraph cases were evaluated. This represented the total number of polygraph tests within a designated region over a 5-year period. Rather than rely on polygraph-elicited confessions or panel decisions to determine accuracy, the researchers searched the suspects' police file for evidence of their guilt or innocence that emerged after they had completed the polygraph test. This included



non-polygraph-related confessions, additional medical evidence or confirmation that the crime had actually not occurred (e.g. finding the 'stolen' property or false accusation). By using such stringent criteria for establishing criterion validity in only one case could the guilt of a suspect be independently verified. The researchers speculated that this small number reflected how the police tended to use the polygraph in their investigations. Suspects were typically only asked to undertake a polygraph exam when there was not enough evidence for a conviction and the investigation was unable to progress further. Because polygraph evidence cannot be submitted as evidence in Canada, they reasoned that the police hoped that the polygraph would lead to a confession, and thus resolve the issue at hand.

In another study that also attempted to overcome the problems associated with establishing criterion validity, Raskin, Kircher, Honts and Horowitz (1988) considered 2,522 polygraph tests administered by the US Secret Service. In addition to using polygraph-elicited confession, the study also required the presence of independent corroboration of the confession by some form of physical evidence. Only 105 (4%) cases were identified as suitable in total. The use of corroborating evidence in this manner does not, however, overcome the inherent problems of using confessions to establish criterion validity, because gaining the additional evidence is not independent from the original test outcome. Had the examinee not failed the polygraph, there would have been no post-test interview, thus no confession, and therefore, no opportunity to recover the physical evidence. In this manner, the corroborating physical evidence is dependent (like a confession) upon the examinee failing the test.

Finally, another concern with field studies is that due to their retrospective nature, the internal validity is likely to be low. Extraneous factors that may affect polygraph outcomes cannot be controlled. It is conceivable, for instance, that the examiner's decision can be 'contaminated' by other information. In real-life criminal investigations polygraph examiners are typically aware of, and have access to, other information related to the suspect. This could include victim statements, medical evidence, police intelligence, impressions of other investigators and records of previous convictions (Saxe, Dougherty & Cross, 1985). Thus it is unclear whether an examiner's decision is based solely upon the physiological data gained during the polygraph test.

So while field studies provide the most direct evidence regarding polygraph accuracy in real-life settings, the value of these findings is limited due to problems with criterion and internal validity.

The scientific merit of field studies can be appraised by considering four factors: sample; sampling strategy; polygraph evaluations; and criterion development. Firstly, the sample must represent the population for generalization. Secondly, cases should be selected randomly without reference to the accuracy of the original examiner's decision or to the quality of the physiological data. Thirdly, the outcome of the polygraph should be exclusively determined by the physiological data, with information about the case facts or other pertinent details withheld from the examiner. Finally, the truthfulness of the subjects must be confirmed by a criterion independent from the outcome of the polygraph examination. It is important to recognise that whilst these criteria may go some

way to improve the quality of field studies, they do not overcome the inherent problems associated with these studies (Honts & Quick, 1995).

In the review conducted by the Organisation of Technology Assessment (1983), nine field studies evaluating the *CQT* and one study investigating the *RIT* were considered. All polygraph tests had been conducted with suspects in criminal investigations for a range of crimes including sexual assault, theft and homicide. Two studies used panel decisions for determining criterion validity, while the remainder used confession. Unfortunately, none of the studies met all the previous mentioned criteria. Kleinmuntz and Szucko (1984), for instance, used 'trainee' examiners and only completed one polygraph chart, as opposed to the standard three. A number of other studies used global chart evaluations rather than blind numerical scoring procedures (Bersh, 1969; Hunter & Ash, 1973; Slowik & Buckley, 1975; Wicklander & Hunter, 1975).

The sensitivity (i.e. correctly identifying deceptive examinees) of these studies ranged from 70 to 98%, with a mean of 86% and a mean false negative rate of 11%. Specificity (i.e. correctly identifying truthful examinees) displayed greater variability with a range from 12 to 94%, the overall mean being 76%, and a mean false positive rate of 19%. Inconclusive polygraph results for both liars and truth tellers ranged between 0 and 25%, with the means being 2% and 5% respectively. Further analysis showed that a 'failed' polygraph result was accurate 82% of the time (i.e. the positive predictive value), whilst a 'passed' polygraph test was correct 87% of the time (i.e. the negative predictive value).



Since this time, an additional four studies have been conducted that met the abovementioned criteria (Honts, 1996; Honts & Raskin, 1987; Patrick & Iacono, 1991; Raskin, et al., 1988). As proposed earlier with the research by Patrick and Iacono (1991) and Raskin et al. (1988) these studies are not without problems. When considered together the sensitivity for these studies, similar to the OTA review, ranged from 73% to 100%, with the mean being 89% and false negative rate only one percent. Specificity was only 59% with the range being between 30% and 83%, and the false positive rate 10%. It was also notable that there was a significantly higher rate of inconclusive test results for ‘truth tellers’ when compared to ‘liars’ (29 versus 10). Ninety percent of failed polygraph tests were correct, whilst 98% of the passed polygraph results were correct.

The National Research Council (2002) evaluated 194 studies and identified only seven field studies evaluating the *CQT* that passed their minimal scientific criteria. This included six studies on the *CQT* and one study evaluating the *CIT*. The inclusion criteria used by this review, in addition to the previously described conditions, also required studies to report the questions, physiological measures, instrumentation and scoring methods used. Again a large variation in the accuracy was observed between the different studies with the Receiver Operating Characteristic (ROC) being between .71 and .99, and the median value being .89.

### *Research evidence from laboratory studies*

*Laboratory studies* are investigations where field methods of polygraph testing are used in simulated, usually criminal, situations. In these studies, the

polygraph is typically used to investigate either “mock” crimes set up by an experimenter with the knowledge and collaboration of some of the subjects, or actual small crimes “induced” by the experimenter (Kircher, Horowitz & Raskin, 1988; Patrick & Iacono, 1989). In these studies the consequence for being ‘caught’ lying usually takes the form of losing the chance to gain a monetary reward. Subjects are occasionally threatened with punishment if their lie is detected, but for ethical reasons these punishments are relatively minor (e.g. loss of course credit for participating in research). Such situations are thus not actual ‘criminal investigations’, and subjects, usually university students, are aware that they are participating in research. The value of this methodology is that criterion validity is assured and internal validity is high (Forman & McCauley, 1986).

Critics, such as Fiedler, et al., 2002 have generally dismissed the value of such studies for estimating field accuracy because of the substantial differences that exist between laboratory settings and real-life situations. In the laboratory, the guilty subjects usually do not have much at stake, and are aware that, whatever the result of the polygraph test, they will be released from further involvement with the ‘crime’. By contrast the stakes are higher in real-life scenarios, and the examinee is likely to be greatly concerned about the outcome of the polygraph test, whether he or she is guilty or innocent. The emotional ‘environment’ between the two testing situations is, therefore, vastly different.

Many researchers assert, however, that the difference between laboratory settings and real-life circumstances is not that great (Anderson, Lindsay & Bushman, 1999; Kircher, Horowitz & Raskin, 1988; Pollina, Dollins, Senter,



Krapohl & Ryan, 2004). It is argued that if laboratory studies use representative subject populations, realistic polygraph testing procedures, 'blind' chart evaluation (i.e. the scorer is not aware of other information related to the examinee) and also involve some meaningful motivation to deceive the polygraph examiner, they can provide useful information for estimating field accuracy.

Two studies have attempted to empirically evaluate whether examinee 'motivation' affects the accuracy of the polygraph. Elaad and Ben-Shakhar (1989) found that examinees who were motivated by a financial incentive were more likely to be detected than were subjects who were not similarly motivated to pass a *CQT* polygraph test. In a later study Honts and Carton (1990), however, reported no difference in the detection rates of motivated and non-motivated subjects.

The OTA (1983) review considered 20 laboratory studies. These included 14 studies that evaluated the *CQT* and six studies that investigated the *CIT*. The majority of these studies had been conducted with university students, four used paid volunteers, and single studies used inmates, psychiatric patients or police candidates. For the *CQT* sensitivity was 64% (range 37 to 100), and the false negative rate 10% (range 0 to 29). Specificity was 57% (range 33 to 100), and the false positive rate 14% (range 0 to 50). Inconclusive polygraph results for both liars and truth tellers ranged between 0 and 50%, with average rates being 22% and 27% respectively. Two studies had notably less than 12 subjects in the lying condition (Dawson, 1980; Ginton, Netzer, Elaad & Ben-Shakhar, 1982). The positive predictive value for these studies was 82%, whereas the negative predictive value was 85%.



In regards to the *CIT*, the sensitivity ranged between 61 and 95%, with the overall mean being 69%. The average false negative rate was 31% (range 5 to 39). Specificity was 91% (range 80 to 100), and the average false positive rate 6% (range 0 to 18).

Since this review, a number of laboratory studies have been completed that are worthy of mention. Patrick and Iacono (1989) evaluated the accuracy of the *CQT* with incarcerated inmates as subjects in a threatening context. In this study, the participants were led to believe that their personal performance could decide benefits or penalties for the sample as a whole. They were told specifically, that if they failed the group would lose all rewards, and they would be named to the other participants. Patrick and Iacono reasoned that because of the potential for physical violence, participants would be highly motivated to pass the polygraph test. Overall, the study reported a high sensitivity rate of 92% and a low false negative rate of 2%. Innocent subjects were, however, only correctly identified in 30% of cases, and were wrongly accused of lying 24% of the time. Inconclusive rates were also greater for innocent subjects (46% versus 6%). This data suggest that the increased threat might be critical factor in false-positive and inconclusive rates. In another study, Forman and McCauley (1986) permitted their volunteers to choose themselves whether they were deceptive or truthful. They reasoned that this manipulation was analogous to an opportunistic crime situation, such as an unlocked car with a valuable item, and the individual must decide whether to take advantage of the opportunity. Again the result indicated a good sensitivity (89%),

but, about half of the truthful subjects (47%) were erroneously classified as deceptive.

Honts and Peterson (1997) filed an *amicus curiae* with the Supreme Court in the US. They reported on nine laboratory studies of the *CQT*, and consistent with the earlier OTA review, there was large variation in the rates reported by the different studies, with sensitivity ranging between 53 to 100% ( $M = 80\%$ ) and specificity between 75 to 90% ( $M = 84\%$ ). The mean for both false negative and positive rates was 8%, with the positive predictive value being 90% and the negative predictive value 91%.

The NAS study considered 52 studies, investigating the *CQT*, *RIT* and *CIT* formats. There was no statistical difference between these testing formats in terms of their ability to discriminate truthful and deceptive subjects. This review included many of the studies also used in the OTA and the Honts and Peterson (1997) reviews. In addition, various unpublished studies by the Department of Defense Polygraph Institute (DoDPI) were also included. The results indicated overall that the accuracy ranged between .81 and .91, with the median being .86 (National Research Council, 2002).

### *A good accuracy study*

To date there is only one study that appears to have overcome the inherent limitations of both field and laboratory studies. In this study, Ginton, Daie, Elaad and Ben-Shakhar (1982) gave an aptitude test to 21 policemen taking part in a

course. The participants were led to believe that this test was potentially important for their careers. The researchers set-up the situation so that each participant had an opportunity to 'cheat', not knowing that the researchers could identify the ones that did so. In total seven participants cheated. All 21 subjects were subsequently recalled and told that they it was believed some of them had cheated on the test and all would be required to complete a polygraph exam. They were further told that the results of the polygraph could affect their careers. Three of the seven cheaters confessed immediately, another two cheaters and one innocent participant refused to take the polygraph. In total, therefore only 15 of the 21 original participants took the polygraph test – two cheaters and 13 non-cheaters.

The original examiner, utilising both chart information and the behaviour of the participant during the test, correctly identified both cheaters and also 11 of the 13 non-cheaters. Two truthful participants were wrongly accused of cheating. Blind re-scoring of these charts by eight other examiners correctly identified the two cheaters 15 out of 16 times. For the innocent participants, there was greater variability. On average these participants were correctly identified 81% of the time, and wrongly accused of cheating 17% of the time. With consideration of the original examiner's decisions even though his false negative rate was 0 and his false positive rate was 17%, when he said an individual was deceptive he was only correct 2 out of 4, or 50%, of the time. Interestingly, one of false-positives when interviewed after the study, reported that he had been focussing (unintentionally) upon a previous incident of cheating (A.Ginton, personal communication, 4<sup>th</sup> August 2004). This suggests that it is perhaps the saliency of an issue that contributes to an individual's reacting to the relevant issue.



The small subject numbers in Ginton's study make it impossible to say anything about the overall accuracy of the polygraph, however, the study does illustrate the inherent difficulties associated with evaluating the polygraph.

### *Surveys of expert opinion*

It is worth noting that proponents and opponents have also presented findings from various surveys as evidence of the polygraph's value or lack thereof (Iacono & Lykken, 1997; Raskin, et al., 1997). A telephone survey by the Gallup Organisation (1984), for example, found that 62% of the polled members (n = 152) from the Society of Psychophysiological Research believed that the polygraph was a 'useful diagnostic tool when considered with other available information'. Only 3% believed that the polygraph as a method of lie detection was of no use. Amato and Honts (1994) replicated these findings in an identical survey study. By contrast, two later surveys conducted by Iacono and Lykken (1997), also with members from the Society of Psychophysiological Research, found that only 36% in one, and 30% in the other agreed that, "the CQT is based on scientifically sound psychological principles or theory" (p.430). This apparent discrepancy in profession opinion may, however, be due to their respective foci rather than reflect a shift in attitudes.

The earlier surveys conducted by the Gallup Organisation and Amato and Honts addressed the utility of the polygraph as an aid to decision making while the latter studies by Iacono and Lykken focused upon the underlying theory of the

CQT. When considering the issue of utility two broad conclusions can be drawn from the literature on ‘lie detection’: Firstly, people are poor lie-catchers and are generally not able to detect deception at a rate better than random chance (Vrij, 2004); secondly, despite the confusing literature, there does seem to be a general trend that under normal circumstances, the polygraph produces better than chance success rates, and, in some circumstances actually appears to be quite accurate. It would seem reasonable, therefore, to conclude that *if nothing else* the polygraph could be a helpful aid in lie detection. With regards to the theory of the CQT, however, as discussed earlier, these are poorly developed and have not been subjected to sustained or rigorous scientific investigation. It would be accordingly difficult to endorse the view that these are based on “scientifically sound psychological principles or theory”. Taken together, it would seem, therefore, that comparing these surveys, as evidence of one view point or another is indeed analogous to comparing ‘apples with oranges’.

### *Unpublished evidence*

Finally, it is valuable to consider unpublished and anecdotal evidence on the polygraph’s accuracy. It is conceivable, for instance, that due to non-significant findings studies have not been published. In an early study which considers such evidence Abrams (1973) reviewed reports of the polygraph’s accuracy dating to 1917, these included a large number of unpublished and anecdotal reports as well as experimental data. He calculated approximate estimates of overall accuracy from this data, noting, however, “it is almost meaningless to total and average these findings because of the great discrepancy in

experimental paradigms and the instruments employed (p. 316)". Nonetheless, he reported that in studies where criterion validity was assured, the polygraph was 100 percent correct. In field reports prior to 1963 Abrams calculated an accuracy rate of 98 percent. In laboratory experiments prior to 1963, Abrams estimated the average accuracy rate of 81 percent. Averaging the results of the reports between 1963 and 1973, Abrams estimate of laboratory and field research accuracy was 83 and 98 percent, respectively. In reality, these data are of limited value because there is such variability in the technique and instruments used, the subjects tested, and also the criteria for establishing ground truth. In addition, many of the biases discussed above are likely to have inflated the accuracy observed in the reports that Abrams included in his review.

The more recent review by the National Academy of Science (2003) utilized a particularly vigorous methodology to gain as much as possible of the published and unpublished polygraph accuracy literature. In addition, to the more traditional literature searches using computerized databases, such as Psycholit, Social Science Citation Index, Medline and so forth, they sent requests to a variety of individuals and organizations that they believed may have relevant research reports. These requests were sent to government organizations and agencies, polygraph websites, and various researchers. The individuals contacted were also asked to forward their request to other professionals who may have potentially useful information. Indeed the NAS review deliberately attempted to limit the possibility of a 'publication bias' by being as inclusive as possible, and sought out studies regardless of whether or not they had undergone peer review.



In total 217 reports were collected of which 194 were identified as unique studies. Notably they highlighted that these studies varied greatly in the quality of research design, choice and standardization of measurement techniques, thoroughness of control of confounding variables, statistical analyses, and various other factors. They evaluated these studies using the criteria describe earlier (i.e. documentation to allow sufficient replication; independent verification of truth, blind scoring, and so on) and found that only 57 studies met their basic scientific criteria. Of these, 31 studies were gained from peer review journals, 7 were either unpublished studies or dissertations and 19 were reports submitted to government agencies or the like (i.e. Department of Defense Polygraph Institute). Taken together, these findings suggest that there is not a great deal of unpublished research on the polygraph's accuracy that is of sufficient scientific merit to contribute to the debate.

## *Conclusions*

The review of laboratory and field studies indicates that there is considerable variability in the accuracy of the polygraph. Much of the data appears in reports that are not peer reviewed. The results generally appear to suggest that while the polygraph is more accurate than 'chance', it is, far from perfect. Yet the quality of the studies varies considerably, and falls far short of what would be ideal. *Field studies*, for instance, cannot guarantee criterion validity and have problems with internal validity, whilst *laboratory studies* have problems with external validity. Survey studies evaluating the opinions of scientists in relevant fields, have further illuminate of the lack of agreement, and

perhaps, confusion within the scientific community regarding the value of the polygraph.

The large variation in the accuracy rates reported by the various studies is particularly noteworthy. This variation is likely to be due to a combination of factors, including differences in the behaviour and personality of the examinees and examiners, and variation in the testing circumstances and formats. Despite attempts to standardise polygraph-testing procedures, each test has a number of unique elements such as the pre-test interaction between the examiner and examinee, which theoretically could affect the test outcome.

## Factors affecting Polygraph accuracy

### *Countermeasures*

Countermeasures are deliberate techniques used by examinees to produce a 'truthful' outcome on a polygraph test. The major types of countermeasures include drugs and alcohol, mental (such as meditation) and physical strategies (such as inducing pain). Numerous publications (e.g. Maschke & Scalabrini, 2000) and web sites (e.g. [www.antipolygraph.org](http://www.antipolygraph.org)) provide detailed information about how such strategies can be utilised to produce a non-deceptive polygraph result. As Maschke and Scalabrini (2000: p.68) suggest when referring to the CQT, "... the key to passing a polygraph test ... is to produce stronger physiological responses when answering control (comparison) questions than when answering the relevant questions." Most of the physical countermeasures described in the literature are consistently designed to augment responses to comparison questions. For instance, there are a number of subtle ways for examinees to induce pain when responding to these questions, such as by pressing on a tack placed in their shoe. Doing so during the data collection phase of the test would theoretically mimic the physiological profile of a truthful examinee.

Mental countermeasures have also been suggested as a method for enhancing responses to comparison questions. Maschke and Scalabrini (2000) suggest using arousing mental imagery during the comparison questions, such as thinking about falling off a cliff.



Numerous studies have also demonstrated that countermeasures can reduce the accuracy of the CQT. Honts, Raskin and Kircker (1994), for example, trained 80 subjects to apply physical means (biting tongue) or mental countermeasures (counting backwards from 100 by sevens), while the comparison questions were being presented during their examinations. The mental and physical countermeasures were equally effective. Each enabled 50% of the subjects to erroneously pass the polygraph test (i.e. 42% of countermeasure subjects correctly classified compared with 75% guilty non-countermeasure subjects). The researchers reported, moreover, that the countermeasures were difficult to detect through observation. These findings are consistent with other studies in the area suggesting that training examinees in a combination of physical and mental countermeasures can substantially decrease the likelihood that deceptive subjects will be detected by the polygraph (Honts, 1986; Honts, Devitt, Winbush & Kircher, 1996; Honts, Hodes & Raskin, 1985; Honts, Raskin & Kircher, 1987, 1994). All these studies have involved only short periods of training for examinees.

Like the research investigating accuracy, there are a number of important limitations worth noting. All the studies have involved mock crime scenarios where the examiners are researchers, and the examinees are volunteers. Again it is questionable whether the results can be applied to real-life circumstances, where the examiner is likely to be a skilled and experienced interviewer, and the consequences of failing the test for the examinee are high. It is possible, therefore, that the results from these studies overestimate the effectiveness of countermeasures.

A curious omission in the research on countermeasures is that no study has attempted to investigate the physiological correlates or features of the actual countermeasures. It is possible that utilising countermeasures such as inducing pain would produce specific physiological profiles that are distinct from other physiological responses. Barland (2003) has presented anecdotal evidence suggesting that specific countermeasures display unique physiological features.

In response to the issue of countermeasures, polygraph practitioners have generally claimed that a skilled examiner is capable of identifying when an examinee is attempting to influence the polygraph recordings through movement or other methods (D. Sosnowski, personal communication, January 24<sup>th</sup>, 2004). This claim would appear more credible if there was actual research demonstrating the known physiological features for specific countermeasures. The polygraph manufacturer, Lafayette, have recently developed ‘movement sensors’, which they claim, can detect the use of physical countermeasures ([www.lafayetteinstrument.com](http://www.lafayetteinstrument.com)). Again, however, these claims have not been substantiated by any published research.

### *Drugs & Alcohol*

With regards to the effect of drugs and alcohol on the polygraph, results have been mixed. In an early study, Waid, Orne, Cook and Orne (1981) found that the anti-anxiety drug, meprobamate, significantly impaired the accuracy of the *CIT*. In a replication of this study Iacono, Boisvenu & Flemming (1984) found,

however, that diazepam (anti-anxiety) and methylphenidate (stimulant) did not affect the accuracy of the *CIT*. Iacono, Cerri, Patrick and Flemming (1992) similarly found that diazepam, meprobamate and propranolol also had no effect. With regards to the *CQT*, Honts and Amato (2002) reported on an unpublished study by Gatchel, Smith and Kaplan (1984) that found that propranolol actually improved the accuracy.

Two studies have investigated the effect of alcohol on the *CQT*. Bradley and Ainsworth (1984) found that being intoxicated at the time of a test did not affect the accuracy. They did report, however, that being intoxicated at the time of the mock crime reduced the accuracy for guilty examinees. By contrast, O'Toole, Yuille, Patrick and Iacono (1994) found being intoxicated, at any time (during polygraph examination or during mock crime) had no effect on the accuracy.

Overall, only a limited number of studies have investigated the affect of drugs and alcohol on the accuracy of the polygraph. Because the subjects in this research were volunteers, and because there were no consequences for failed polygraph outcomes (other than losing a small financial reward), like much of the research investigating countermeasures, it is questionable as to whether the findings can be generalized to real-life settings. The evidence seems to suggest that drugs and alcohol have only a limited impact on the polygraph. Considering the dearth of robust research, however, it would be somewhat premature to conclude that these substances have no effect.



## *Expectations*

Research within social psychology has shown that the expectations an individual holds can unconsciously affect their behaviour and also the behaviour of others. Perhaps the most well known example of this phenomenon is the so-called ‘Pygmalion effect’. This describes how a teacher’s initial expectations about specific students can affect the student’s future performance and behaviour in the classroom and on standardised tests. It could be similarly expected that an examiner’s expectation could influence the outcome of a polygraph test. Darley and Gross (1983) identified two forms of ‘biases’ that could effect a polygraph examiner’s judgements and behaviour – the ‘*cognitive confirmation effect*’ and the ‘*behavioural confirmation effect*’.

The ‘*cognitive confirmation effect*’ describes how an examiner’s beliefs and expectations may influence the interpretation of ambiguous information. For example, if an examiner believes that an examinee is ‘innocent’, he may interpret information in a manner that confirms this assumption. For examinees considered deceptive or guilty, the converse could be expected to happen. Indirect evidence for the ‘cognitive’ confirmation bias can be derived from a number of sources. In a small study for example, Barland (1975) reported that in 17 out of 19 cases polygraph examiners who gave evaluations on an examinee’s truthfulness based on information available to them prior to the administration of a polygraph test, maintained these conclusions after administering a polygraph test.

A second and perhaps more dramatic example of this phenomenon was provided by a CBS production of “60 minutes” broadcast in 1986. In this television programme, three different polygraph examiners were independently approached to test for an alleged theft of a camera and lens from an office employing four employees. The polygraph examiners were told that only four employees had access to the equipment, which, therefore, must have been stolen by one of them. They were also told which one was most likely to have stolen the equipment, but that there was no evidence to prove this. In reality, no equipment was stolen, and each examiner was given a different name for the thief. All of the polygraph, examiners examined each of the four employees using the *CQT* procedure and all reached the confident conclusion that the employee that had been named to them as the suspected thief had lied during the polygraph investigation, while the three others had spoken the truth. This anecdotal example provides a vivid demonstration of the possible impact of examiner biases on the outcome of *CQT* examinations, but again only a limited number of studies have empirically evaluated this phenomenon (Ben-Shakhar, 2002).

Elaad, Ginton and Ben-Shakhar (1994), in a study investigating the scoring of charts, manipulated the prior expectations of polygraph examiners by telling half the examiners that the suspect had confessed. The results indicated that prior expectations did have an impact on the examiner’s judgements, although only when the charts were initially scored to be inconclusive. There was no effect for charts that were originally evaluated as either being deceptive or being truthful. In a follow-up study, Elaad, Ginton and Ben-Shakhar (1998) again manipulated the expectations of the polygraph examiner, although this time over the entire *CQT*

examination. In this study seven experienced polygraph examiners tested four innocent subjects suspected of cheating on an aptitude test. The examiners were led to believe that two of the subjects were probably guilty, and the other two were probably innocent. This manipulation had no discernable effect. It is of course possible that the examiners in the study were sceptical about the true purpose of the research. The evidence is too limited overall to enable one to draw any clear conclusions about whether examiner's expectations affect polygraph outcomes.

The '*behavioural confirmation effect*' refers to the effect of expectations on the behaviour of the individual towards others. In the context of a polygraph examination, the effect occurs when an examiner develops a 'gut feeling' about the suspect's guilt or innocence, and then subconsciously influences the measurements to match this belief. For example, an examiner may present the questions in a different manner when he believes that the examinee is deceptive from when he believes he is testing a truthful suspect. This theoretical limitation has been recognised by several proponents, including Honts and Perry (1992: p. 372) who point out:

“An examiner motivated to produce a deceptive result might ask over general or provocative relevant questions, and spend a great deal of time on their review and presentation. Subsequently, this unethical examiner could ask very narrow, specific, or inappropriate control questions and spend very little time on their review and presentation. An examiner predisposed to produce a truthful result could take the opposite approach.”



The American Polygraph Association (APA) attempts to address problems of ‘examiner bias’ by emphasising the importance of an impartial standardised approach ([www.polygraph.org/standards.htm](http://www.polygraph.org/standards.htm)). It is perhaps also noteworthy, that while this concern is plausible because of the lack of research that has conducted, it remains only speculation.

### *Personality*

A limited number of studies have evaluated the impact of personality characteristics on the accuracy of the polygraph. Investigating ‘normal’ personality traits, Giesen and Rollison (1980) found that subjects who described themselves as anxious responded stronger than subjects identifying themselves as non-anxious on the *CIT*. Bradley and Kohn (1987) similarly showed that subjects high in Machiavellianism (i.e. “able to focus more directly on the relevant aspects of a situation”, p.747) displayed greater physiological responsivity when guilty than when innocent. Early studies, however, by Janisse and Bradley (1980) and Oksenberg (1970) had failed to find an effect for Machiavellianism. Other studies have generally failed to find effects for normal personality characteristics. Gudjonsson (1982), for example, found no relationship between personality traits as evaluated by numerous inventories, and the accuracy of the *CIT*. Bradley and Rettinger (1992) also found no difference between subjects high and low in their propensity to monitor their own social demeanour, a characteristic of social anxiety.

Waid, Orne and Wilson (1979) investigated the impact of socialisation on the accuracy of the *CQT* in a mock crime scenario. They found that the 3 guilty subjects who were misclassified (i.e. false negative) scored significantly lower on a socialisation scale than the 12 guilty subjects who were correctly classified. By contrast, innocent subjects who were false positives had significantly higher socialisation scores than the subjects who were correctly classified, and also, the false negatives. The study tentatively concluded that poorly socialised individuals might be prone to false negative errors, whilst highly socialised individuals may be prone to false positive errors. This finding has not been replicated by other studies. Raskin, Barland and Podlesny (1976), for example, failed to find any relationship between outcomes on the *CQT* and psychopathy (a construct related to socialisation) as defined by the Pd scale of the Minnesota Multiphasic Personality Inventory (MMPI). In another study, Raskin and Hare (1978) reported on a laboratory experiment that utilised a mock crime scenario with psychopathic and non-psychopathic criminals as subjects. Consistent with the Raskin et al. (1976) study, there was no difference between the groups, and indeed, all the deceptive psychopaths were correctly identified. Hammond (1980) also found no difference between a student sample, psychopaths and alcoholics in a mock crime scenario. Similarly, neither Honts, Raskin and Kircher (1985) or Patrick and Iacono (1986, 1989) found that socialisation or psychopathy affected the accuracy of the polygraph.

The limited number of studies that have investigated the impact of personality have tended to show that this is probably not an important factor

influencing polygraph results, although again, it would be premature to conclude that personality has no effect as two studies did find an effect.

### *Intelligence, Age & other Demographic Variables*

Little research has examined the effect of ethnic background, education and intelligence on polygraph testing. Intelligence could conceivably affect the accuracy of polygraph tests, as more intelligent examinees might be inclined to educate themselves about polygraph testing and also utilise countermeasures. Barland and Raskin (1975) found however, that intelligence or education level did not affect accuracy. In a separate analysis of this data, Raskin (1976) actually reported that the majority of false positives within the initial research had occurred among subjects with college degrees. It is arguable, however, that examinees in the seventies and eighties had the same access to detailed information about polygraph testing as people do today. For example, simply typing 'how to beat the polygraph' on the search engine [www.google.com](http://www.google.com) generates 20 900 web pages (October 2004). It remains plausible, therefore, that more intelligent examinees today are better informed, and thus, more likely to successfully utilise countermeasures.

A small number of studies have investigated the impact of demographic variables on the accuracy of the polygraph. Buckley and Senese (1991) found that race and gender of the examinees did not affect the blind re-scoring of polygraph charts. Reed (1993a; 1993b), in a laboratory study utilising 381 examinees, found that there was no effect for any of the examined variables (i.e. race, age, gender,



income, education level) for either examinees or examiners. Two studies by Patrick & Iacono (1989, 1991) have, however, found that age differentiated false-positives and true-negatives. Younger subjects were more likely to be a false positive. These variables have only been minimally investigated and much of this research is plagued by the same problems as the studies that have investigated accuracy.

## Reliability

*Reliability* refers to the degree to which a test consistently produces the same results across different subjects, contexts and with different examiners. There are several aspects to reliability that are important for polygraphy. *Test re-test reliability* refers to the extent that the test used to examine the same subject for the same purpose yields the same result. *Inter-rater reliability* refers to the extent that different examiners arrive at the same conclusions. Correlations are used to present such data. Tests can be highly reliable, though not valid. A test cannot be considered valid, however, if it is unreliable (Grimm & Yarnold, 2000; Tabachnick & Fidell, 2001). To date the majority of published research on polygraph reliability has investigated the inter-rater reliability of chart scoring, and generally these studies have demonstrated high rates of reliability (.85 to .93) (e.g. Honts & Raskin, 1988; Horvath, 1977; Patrick & Iacono, 1991). A single study has evaluated test-retest reliability. Dollins, Cestaro and Pettit (1995) investigated the accuracy of repeated *CIT* polygraph tests over a six-day period. The results showed perfect agreement between the two occasions.

## What needs to be done?

Post-conviction polygraphy has in recent years been proposed as a valuable treatment and supervision tool with sex offenders. The polygraph when used in this context appears to have the potential of significantly improving the overall management of community-based sex offenders. Little research, however, has investigated the claims made by proponents and the accuracy of the polygraph remains a controversial issue within the wider scientific community. Much of this latter debate has focused on the value of the utilised methodologies, namely field and laboratory studies. Notably no previous research has evaluated the accuracy of the polygraph as used in a post-conviction capacity with sex offenders. The purpose of this thesis is two-fold: firstly, to investigate the utility of post-conviction polygraphy with community-based sex offenders; and secondly, to examine the accuracy of the polygraph in this context.

### *Study 1*

The initial study incorporated a prospective design with comparison groups to evaluate the impact of the polygraph testing on the engagement or avoidance of risk behaviours. Participants were seen on three occasions. On the first occasion subjects completed a number of questionnaires and a brief interview, during which four high-risk behaviours were identified. Based on location participants were divided into two groups. One group was told to expect a polygraph test focussed on their specified high-risk behaviours, whilst the second group were not told to expect a polygraph test.



On the second occasion participants were again interviewed about their identified behaviours. All the volunteers were then asked to submit to a polygraph test focussed upon these. After the polygraph test, participants who had disclosed engaging in high-risk behaviours were now told to expect a polygraph test in three months time, whilst the participants that did not disclose any high-risk behaviour and passed their polygraph were not. The procedure for the final part of the study was identical in format to the second occasion. The following hypotheses were tested:

- that regular polygraph examinations would act as a deterrent against engaging in high-risk behaviours;
- that offenders expecting a polygraph test would engage in less serious risk behaviour;
- that the polygraph would assist with gaining previous unknown information about high-risk behaviours in sex offenders.

This part of the study also investigated the impact of risk, personality and psychological distress on the nature of disclosure of behaviours. The following hypotheses were tested:

- that risk would be associated with greater numbers of behaviours reported;

- that antisocial personality disorder would be associated with reporting larger numbers of risk behaviours;
- that personality disorder would be associated with engaging in larger amounts risk behaviour;
- that psychological distress would be associated with reporting engaging in larger numbers of risk behaviours;
- that the normal personality traits of neuroticism and agreeableness would be associated with reporting larger numbers of risk behaviours.

### *Study 2*

The next study utilised a unique methodology to investigate the accuracy of post-conviction polygraph maintenance tests with adult sex offenders. Four conditions needed to be fulfilled for this methodology to overcome the limitations of previous research:

- Firstly, there should be a conclusive criterion for establishing truthful and deceptive examinees;
- secondly, there needed to be an independence between the criterion and the polygrapher's judgement;

- thirdly, the sample should be representative of the examinees for which the polygraph is being used;
- finally, the testing conditions should be ‘real-life’ or closely resemble those of a true examination, in particular any deceptions should be authentic and there should be meaningful consequences associated with the test outcome.

The second study recruited community-based sex offenders in the US mandated to complete regular polygraph maintenance examinations as part of their supervision and treatment package. During their regular maintenance exam in addition to their usual questions an additional question about drug use was asked. Immediately following this test the offenders were asked to supply a hair sample that was subsequently analysed for drugs. By having blind examiners score the charts it is then possible to compare the objective drug test result with the information provided during the testing process regarding drug usage and also the overall polygraph outcome. Drug-use whilst on probation in the US is prohibited and can potentially incur significant sanctions, including revocation of probation. Utilising this methodology overcomes, therefore, the difficulties associated with previous studies.

The impact of personality, psychological distress and demographic characteristics were examined in regards to accurate and inaccurate outcomes. In addition, the volunteers completed a questionnaire asking about previous



polygraph outcomes (whether correct or wrong), the use of countermeasures and whether they had ever made false admissions regarding their behaviour during a polygraph test. The participants were also asked about the usefulness of the polygraph with avoiding risk behaviour and offending.

## CHAPTER 6

### Study 1

#### *Method*

##### *Participants*

One hundred and sixteen convicted sex offenders taking part in community-based treatment programmes in three probation service areas in the UK were approached for the study, of whom 50 (43%) agreed to participate. Forty-nine subjects were White British and one was of Asian origin. Twenty-one reported being married or in a stable relationship, and 29 were either separated or single. The mean age of the sample was 41( $SD = 10.7$ ) with a range of 22 to 67 years.

In regards to offence characteristics, the sample consisted of 45 offenders against children (including offences involving possession and manufacture of child pornography), and 5 men who had offended against adult victims.

##### *Polygraph Examination*

The polygraph examinations were conducted by one of two polygraphers, both of whom were members of the American Polygraph Association, and who were accredited to conduct Post-Conviction Sex Offender Polygraph

Examinations. The examiners used standard polygraph instruments that provided measures of respiration, skin resistance, and cardiovascular activity. Each examination consisted of three phases: a pre-test interview; a test phase; and a post-test interview.

The examiner explained the polygraph examination during the pre-test interview, conducted an acquaintance test, and then asked the participant whether he had engaged in any of his high-risk behaviours over the preceding three months. If the subject disclosed having done so, he was asked when and how often these had occurred. The examiner then reviewed the relevant and comparison questions. The MGQT protocol, as outlined in chapter 3, was used. In most cases, the relevant question was structured in the following manner: “Since (date of initial session), other than what you have told me, have you (e.g. had any unsupervised contact with someone under the age of 16 years / masturbated to thoughts of children)?” A typical comparison question was: “Since you have been on probation have you ever lied to (your treatment provider/ probation officer/ someone who trusts you)?” During the examination phase, the examiner monitored the subject’s physiological responses to test questions.

The examiner scored the charts with the numerical scoring procedure described earlier for the MGQT. The numerical scoring of the charts was performed prior to the post-test interview. The subject was then told the result of his test (i.e. whether truthful, deceptive or inconclusive). The examiner’s strategy in the post-test interview was dictated by his evaluation of the subject’s truthfulness. If the subject was deceptive, he was provided with an opportunity to



explain the result. All post-test interviews were conducted in a low-key, non-threatening manner. After this interview, the examiner and the participant returned to the initial interview room, where the examiner reported the outcome of the test and also discussed any disclosures the participant made during the polygraph process.

Overall, the initial polygraph test took between 1 to 2 hours, whilst the second polygraph test was slightly shorter, because less time was needed to explain the procedure, instruments and to address the participants' concerns and questions.

### *Ethical Approval*

Ethical approval was gained through the Northumberland, Tyne and Wear Local Research Ethics Committee. The National Probation Service accepted this committee's decision for the research to be conducted in the probation service areas outside Northumberland.

### *Materials*

#### *Personality.*

The Revised NEO-Personality Inventory (NEO-PI-R) is a self-report questionnaire, developed by Costa and McCrae (1992) to assess normal personality dimensions based on a five-factor model. It consists of 243 items, which are rated on a five-point Likert scale, anchored by Strongly Disagree and Strongly Agree

options. Three validity items are included. The test takes between 30 and 45 minutes to complete. The inventory gives a score for the five domains that are based on 48 questions per domain, as well as assessing six facets within each domain using 8 items per facet.

The five domains and their facet scales are:

1. Neuroticism (N): This domain covers emotional and stability from one extreme to maladjustment or neuroticism at the other. High scorers are more likely to experience pervasive feelings such as guilt, fear, disgust, anger and embarrassment. The facet scales included - Anxiety, Hostility, Depression, Self-consciousness, Impulsiveness, and Vulnerability.
2. Extraversion (E): This domain reflects extroversion at one extreme and introversion at the other. High scorers tend to like large groups, are outgoing and generally sociable, while low scorers tended to be introverted, reserved and independent. The facet scales are - Warmth, Gregariousness, Assertiveness, Activity, Excitement-Seeking, and Positive Emotions.
3. Openness to Experience (O): This domain contrasts openness to new experiences, thoughts and processes, with a rejection of such. High scorers tend to be curious and tolerant of differences, whilst low scorers are more conventional and conservative. The facet scales are - Fantasy, Aesthetics, Feelings, Actions, Ideas, and Values.

4. Agreeableness (A): This domain contrasts a type of 'easy-going' with 'hard headiness'. High scorers tend to be altruistic, optimistic and sympathetic to others. They are more popular, but may struggle to stand up for their beliefs and are seen by some as gullible. Low scorers tend to be more egocentric, arrogant, sceptical and competitive. The facet scales are - Trust, Modesty, Compliance, Altruism, Straightforwardness, and Tender-Mindedness.
5. Conscientiousness (C): This domain contrasts a form of impulse control as reflected in the process of planning, organising, and carry out tasks. Individuals with high scores on this domain tend to be more scrupulous, punctual and reliable than low scorers. The facet scales are - Competence, Self-discipline, Achievement-Striving, Dutifulness, Order, and Deliberation.

*T* scores, with a mean of 50 and standard deviation of 10, were calculated for each of the five dimensions and the 30 facets using adult normative data. Numerous studies have been conducted using the NEO-PI-R, and it has consistently shown good reliability and validity (Clarkin, Hull, Cantor & Sanderson, 1993; Costa, McCrae, & Dye, 1991; Trull, 1992; Wilberg, Urnes, Friis, Perdersen & Karterud, 1999). Internal consistencies for the facets have ranged from .56 to .81 and .86 to .92 for the five broader domains (Costa, McCrae & Dye, 1991). The Cronbach alpha coefficients in this sample for the five domains ranged from .6 (Openness) to .91 (Neuroticism). ). In the present study, the NEO profiles that were missing more than 40 items were deemed to be invalid and excluded from the analysis.



### *Personality Disorder.*

The Structured Interview for Personality Disorders (SCID-II) is a semi-structured diagnostic interview for assessing the 10 DSM-IV (American Psychiatric Association, 1994) Axis II personality disorders, as well as Depressive Personality Disorder and Passive-Aggressive Personality Disorder (included in DSM-IV's Appendix B). SCID-II can be used to make Axis II diagnoses, either categorically (present or absent) or dimensionally (by noting the number of personality disorder criteria scored for each diagnostic category). The SCID-II format used for this study consisted of two phases. The first phase involved a 118-item self-report inventory that investigated each of the 10 specific personality disorders and two appendix categories. Participants indicate 'Yes' or 'No' to whether they have experienced characteristic symptoms of each personality disorder. Three studies have demonstrated that this component of the SCID can be used as a valid screening tool for personality disorder symptomatology (Eksilius, Linstrom, von Knorring, Bodlund & Kullgren, 1994; Jacobsberg, Perry & Frances, 1995; Nussbaum & Rogers, 1992). Ball, Rounsaville, Tennen and Kranzler (2001) reported good internal consistency (above .6) for all the scales, except Schizoid (.35). In the present study Dependent (.43), Obsessive-Compulsive (.43) and Schizoid (.17) had Cronbach Alpha coefficients less than the lowest acceptable value (.6; Tabachnick & Fidell, 2000) and were excluded from further analyses. The alpha coefficients for the remaining scales ranged from .65 (Schizotypal) to .85 (Antisocial).

The second phase involved a clinical interview that was focussed on the items that the subjects responded 'yes' to, each criterion is rated on a 3-point scale (1 = not present; 2 = sub threshold; 3 = threshold). So that I could examine the interrater agreement of the PD ratings, two researchers (both trained to Master's degree level in clinical or forensic psychology) interviewed 15% of the sample (7 participants) together. The mean kappa for the first and second rater was computed to be .79.

### *Psychological distress.*

The Brief Symptom Inventory (BSI) was used to evaluate psychological distress associated with AXIS I disorders (Derogatis & Spencer, 1982). The BSI consists of 53 items covering nine symptom dimensions. These are: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation and Psychoticism; and three global indices of distress are also included. These are: Global Severity Index, Positive Symptom Distress Index, and Positive Symptom Total. The global indices measure current or past level of symptomatology, intensity of symptoms, and number of reported symptoms, respectively. Respondents are asked to rank each item (e.g. "your feelings being easily hurt") on a 5-point scale ranging from 0 (not at all) to 4 (extremely). These rankings reflected the intensity of distress during the previous seven days. The global indices are calculated by summing the values for the items included in that dimension and dividing the number of items endorsed. Raw scores are converted to age and gender-corrected standard t-scores, using adult

nonpatient community norms (mean = 50, SD = 10). A t-score greater than 63 is used to identify clinical cases (Derogatis, 1993). The BSI has been effectively used as a screening instrument in medical settings (Derogatis & Spencer, 1983). Derogatis (1993) has reported good internal consistency with the Cronbach alpha coefficients for the nine scales, ranging from .71 to .85. In the present study the alpha coefficients ranged from .69 (Psychoticism) to .91 (Somatization). Because this study was interested in the impact of psychological distress, the Global Severity Index (GSI) was used. The GSI is the most sensitive single indicator of the respondent's distress level, combining information about the numbers of symptoms and intensity of distress (Derogatis, 1993).

### *Risk.*

The Static-99 is an actuarial risk prediction instrument designed to estimate the probability of sexual and violent recidivism for adult males who have already been convicted of at least one sexual offence against a child or non-consenting adult (Hanson & Thornton, 1999, 2000). Static-99 is based on follow-up studies from Canada and the United Kingdom with a total sample size of 1,301 sexual offenders. The measure contains ten items that address static (historical) factors that have been empirically shown to be associated with sexual recidivism. These include:

1. Prior sexual offences;
2. Prior sentencing dates;
3. Any convictions for non-contact sex offences;



4. Current convictions for non-sexual violence;
5. Prior convictions for non-sexual violence;
6. Unrelated victims;
7. Stranger victims;
8. Any male victims;
9. Young (<25 years);
10. Single (has not lived with lover for more than 2 years?);

Each item, excluding 'Prior sex offences', receives a score of 1 for present or 0 for absent. For 'Prior sex offences' an individual can receive a score between 0 and 3 depending on the number of previous sex offence convictions. An interview is not required to complete the Static-99, as the data can be gathered from a participant's probation and / or treatment file. The scores for all items are summed and individuals receive a score out of 12. The Static-99 scores can be translated into four risk categories – Low (0 to 1); Moderately-Low (2 to 3); Moderately-High (4 to 5); and, High (6+). The Static-99 has shown moderate predictive accuracy (ROC = .71) in diverse samples of sexual offenders drawn from the United States, Canada and the UK (Hanson & Harris, 2001; Hanson & Thornton, 1999). Numerous studies have investigated inter-rater reliability showing rates between .87 and .96 (Hanson, 2001; Harris, Rice, Quinsey, Boer & Lang, 2002). In the present study 10 cases (20%) were randomly selected and blindly re-scored by a Chartered Forensic Psychologist. There was perfect agreement.

### *Polygraph experience.*

The Polygraph Experience and Disclosure Questionnaire (PEDQ) is a 7-item survey developed specifically for the purpose of this study which covered the following areas: the helpfulness of the polygraph with remaining offence-free; avoiding risk behaviours; and disclosing information to their treatment provider. The questionnaire also asked whether the participant had deliberately attempted to lie during the polygraph and whether this was accurately identified. See appendix D.

### *Procedure*

Participants were approached to take part in this study while attending their regular treatment programme at their probation service or treatment facility. These groups varied in size from 6 to 16. All participants were told that the purpose of the study was to investigate the use of the polygraph ('lie detector') with sex offenders. They were informed that the specific aim of the study was to determine whether the polygraph could assist offenders avoid high-risk behaviours. Subjects were then either told that they *would* be required, or that they *may* be asked, to take a polygraph test as part of the study. All subjects were told that if they were asked to take a polygraph test they would *not* be asked about any specific illegal or criminal behaviour. Participants were then given an information sheet (see Appendix B) that detailed the nature and purpose of this research. This information sheet was slightly different. This reflected the two conditions used in

this study. One sheet stated that participants would be required to take a polygraph test, while the second suggested that they might be asked to take one. Subjects were told that participation was completely voluntary and that they could withdraw at any stage without giving a reason. Individuals that volunteered either commenced their participation immediately or arranged a suitable time to participate in the research. Throughout the study, all volunteers were seen individually, and were reimbursed for any travel expenses incurred due to participating in the study (see figure 2 for study design).

### *Session 1*

As noted earlier, there were three parts to this study. In the first session participants initially signed a consent form detailing the nature of the research and limits of confidentiality (see Appendix B). The volunteers then completed the NEO, BSI and questionnaire component of the SCID-2 by themselves in a quiet room. After completion of these questionnaires, the subjects were initially interviewed about their responses on the SCID-2 and then regarding their current circumstances.

Prior to this interview, four ‘high-risk’ behaviours had been identified. These behaviours had been identified through discussion with the subject’s treatment provider and/or perusal of their treatment or probation file (either when the subject was completing the questionnaires or prior to the arranged appointment). These behaviours were chosen to reflect a spectrum of seriousness from low to high serious risk behaviours. ‘Seriousness’ was determined firstly by



how relevant the behaviour was considered to be to the individual's offending behaviour, and secondly, the extent of purposeful activity required to engage in the specific behaviour. These behaviours were then classified into three categories – low, medium and high seriousness. 'Low serious' behaviours, for instance, were comparatively passive behaviours such as masturbation to deviant fantasies. 'Medium serious' risk behaviours involved slightly more active behaviours, such as visiting areas specifically because children were there. 'High-serious' risk behaviours reflected active attempts to acquire victims, for example, having unsupervised contact with a potential victim (see Table 5 for list of the behaviours and their ratings).

Table 5.

*Behaviours rated as 'low', 'medium' and 'high' serious.*

Behaviour	Risk Level
Masturbating to deviant fantasies	Low
Associating with other sex offenders outside of group	Low
Collecting pictures of children for masturbation (e.g. magazine or clothing catalogues)	Low
Watching television shows involving children for sexual arousal	Low
Excess alcohol consumption and/ or illegal drug use	Medium
Going to areas to view children for sexual pleasure (e.g. arcades, playgrounds, gyms)	Medium
Corresponding with other sex offenders in relation to sex offending (e.g. exchanging stories)	Medium
Supply alcohol to minors	Medium
Accessing child pornography and using chat-rooms on the internet	Medium
Wandering alone at night	Medium
Aimless driving or cruising	Medium
Sexual activity in public place	High
Unsupervised contact with children or vulnerable adults	High
Travel to countries where unsupervised access to children is increased	High
Actively seeking victim	High
Attempting to set-up situation to have contact with children (e.g. applying for inappropriate work, offering to baby-sit)	High

During the initial session the identified behaviours were presented and discussed with the participant. If he accepted that the identified behaviours were ‘high-risk’ (that is, increased his risk of a re-offence) and should be avoided, it was agreed that they would be monitored over the next three months.

All participants were told that they were expected *not* to engage in any of the identified behaviours. If they did, however, then it was expected that they would report doing so to their treatment provider or probation officer. All participants were also informed that any disclosures regarding their high-risk behaviours during the study would be reported to their treatment provider or to their probation officer (see Appendix C for sample report).

Through the recruitment process, all the groups approached had been randomly allocated into one of the two conditions - 'a polygraph aware group' and a 'polygraph unaware group'. During the recruitment process participants were initially informed which group they were in. During the first session in the volunteers in the Polygraph Aware group were specifically told to expect a polygraph examination in their next session (3 months time) focussed on the agreed upon high-risk behaviours. The Polygraph Unaware group, however, were told that their behaviours would be reviewed in the next session and were not explicitly told to expect a polygraph. Finally, the participant's probation files were accessed to gather demographic information and the necessary information to complete the Static-99.

### *Session 2 (Time 1)*

After three months (Time 1) the subjects were again seen for an interview. This took place either at their probation office, or in a treatment facility, or in a room at the University of Newcastle-upon-Tyne. All subjects were initially reminded that any disclosures regarding risk behaviour would be included in a



report to their treatment provider or to their probation officer. Their high-risk behaviours were then reviewed, and they were asked whether they had engaged in any of these behaviours, and if so, how often and when. After this interview, all the participants were asked to complete a polygraph examination focussed upon these behaviours, regardless of which group they were in. They were then taken to another room and introduced to the polygraph examiner. The polygraph examiner had been given a list of the subject's high-risk behaviours prior to this meeting.

After the polygraph examination was completed the polygraph examiner and subject returned to the original interview room. Here the examiner presented the results of the polygraph examination (i.e. deception indicated or no deception indicated) and also reported any disclosures regarding high-risk behaviours that the participant may have made during this time. The participant was then asked to confirm the details reported by the polygraph examiner. The polygrapher then left the room and the participant was provided the opportunity to discuss the results with this researcher. If the participant disclosed having engaged in the high-risk behaviour, he was asked whether he had reported doing so to his treatment provider or probation officer. If he reported having done so, this was confirmed with the relevant professional at a later date. Finally, if the participant had reported engaging in high-risk behaviours at any stage (i.e. during the initial interview or the polygraph examination) he was told that the same behaviours would be reviewed again in the next session. He was also told that he would also be asked to complete another polygraph examination (thus forming a new 'Polygraph Aware' group). Subjects that did not report any high-risk behaviour and who also passed the polygraph test, were simply told to 'keep up the good

work', thereby forming a new 'Polygraph Unaware' group. An appointment for the next session was then arranged. The research design is illustrated in Figure 2.

### *Session 3 (Time 2)*

Session 3 was identical to the format of session 2. Again all participants were interviewed and asked whether they had engaged in any high-risk behaviour. After this interview, all subjects immediately underwent a polygraph examination in a separate room that focused on whether they had engaged in any of their high-risk behaviours over the previous three months. After the participant had been debriefed he then completed the PEDQ (see Appendix D).

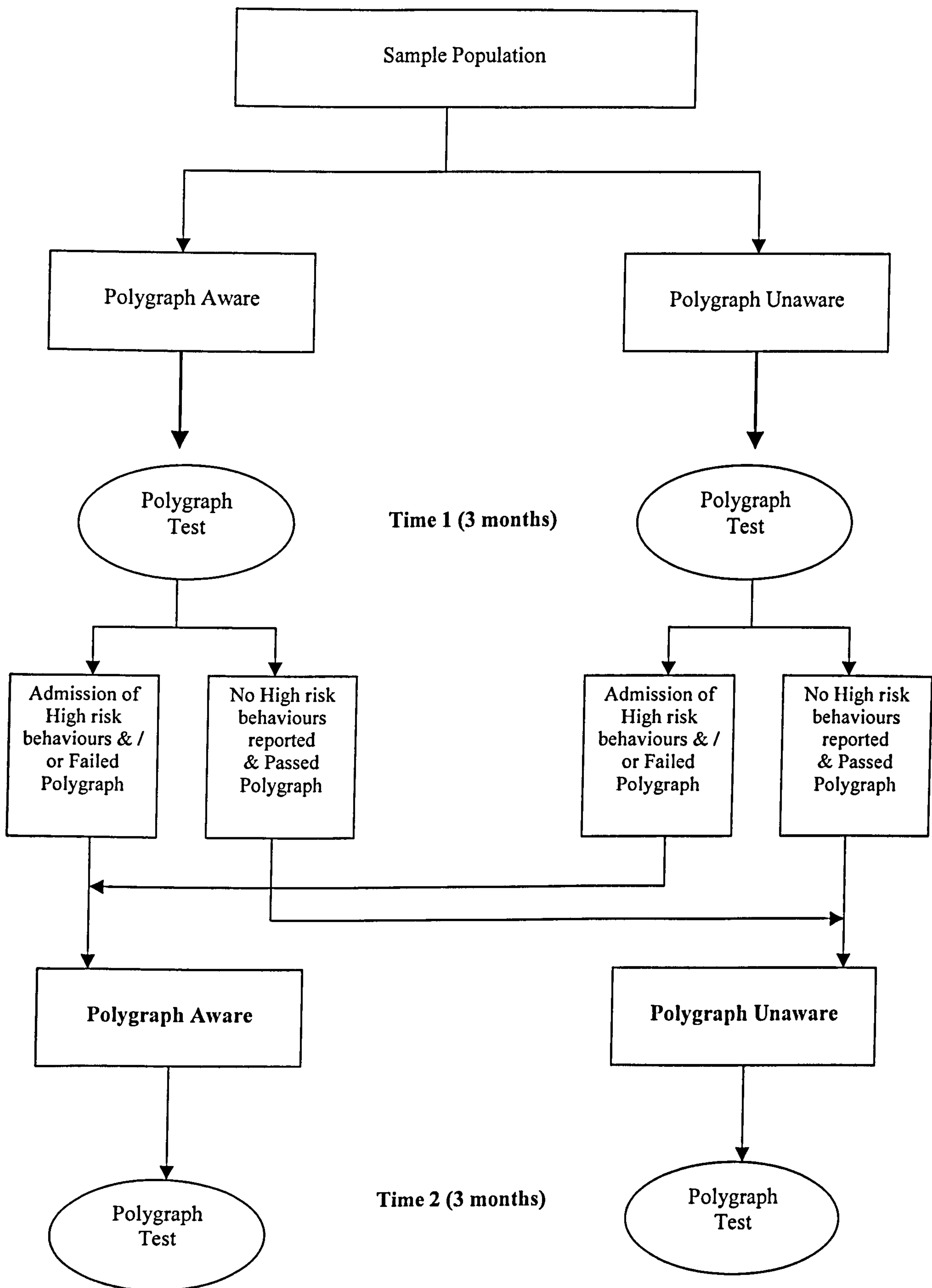


Figure 2. Study 1 design



## *Results*

Data entry was verified by a random check across all instruments in the study of 10% of the data. Less than .01% error was found. SPSS was used for data entry and analysis. For an 80% chance of detecting a difference in the number of risk behaviours reported between the groups, 41 offenders were needed in each group for a total sample size of 82 offenders.

### *Participants*

The 'polygraph aware' group consisted of 27 men with a mean age of 43 (SD = 11.5), of whom 24 (89%) had offended against children and 3 (11%) against adult women. The 'polygraph unaware' group comprised 22 men with a mean age of 39 (SD = 9.5), of whom 20 (91%) had offended against children and 2 (9%) against adult women. One subject was excluded from the study because of acute mental illness. There was no significant difference between these groups in terms of age,  $t(47) = 1.24$ ;  $p > .05$ .

### *Personality characteristics of sample*

Two participants produced invalid NEO-PI-R profiles and were excluded from the analysis. Table 6 displays the mean scores and standard deviations for the NEO-PI-R domain and facet scores. As a group, the Five Factor Model (FFM) personality profile of these sex offenders was characterized by high neuroticism

(N), suggesting that they have a high level of maladjustment and emotional instability. Inspection of the facet scores within the N domain provides an opportunity to examine the specific areas where sex offenders are likely to have difficulties. Here the scores indicate that offenders experience trait anxiety, depression and hostility. They are likely to struggle to cope with stress, and have difficulty controlling their cravings and impulses. The volunteers also scored highly on the self-consciousness facet suggesting that they experience pervasive feelings of guilt and shame. At the other end of the spectrum, the sample was characterized by low conscientiousness, indicating a tendency to be unreliable. Facet scores indicate they are likely to become easily discouraged when confronted with difficult tasks, have poor self-confidence, and make generally hasty or impulse decisions without having thought through the consequences.

The low score on the straightforwardness facet indicates that sex offenders are more willing to manipulate others through deception. Perhaps a somewhat unexpected finding was that the sex offenders scored highly on the modesty and tender-mindedness facets of the agreeableness domain. This suggests that they are generally humble and self-effacing and have a high ability to feel concern for others.

To ensure that there were no differences on these personality characteristics between the two conditions a one-way between-groups multivariate analysis of variance (MANOVA) was performed. The five NEO-PI-R domains were the dependent variables, whilst the independent variable was the condition (polygraph aware & non-aware). Preliminary assumption testing was conducted to check for

normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There was no significant difference between the two groups on the combined dependent variable demonstrating that there were no differences in the NEO domain scores for the two groups [ $F(5, 41) = .85, p > .05$ ]. The observed power was low at .27, and the effect size was small at .093.



Table 6

*Means, Standard Deviations and Ranges for the NEO-PI-R Domain and Facet*

NEO-PI-R Domain and Facet Scales	<i>No Personality disorder(SD)</i>	<i>Personality disorder(SD)</i>	<i>Overall Mean<sup>a</sup> (SD)</i>
<b>Neuroticism</b>	<b>57.8 (13.5)*</b>	<b>70.1 (13.7)*</b>	<b>63.8 (14.8)</b>
N1 Anxiety	58.7 (10.5)	66.9 (11.4)	62.7 (11.6)
N2 Anger Hostility	49.0 (10.3)**	67.7 (12.9)**	58.1 (14.8)
N3 Depression	63.0 (10.2)	71.9 (12.2)	67.4 (11.9)
N4 Self-consciousness	58.9 (14.1)	66.8 (12.7)	62.8 (13.9)
N5 Impulsiveness	51.0 (7.0)**	60.2 (11.1)**	55.5 (10.3)
N6 Vulnerability	58.0 (11.0)**	69.7 (12.8)**	63.7 (13.2)
<b>Extraversion</b>	<b>45.7 (8.4)</b>	<b>47.3 (11.7)</b>	<b>46.5 (10.1)</b>
E1 Warmth	45.6 (9.5)	41.0 (10.6)	43.4 (10.2)
E2 Gregariousness	47.6 (10.6)	44.7 (11.7)	46.2 (11.1)
E3 Assertiveness	44.9 (12.0)	42.3 (10.8)	43.6 (11.4)
E4 Activity	46.3 (5.1)	44.4 (6.6)	45.4 (5.9)
E5 Excitement Seeking	43.7 (10.1)	50.9 (12.5)	47.2 (11.8)
E6 Positive Emotions	47.0 (7.1)	43.0 (11.3)	45.1 (9.5)
<b>Openness</b>	<b>51.4 (11.2)</b>	<b>45.6 (11.8)</b>	<b>48.6 (11.8)</b>
O1 Fantasy	51.1 (9.0)	56.8 (11.4)	53.9 (10.5)
O2 Aesthetics	49.8 (13.6)	47.6 (12.0)	48.7 (12.8)
O3 Feelings	49.4 (11.0)	46.9 (13.8)	48.2 (12.4)
O4 Actions	48.7 (10.3)	41.5 (12.6)	45.2 (12.0)
O5 Ideas	51.5 (12.0)	41.7 (14.7)	46.7 (14.1)
O6 Values	50.7 (7.9)	50.3 (7.6)	50.5 (7.7)
<b>Agreeableness</b>	<b>55.7 (6.7)*</b>	<b>44.6 (11.2)*</b>	<b>50.3 (10.6)</b>
A1 Trust	50.7 (7.0)**	37.4 (12.9)**	44.2 (12.2)
A2 Straightforwardness	48.6 (6.9)**	36.2 (11.6)**	42.6 (11.9)
A3 Altruism	48.5 (9.4)	43.9 (11.7)	46.2 (10.7)
A4 Compliance	55.2 (7.2)**	43.0 (13.0)**	49.2 (13.1)
A5 Modesty	55.7 (8.1)	55.0 (12.8)	55.3 (10.5)
A6 Tender-mindedness	55.1 (8.0)	53.9 (8.3)	54.5 (8.1)
<b>Conscientiousness</b>	<b>43.1 (8.4)</b>	<b>38.3 (10.6)</b>	<b>40.7 (9.8)</b>
C1 Competence	40.2 (8.4)	32.1 (11.8)	36.3 (10.9)
C2 Order	46.1 (9.8)	41.4 (12.0)	43.8 (11.0)
C3 Dutifulness	42.5 (8.5)	38.7 (14.4)	40.6 (11.8)
C4 Achievement Striving	43.3 (8.4)	42.4 (11.1)	42.9 (9.7)
C5 Self-discipline	43.8 (9.1)	34.5 (11.5)	39.2 (11.3)
C6 Deliberation	45.8 (7.1)	37.6 (12.4)	41.8 (10.7)

*Scales for overall sample, PD and non-PD groups (n = 47)*

. <sup>a</sup> T-scores (M = 50; SD = 10).

\* P < .05. \*\* p < .004.

### *Personality disorder.*

On the SCID-II 26 subjects (53%) were evaluated as meeting the criteria for a personality disorder (PD). The mean number of PDs diagnosed was 1.08 (SD = 1.65). There was no significant difference between the groups in terms of number of PDs diagnosed [ $t(47) = 1.05, p > .05$ ]. The distribution of DSM-IV PDs was as follows: 27% antisocial; 20% avoidant; 12% depressive and paranoid; 10% borderline; 8% narcissistic; 4% passive-aggressive; 2% histrionic and schizotypal.

A chi-squared analysis indicated there was no difference in the proportion of individuals diagnosed with Antisocial PD in the aware ( $n = 7$ ) and non-aware ( $n = 6$ ) groups [ $\chi^2 = .011; df = 1; p > .05$ ]. To evaluate whether there was a difference in PD between the two groups the scale scores were combined to create an overall PD symptomatology score. An independent samples t-test did not reveal any difference between the conditions on this variable, [ $t(47) = 1.03, p > .05$ ].

### *Relationships between personality disorder and the FFM*

To examine the relationship between the FFM and PD, a dimensional score was created for each of the DSM categories by collating the number of symptoms the individual endorsed during the self-report component of the SCID-2. Table 7 provides the correlations between the nine PDs and the FFM domains. Neuroticism and Agreeableness was strongly associated with most PDs.

Extraversion was negatively correlated with depressive and schizotypal PDs, while Openness was negatively associated with antisocial.

Table 7.

*Pearson correlations between the FFM domains and the DSM-IV Personality Disorders*

	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Avoidant	.69**	-.29	-.15	-.2	-.18
Passive-aggressive	.65**	-.171	.05	-.47**	-.09
Depressive	.74**	-.31*	.12	-.34*	-.06
Paranoid	.52**	.03	-.19	-.63**	-.27
Schizotypal	.43**	-.38*	-.09	-.26	-.01
Histrionic	.54**	.12	.26	-.51**	-.01
Narcissistic	.48**	-.15	.06	-.51**	-.15
Borderline	.56**	.17	-.01	-.46**	-.11
Antisocial	.19	.22	-.40**	-.44**	-.11

\*\* p < 0.01 level

\* p < 0.05 level

*Comparison of the FFM between PD and non-PD sex offenders*

A one-way between-groups multivariate analysis of variance was performed to investigate the difference between the PD and non-PD subjects on the NEO-PI-R domain scores. The independent variable was whether the subject had been diagnosed with a PD. There was a statistically significant difference between the PD and non-PD groups on the combined dependent variables:  $F(5, 35) = 6.03$ ,  $p < .05$ . When the results for the dependent variables were considered separately,



two differences reached statistical significance using a Bonferoni adjusted alpha level of .01, these were N:  $F(1, 39)=8.24, p=.00$ ; and A:  $F(1, 39)=14.7, p < .01$ . An inspection of the mean scores indicated that offenders with a PD reported higher levels of neuroticism (70) and lower levels of agreeableness (44), than non-PD participants ( $N = 57$  &  $A=55$ ). Post hoc power calculations showed low power for this analysis (0.43).

To further examine group differences a second MANOVA was performed using the facet scores. Due to the modest sample size only the facets from the Neuroticism and Agreeableness domains were included. Again, there was a statistically significant difference on the combined dependent variables:  $F(12, 28) = 3.74, p < .05$ . Power was calculated post hoc and was found to be 0.25. When considered separately, six comparisons reached statistical significance using a Bonferoni adjusted alpha level of .004, these were: Anger & Hostility,  $F(1,39) = 26.4, p < .004$ ; Impulsiveness,  $F(1,39) = 10.1, p < .004$ ; Vulnerability,  $F(1,39) = 9.8, p < .004$ ; Trust,  $F(1, 39) = 16.9, p < .004$ ; Straightforwardness,  $F(1,39) = 14.7, p < .004$ ; and, Compliance  $F(1,39) = 11.2, p < .004$ . Personality disordered individuals reported higher levels of Anger and Hostility (68 versus 49), Impulsiveness (60 versus 51) and Vulnerability (70 versus 58) than the non-PD participants. The PD group also reported lower levels of Trust (37 versus 51), Straightforwardness (36 versus 49) and Compliance (43 versus 55). Overall, These findings should be regarded with a degree of caution due the low Power.

### *Psychological Distress & Risk.*

An independent samples t-test was conducted to compare the Global Severity Index to determine whether there was any difference in level of psychological distress reported between the two conditions. There was no difference between the groups on this variable,  $t(47) = .08, p > .05$ . Overall, using adult outpatient norms, the sample was experiencing a moderate degree of psychological distress (GSI = 67, 61<sup>st</sup> percentile).

Scores on Static-99 for the overall sample ranged between 0 and 9, with a mean of 2.8 (SD = 2.2). Thirteen subjects were categorised as low risk, 21 as medium-low risk, 8 as medium-high risk and 5 as high risk. There was no difference in evaluated risk between the aware (M = 2.7, SD = 2.4) and the non-aware conditions (M = 2.9, SD = 2;  $t(46) = -.38, p > .05$ ).

#### Time 1(First Polygraph)

##### *High risk behaviours reported*

Two subjects were recalled to prison in the three-month period leading up to the first polygraph examination, while 14 (30%) of the remaining 47 defaulted from the time 1 assessment and polygraph examination. Another subject, who was in the 'polygraph unaware' group, withdrew from the study after being asked to take a polygraph test. In total, therefore, 32 men completed a polygraph examination at Time 1. This number represents 64% of those who initially agreed

to take part in the study, and 28% of the overall sample of sex offenders approached.

Of the 32 men, one subject had reported engaging in high-risk behaviour to his treatment provider. At Time 1, 31 (97%) offenders reported having engaged in at least one of their high risk behaviours during the preceding three months, either during the assessment interviews or in the course of the polygraph examination. In total 76 high-risk behaviours were reported (Table 8). The mean number of different risk behaviours reported per offender was 2.45 (SD = 1.7), with a median of 2 and a range of 0 to 7 (although only 4 behaviours were established for each offender, 3 men disclosed additional high risk behaviours during the examination). The one offender who did not disclose any risk behaviours failed his polygraph, and gave no explanation for this. A Mann-Whitney U test was conducted to compare the numbers of high-risk behaviours reported by the two groups. There were no significant differences in the number of risk behaviours reported by men in the 'polygraph aware' (M = 2.4, SD = 1.6) and the 'polygraph unaware' groups (M = 2.7, SD = 1.9;  $Z = -.11$ ,  $p > .05$ ).

Table 8

*Number (%) of men reporting different high risk behaviours at Time 1 (n=32 subjects)*

Behaviour reported – Time 1	Number (%)
Masturbation to deviant fantasies	27 (84)
Unsupervised contact with children or vulnerable adults	9 (28)
Attempting to set-up situation for contact with children (e.g. offering to baby-sit)	8 (25)
Going to areas to view children for sexual arousal (e.g. gyms; playgrounds; video arcades)	8 (25)
Collecting pictures of children for masturbation (e.g. magazines, clothing catalogues, art books)	7 (22)
Watching TV shows involving children for sexual arousal	5 (16)
Alcohol consumption in excess of a prescribed limit	3 (9)
Associating with other sex offenders outside group treatment	2 (6)
Supplying alcohol and cigarettes to minors	1 (3)
Accessing internet sites for child pornography/ teenage chat rooms	1 (3)
Travel to other countries where access to children more likely	1 (3)
‘Aimless’ driving	1 (3)
Actively seeking victims in public toilets	1 (3)
Sexual activity in public place	1 (3)
Wandering alone at night	1 (3)
<b>TOTAL</b>	<b>76</b>



Because nearly all the participants had disclosed risk behaviours at time 1, to examine the impact of the variables of interest in this study, the sample was divided into two groups on the basis of the number of behaviours disclosed. Group 1 had reported a 'low' number of behaviours (0 to 2) and group 2 had disclosed a 'high' number of risk behaviours (3 +). Similarly, risk (static), personality disorder (overall sum and antisocial personality disorder), psychological distress (GSI) and the two normal personality domains (neuroticism and agreeableness) were divided into high and low groups based on the means. Table 9 displays the categories and the chi squared results. Only one variable was significant, Antisocial Personality Disorder. The volunteers scoring highly on this variable were more likely to report engaging in a greater number of risk behaviours. Post hoc calculations revealed that power was low for this analysis (0.38).

Table 9

*Categorical variables (high versus low) and risk behaviour categories.*

Variables	Number of risk behaviours reported		
	Low	High	$\chi$
Risk (Static 99)			
Low	6	5	$\chi (1) = .001, ns$
High	11	9	
Psychological Distress (Global Severity Index, GSI)			
Low	8	8	$\chi (1) = .31, ns$
High	9	6	
Personality Disorder (Sum)			
Low	11	7	$\chi (1) = .68, ns$
High	6	7	
Antisocial Personality Disorder (Sum)			
Low	16	8	$\chi (1) = 6.01, p < .05$
High	1	6	
Neuroticism (NEO)			
Low	6	7	$\chi (1) = .68, ns$
High	11	7	
Agreeableness (NEO)			
Low	9	6	$\chi (1) = .31, ns$
High	8	8	

### *Time of disclosure*

All participants had three opportunities to disclose risk behaviours: at the interview prior to the polygraph test, to the polygraph examiner during the pre-test interview, and to the polygraph examiner during the post-test interview. At the initial interview, 14 subjects (44%) disclosed having engaged in risk behaviour, 24 subjects (75%) reported risk behaviours to the polygraph examiner at the pre-test interview (either in addition to what they had told me or in 11 (34%) cases after having denied any such behaviours). Twenty-five subjects (78%) failed the polygraph, of whom 20 (63%) reported additional or new information regarding their behaviour during the post-test interview (see Figure 3). In total, 30 (94%) participants reported *additional* information regarding their high-risk behaviours during the polygraph process.

Research iv

pre-poly iv

poly exam

post-test iv

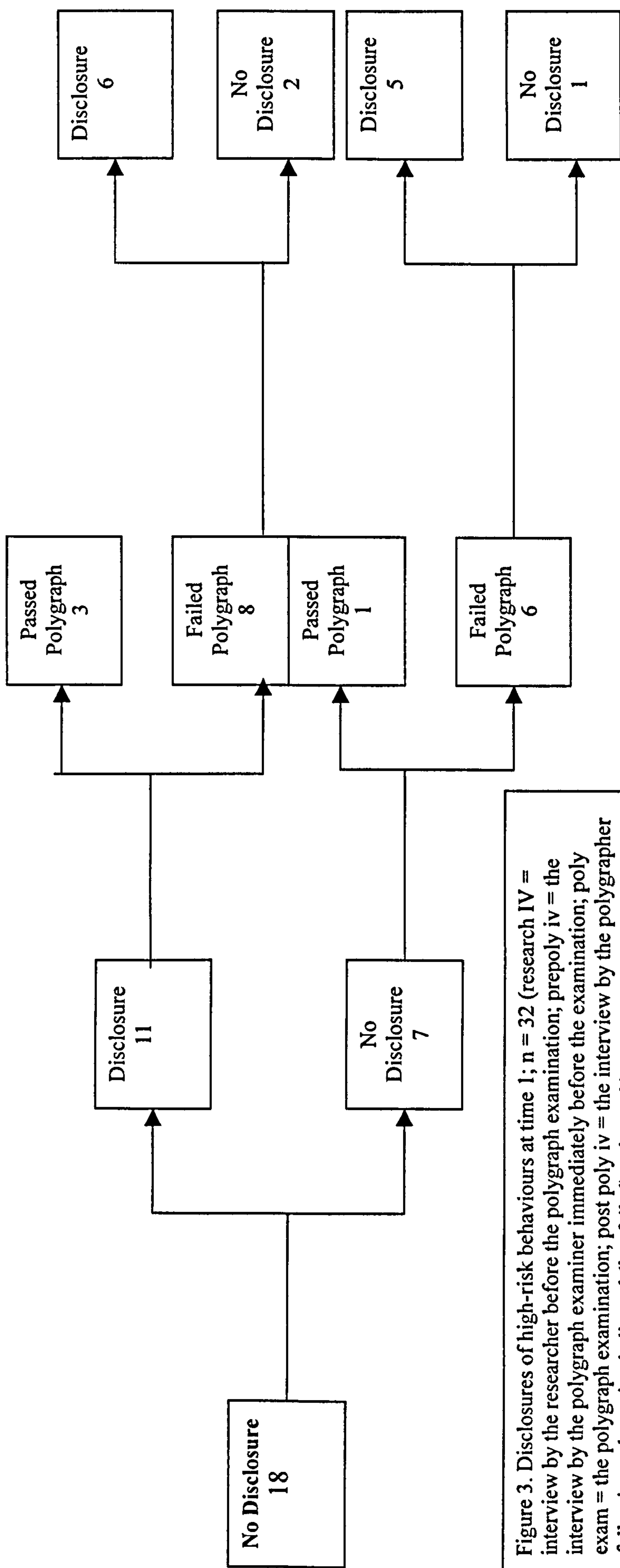
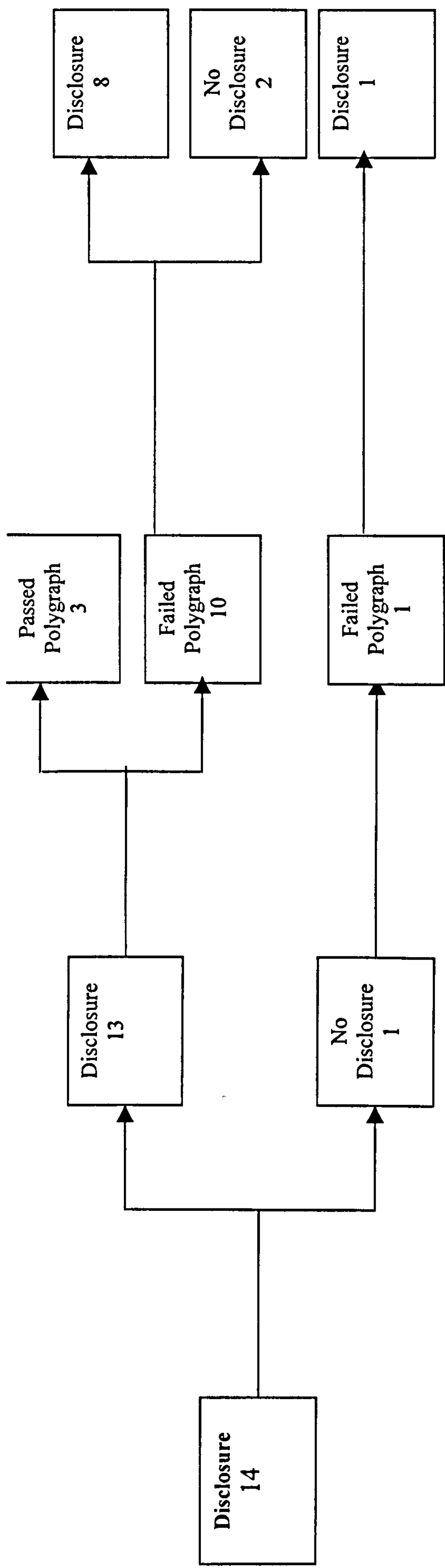


Figure 3. Disclosures of high-risk behaviours at time 1; n = 32 (research IV = interview by the researcher before the polygraph examination; prepoly iv = the interview by the polygraph examiner immediately before the examination; poly exam = the polygraph examination; post poly iv = the interview by the polygrapher following a deception indicated (i.e. failed) polygraph).



When considering only the polygraph testing procedure it is possible to classify the offenders into four groups based on the times of their disclosure of risk behaviour: (i) a '*false denier*' group (n = 6 / 18.7%) where the offender initially denied engaging in risk behaviours, failed the polygraph test, and then confessed to having done so; (ii) a '*false confessor*' group (n = 14 / 43.7%) where the offender confessed some risk behaviour during the pre-test phase, failed the polygraph test, and then confessed additional behaviour; (iii) a '*true confessor*' group (n = 6 / 18.7%) where an offender confessed behaviour during the pre-test and passed the polygraph; and finally, (iv) a '*non-confessor*' group (n = 6 / 18.7%) where the offender either confessed some behaviour or no behaviour in the pre-test interview and failed the polygraph and did not disclose further information.

A question of interest was whether there was a difference in the risk and antisocial personality characteristics of the offenders that attempted to 'hide' risk behaviours (i.e. the false denier and false confessor groups) when compared to offenders that did not do so (i.e. true confessors). It is, for instance, conceivable that greater risk levels and/or antisocial personality features would be associated with a preparedness to conceal having engaged in such behaviours. The two denier groups were collapsed into an overall 'denier' group (n = 20). Using Mann-Whitney U test there were no significant differences, however, between these groups in terms of risk [ $Z = -1.66, p > .05$ ] or antisocial personality characteristics [ $Z = -.62, p > .05$ ].

*Seriousness of behaviours*

To determine whether there was a difference in the ‘nature’ of behaviours reported by the different conditions the ‘seriousness’ of the behaviours disclosed during the polygraph process was considered. For the purpose of the analysis, only the most serious risk behaviour reported by each subject was counted. At this time, the most serious behaviour reported by 11 men (34%) was evaluated as ‘low’ serious, while another 11 (34%) men reported medium serious behaviour, and 8 (25%) high serious.

To determine whether the polygraph had an impact upon the type of behaviour engaged in by participants the two conditions (‘aware’ versus ‘non-aware’) were compared (see table 2). A Mann-Whitney test revealed no significant differences between these groups,  $Z = -.64, p > .05$ .

Table 10

*Comparison between conditions and seriousness of behaviour disclosed*

Risk behaviour category	Condition		Total (%)
	Aware	Non-aware	
No behaviour	1 (3)	1 (3)	2 (6)
Low serious	6 (19)	5 (16)	11 (34)
Medium serious	4 (12)	7 (22)	11 (34)
High serious	6 (19)	2 (6)	8 (25)
TOTAL	17	15	32

A second comparison between the conditions was carried out excluding the five participants who failed the polygraph and did not disclose additional information (see table 11). Again no significant differences between the two conditions were found,  $Z = -.75$ ,  $p > .05$ .

*Personality, personality disorder, risk & psychological distress*

To examine whether normal personality traits (Neuroticism & Agreeableness), personality disorder, risk and psychological distress differed between those who had reported different types of high risk behaviour, a multivariate analysis of variance (MANOVA) was performed. The two participants that did not report any risk behaviour during the polygraph process were excluded from the analysis. There was no statistically significant difference between the groups reporting differing levels of seriousness on the combined dependent variables:  $F(8, 48) = .78$ ,  $p > .05$ ; Wilks' Lambda = .78.

Table 11

*Comparison between conditions and seriousness of disclosed behaviours excluding failed non-disclosing participants*

Risk behaviour category	Aware	Non-aware	Fail & no further disclosure
No behaviour	0	1	1
Low serious	5	4	2
Medium serious	3	7	1
High serious	5	2	1
TOTAL	13	14	5

To determine whether there was a change in the seriousness of behaviour disclosed between the times when participants could have disclosed information, the pre and post-test disclosures were compared. In the pre-test 8 (53%) disclosed low serious behaviour, 3 (20%) medium serious and 4 (26%) high serious behaviours. In the post-test, 9 (60%) made low serious disclosures, 1 (7%) medium serious and 5 (33%) high serious disclosure. Comparison between these times using Wilcoxon Signed Ranks test did not reveal any differences in the seriousness of behaviours disclosed,  $Z = 0$ ;  $p > .05$ .



## *Time 2 (Second polygraph test)*

### *High risk behaviour*

Because all but one of the offenders admitted to having engaged in high risk behaviours at Time 1, all were told to expect a polygraph examination at Time 2, including the seven men who had passed the polygraph.

Eleven subjects (22% of overall participants, and 34% of those who were polygraphed at Time 1) did not attend at Time 2. Of the 21 who remained in the study, 15 (71%) reported a total of 33 high risk behaviours, with a mean of 1.6 (SD = 1.4), a median of 1, and a range of 0 to 5; six (29%) men reported no high risk behaviours at all and passed their polygraph tests. Table 12 displays the numbers and types of behaviours reported at Time 2.

Table 12

*Number (%) of men reporting different high risk behaviours at Time 2 (n=21 subjects)*

Behaviour reported – Time 2	Number (%)
Masturbation to deviant fantasies	15 (72)
Unsupervised contact with children or vulnerable adults	5 (24)
Alcohol consumption in excess of a prescribed limit	3 (10)
Attempting to set-up situation for contact with children (e.g. offering to baby-sit)	2 (10)
Associating with other sex offenders outside group treatment	2 (10)
Going to areas to view children for sexual arousal (e.g. gyms; playgrounds; video arcades)	1 (5)
Collecting pictures of children for masturbation (e.g. magazines, clothing catalogues, art books)	1 (5)
Watching TV shows involving children for sexual arousal	1 (5)
Accessing internet sites for child images &/or teenage chat rooms	1 (5)
Sexual activity in public place	1 (5)
TOTAL	33

Of the 15 subjects who reported high risk behaviours, 9 (60%) had already disclosed these to their supervisors compared to just 1 who had done so at Time 1. The number of behaviours reported by each man at Time 1 and 2 was significant [ $t(20) = 2.55, p < .05$ ]. Participants that attended both polygraphs reported significantly less behaviour at time 2 (58 versus 33 behaviours). Post hoc analysis revealed that power was low for this analysis (0.47).

### *Time of disclosure*

It can be seen from Figure 4 that at Time 2, 13 men (62%) made disclosures at the initial interview, of whom 7 (33%) made no further disclosures and passed their polygraph tests, while 6 (29%) men made no disclosures at all and also passed the polygraph. Considering the four groups described earlier shows that at Time 2 no participants were *false deniers*, 4 (19%) were *false-confessors*, 11 (52%) *true confessors* and 6 (28%) *non-confessors*. At Time 2, therefore, 17 (80%) men either disclosed all their risk behaviour prior to the polygraph, and passed or disclosed no behaviour and passed, by comparison to only 12 (37%) men who did this at Time 1.

In regards to polygraph outcome then 15 (71%) subjects passed their polygraph examinations at Time 2 compared with 7 (29%) at Time 1. Comparing the polygraph results of the 21 men that attended at Time 2 revealed that: 3 (14%) had passed their polygraphs on both occasions; 6 (28%) had failed on both occasions; and 12 (56%) had failed the initial test and passed the second polygraph test. These findings appear to suggest that participants are more likely to pass the polygraph on the second occasion. It is not possible to statistically examine this question, however, due to small cell numbers.

### *Seriousness of behaviours*

Of the 21 men who were examined on both occasions: 12 (57%) reported a decrease in the level of seriousness of the behaviours; in 5 (24%) there was no

change; and in 4 (19%) cases there was an increase. At Time 1, 1 of the 21 men failed his polygraph examination and made no further disclosures, while at time 2 this was the case for 2 men (these were different men on the two occasions). If these 3 men are excluded, then 11 (61%) participants reported a decrease in seriousness, 3 (14%) reported no change and 4 (22%) reported an increase. There was a significant reduction in the seriousness of disclosed behaviours for the participants who attended on both occasions using Wilcoxon Signed Rank Test,  $Z = -2.2, p < .05$ . Power was low for this analysis at 0.15.

The participants were divided into groups reflecting the change in their risk behaviours between Times 1 and 2: Group 1: Participants reporting decrease seriousness ( $n = 11$ ), and group 2: Participants reporting no change or an increase in the seriousness of risk behaviour ( $n = 7$ ). Excluding the five participants that failed the polygraph and did not disclose further information, the groups were compared on risk, Antisocial personality characteristics, Neuroticism and Conscientiousness. There were no significant findings.



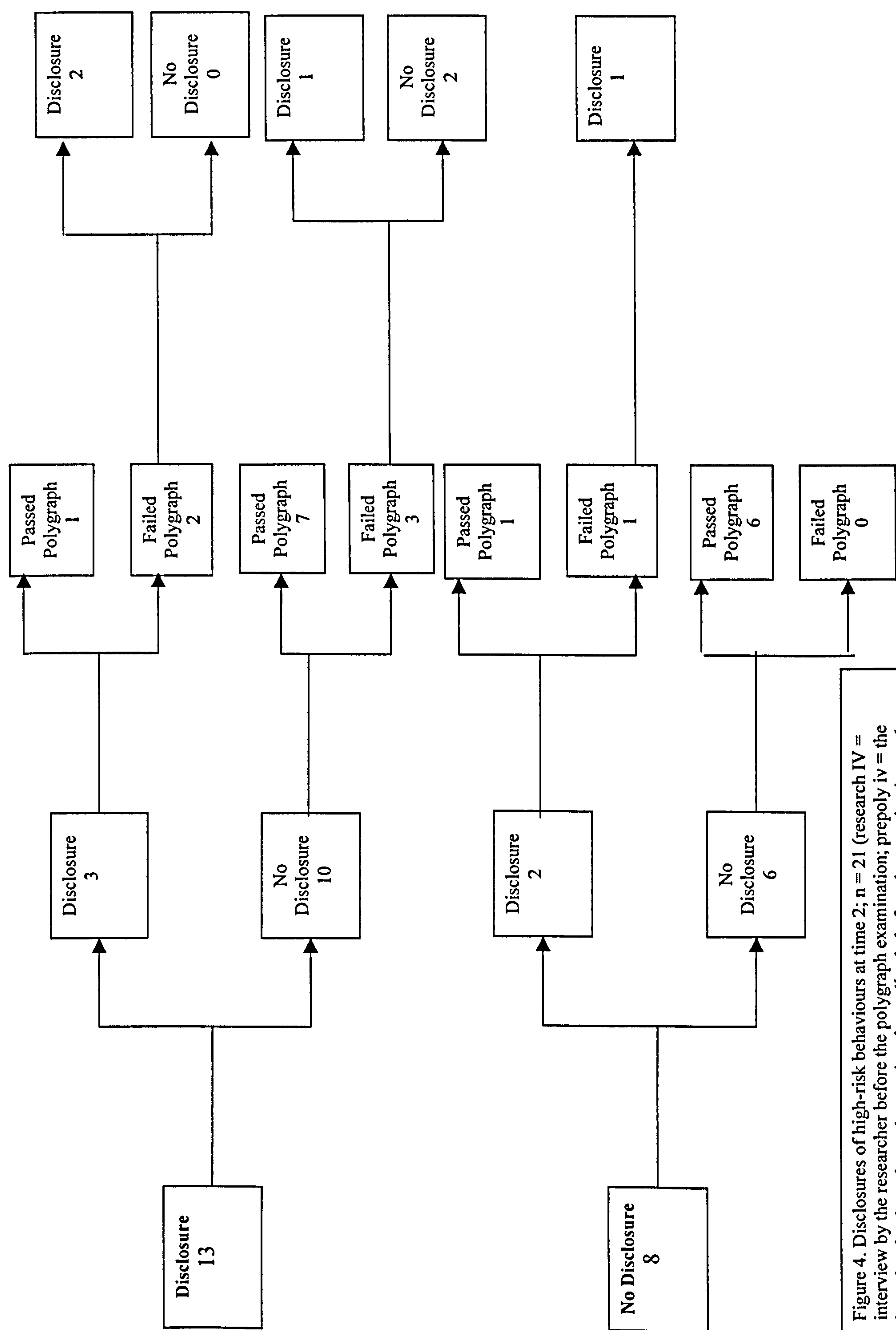


Figure 4. Disclosures of high-risk behaviours at time 2; n = 21 (research IV = interview by the researcher before the polygraph examination; prepoly iv = the interview by the polygraph examiner immediately before the examination; poly exam = the polygraph examination; post poly iv = the interview by the polygrapher following a deception indicated (i.e. failed) polygraph).

### *Self-report*

At Time 2, the 21 subjects completed a short questionnaire about their perceptions of the polygraph (PEDQ). Twenty men (95%) reported that the polygraph had been of benefit to them, although only 11 men (52%) reported that knowledge of an impending polygraph examination assisted them with avoiding their identified risk behaviours 'most' or 'all of the time' during the research period. Eleven men (52%) admitted that knowledge of the polygraph led them to report high-risk behaviour to their supervisors on at least some occasions that they, otherwise, would not have disclosed.

### *Non-attendees*

Seventeen participants did not attend for their first polygraph examination at Time 1, of whom two had been returned to custody (one for a sexual offence and the other for a breach of his probation conditions). Those who did not attend did not differ from those who did in terms of risk as determined by Static-99, normal personality, personality disorder symptomatology or psychological distress. Similarly, the 11 subjects who defaulted at Time 2 did not differ from the ones who attended. It is of interest that two participants gave as reasons for dropping out of the study, the reprimands they received from their supervisors in relation to the behaviours they reported at Time 1.

### *Main findings*

The findings showed that nearly all the participants reported engaging in high-risk behaviour during the research. Whilst the majority of these behaviours involved deviant fantasies, a significant portion included more serious victim acquisition-type behaviours, including actual contact with previous victims. Only one participant had reported his high-risk behaviour to his immediate supervisor. Comparison between the two conditions showed that there was no significant difference in the number or seriousness of behaviours reported. This finding initially appears to imply that 'knowing' about the impending polygraph test had little impact upon the participant's behaviour. At the second polygraph examination, however, participants reported significantly less behaviour and the behaviours reported were less serious. Also at this time volunteers were more likely to 'pass' the polygraph (i.e. be truthful) and disclose all their risk behaviour prior to the polygraph test. This suggests that the polygraph may have acted as a deterrent for them only after they had experienced an examination. This appears partly consistent with self-reports that indicate that the majority found that the polygraph was of some benefit. The high drop out rate possibly suggests that those motivated not to reoffend found polygraphy useful, while those less motivated sought to avoid it.

## Study 2

### *Method*

#### *Overview*

The purpose of the second study was to investigate the accuracy of the polygraph as used in a post-conviction context with sex offenders. Subjects were engaged in treatment and were mandated to undertake bi-annual polygraph tests focussed upon offending and other risk behaviours. Participants were seen for a single occasion where they completed a number of questionnaires and were interviewed about their current circumstances. The participant's regular polygraph maintenance test was used for the study, however, in addition to the regular issues covered in this test the examiner included 'drug use' over the preceding three months as a relevant question. Immediately after the polygraph test a hair specimen was collected and subsequently analysed for drugs. The polygraph charts were examined by a number of independent and 'blind' examiners, with their decisions about deception or non-deception compared with the objective information of the drug test.

#### *Participants*

Three hundred and twenty-one sex offenders participating in community-based treatment programmes throughout Georgia in the US were approached, of which 175 (55%), including 3 females, agreed to take part. The age ranged from



18 to 82 years ( $M = 40$ ,  $SD = 12.6$ ). One hundred and forty-four were Caucasian (81%), 28 African-American (16%), one was Hispanic, one American-Indian, and two were from other backgrounds. Sixty-seven (38%) identified themselves as presently single or separated, 79 (45%) as either married or co-habiting with a partner, and 30 (17%) reporting being in a relationship though not co-habiting.

With regards to education, 30 (17%) participants reported that their highest educational achievement was junior high school, 49 (28%) had completed high school or an equivalent diploma, 48 (27%) had completed or attended a technical or business school, 36 (20%) had completed or commenced a university degree, and 9 (5%) had completed or commenced a post-graduate university degree.

Excluding their current sex offender treatment 68 (40%) participants reported having received previous psychiatric or psychological treatment. Of these 39 (22%) reported problems with depression, 16 (9%) with anxiety, 9 (5%) with manic depression and 4 (2%) reported having been diagnosed with schizophrenia.

One hundred and sixty-six participants (94%) had been convicted of a sexual offence, 4 (2%) had not been convicted and were awaiting trial. Two participants had not been convicted or charged for any sexual offences. One was receiving treatment for professional misconduct and another for 'sex addiction'. The mean time on probation was 33.7 months ( $SD = 31$ ), with a range from 1 to 204 months. During this time a total of 29 (17%) had been imprisoned for probation violations. In regard to offence characteristics 150 (87%) had been convicted of contact offences, this including 137 (80%) that had offended against

child victims, 12 (7%) against adult victims and 1 against both. Sixteen (9%) participants had been convicted of non-contact offences, including manufacture or possession of child pornography, indecent exposure and voyeurism (i.e. 'peeping – tom'). The mean length of time in sex offender treatment was 23.5 months (SD = 23), with the range being from 1 to 120 months.

Twenty-eight (16%) participants reported having a previous drug-related conviction. Ninety-eight participants reported having had at least one drug test whilst on probation, the mean number of drug tests was 4.2 (SD = 13.1).

### *The polygraph examination*

The polygraph examiners that regularly conducted the sex offender testing for the services participating in the study were used. In total four examiners conducted polygraph tests for the study. All were members of the American Polygraph Association (APA) and accredited to conduct post-conviction polygraph tests. The examiners were instructed to conduct the examination as they normally would and to treat the drug issue as if they had been asked to investigate the area by the participant's treatment provider or probation officer. They were given a list of the names of the volunteers, and were asked to identify these individuals prior to them attending for the polygraph test.

Excluding the last seven days the examiners were told to determine whether the participant had used any of the following substances in the last three months: heroin, cocaine, cannabis, phencyclidine (PCP), methamphetamine and



amphetamine. They were each given a sheet with the different terms often used for these substances (e.g. dope, grass, speed, smack, skunk, angle dust etc.) and were asked to also refer to these terms when questioning subjects to ensure that participant's understood the drug terms.

The polygraph tests were conducted at the examiner's office or at a room at the participant's treatment facility. The MGQT protocol was used. The typical pre-test format involved asking the subject about their behaviour since the last polygraph examination. This interview would focus upon the participant's idiosyncratic offending and risk behaviours, in addition to more general sexual behaviour, such as the use of pornography and nature of masturbatory fantasies. Drug and alcohol use was also explored in the pre-test interview. The typical relevant questions used addressed whether the participant had re-offended or had engaged in high-risk behaviours (unless the participant's treatment provider or probation officer had specifically requested an alternative relevant question). Probable-lie comparison questions were used. In regards to the drug issue the suggested relevant question was:

“Other than what you have told me about, have you used any other type of drugs in the last three months?”

After completion of the polygraph test the examiner evaluated the charts. If the outcome was ‘deception indicated’ and the participant denied any drug use, another polygraph examination was immediately conducted. This second examination utilised the DoDPI Zone protocol (as detailed in chapter 4) and

focussed specifically on drug use over the last three months. This 'breakout' exam took between 15 to 20 minutes to complete. Again the examiner evaluated the charts, and if these were 'deception indicated' the examiner would interview the participant about drug use.

At the completion of the polygraph exam the participant complete the Depression Anxiety Stress Scale (DASS-42) and the examiner collected a hair specimen. This hair sample was sent to a laboratory and analysed for the previously mentioned substances. The result of the analysis was sent to the researcher and not shared with participant's polygraph examiner, treatment provider or other professionals involved with the individual.

### *Ethical Approval*

The Northumberland, Tyne and Wear Local Research Ethics Committee were approached for ethical approval, however, because the research was being conducted in the United States it was beyond their jurisdiction. The committee did report that the study would have been considered favourably. I then approached the Institutional Review Board (IRB) at the University of Georgia. They recommended that I approach the individual organisations that conducted the study with, and their IRB granted ethical approval.



## Materials

### *Risk & Personality*

Static-99 and Revised NEO-Personality Inventory (NEO-PI-R) as used in Study 1, was again used to evaluate risk of a re-offence and personality. Alpha coefficients for the NEO domains were between .67 (Openness) and .84 (Neuroticism).

### *Intelligence.*

The National Adult Reading Test (NART-2) is designed to provide an estimate of intellectual ability (Nelson & Willison, 1991). The NART-2 consists of a list of 50 words printed in order of increasing difficulty. The words are all irregular with respect to common rules of pronunciation, in order to minimise the possibility of reading by phonetic decoding rather than word recognition (e.g. ACHE, NAÏVE, ZEALOT). The participant reads aloud down the list of words and the test administrator records the number of errors made. Slight variations in pronunciation are acceptable when these are due to regional accents. These scores are then totalled to give an overall score out of 50. The NART-2 takes up to 10 minutes to complete.

Research has shown the NART-2 to be a valid measure of general intelligence (e.g. Ryan & Paolo, 1992; Weins, Bryan & Crossen, 1993). The NART-2 is considered to be among the most reliable tests in clinical use. When

internal consistency is considered, reliability estimates are above .9 (Crawford, Stewart, Garthwaite, Parker & Besson, 1988b). A test-retest reliability of .98 has been reported (Crawford, Stewart, Cochrane, Parker & Besson, 1989a). The NART-2 can be used with the general population and in a range of clinical populations, including dementia, depression and sex offenders (Crawford, Parker, Stewart, Besson & De Lacey, 1989; Crawford, Stewart, Cochrane, Parker & Besson, 1989; Fazel, Hope, O'Donnell & Jacoby, 2002). For North American samples Ryan and Paolo (1992) have developed regression equations to predict WAIS-R IQ scores from NART-2 error scores. These are as follows:

- Estimated Full Scale IQ =  $131.3845 + (\text{NART Errors}) (-1.124)$
- Estimated Verbal IQ =  $132.3893 + (\text{NART Errors}) (-1.164)$
- Estimated Performance IQ =  $123.3845 + (\text{NART Errors}) (-0.823)$

### *Psychological distress.*

The Depression, Anxiety and Stress Scale (DASS-42) is a 42 item self-report inventory that evaluates three factors – Depression, Anxiety, and Stress. Each of the three factors contains 14 items, divided into subscales of 2-5 items with similar content. The Depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia and inertia. The Anxiety scale assesses autonomic arousal, skeletal muscle effects, situational anxiety and subjective experience of anxious effect. The Stress scale is

sensitive to levels of chronic non-specific arousal. It assesses difficulty relaxing, nervous arousal, and being easily upset/agitated, irritable/over-reactive and impatient. Subjects are asked to use 4-point severity/frequency scales to rate the extent to which they have experienced each state over the past week. Scores for Depression, Anxiety and Stress are calculated by summing the responses for the relevant items (Lovibond & Lovibond, 1995). Exploratory and confirmatory factor analyses have sustained the proposition of the three factors (Brown, et al., 1997). The DASS anxiety scale correlates .81 with the Beck Anxiety Inventory, and the DASS Depression scale correlates .74 with the Beck Depression Scale (Lovibond & Lovinbond, 1995b). Reliability of the three scales is considered adequate and test-retest reliability is likewise considered adequate with .71 for depression, .79 for anxiety and .81 for stress (Brown, Chorpita, Korotitsch & Barlow, 1997). In the present study the alpha coefficients were between .84 (Anxiety) and .95 (Depression).

#### *Previous experiences with the polygraph.*

A 12-item survey, the Previous Experiences of the Polygraph Questionnaire (PEPQ), was developed for the study to gather descriptive information about participants' previous experiences and perceptions of the polygraph. The questionnaire is divided into three broad sections. *Section One* addresses false-positive and false negative rates, false admissions, and the use of countermeasures, *Section Two* the extent to which the participant considered the polygraph to be helpful in assisting him to avoid risk behaviours and re-offending and to engage in treatment, and *Section Three* the participant's perception of



polygraph accuracy overall. The PEPQ took approximately 10 minutes to complete. See appendix F for PEPQ.

### *Hair analysis.*

‘Hair drug testing’ is the process of releasing embedded drugs and drug metabolites through the chemical destruction of the hair fibre and then examining the resulting extract using radioimmunoassay (RIA) and gas chromatography/mass spectrometry (GC/MS) techniques (Travis, 1995). It remains unclear exactly how drugs enter the hair, although the most likely routes include:

- diffusion from blood into the hair follicle and subsequent binding to hair cell components,
- excretion in sweat which bathes hair follicles and hair strands,
- excretion of oily secretions into the hair follicle and onto the skin surface (Ditton, 2002).

Studies have reported sensitivity and specificity rates between 93 to 96% for the following substances: heroin (e.g. Goldberger, Caplan, Maguire & Cone, 1991), cocaine (e.g. Garside & Goldberger, 1996; Kauert & Röhrich, 1996), PCP (Sakamoto & Tanaka, 1996), cannabis (Cirimele, 1996), methamphetamine and amphetamine (Nakahara & Kikura, 1996). Hair testing is considered most reliable for detecting drug use that occurred 7 to 90 days previously (Travis, 1995).

Approximately 3.5 centimetres of hair (or 90 to 120 strands) needs to be gathered



from individual's head. Other types of hair (i.e. arm, chest) can also be used if necessary.

## Procedure

Participants were approached while attending their regular treatment group. These groups varied in size from 5 to 17. All treatment programmes used polygraphy, and required individuals to complete a minimum of two polygraph tests per year. Participants were initially told that the broad purpose of the research was to investigate the accuracy of the polygraph in a post-conviction context. Individuals were then given an information sheet that described the nature and purpose of the study (see Appendix D). Participants were told that they would be asked to complete a number of questionnaires and for their next polygraph test an additional question about drug use would be included. At the completion of this test they would be asked to supply a hair sample that would be analysed for drugs. The subjects were told that the result of the drug test was confidential, although any admissions regarding drug use to the polygraph examiner would be treated as a 'normal admission' and be reported to the relevant professionals. Because probationers were required to pay for their polygraph tests as an incentive to participate subjects were offered a \$25 discount on the polygraph. Participants were told that participation was voluntary, and they could withdraw at any stage without consequence or providing a reason. Volunteers either commenced participation immediately after the introduction or arranged a suitable time to start.

Prior to commencing the study all participants signed a consent form that detailed the nature of the research and the limits of confidentiality (see Appendix D). Participants were seen on a single occasion by the researcher for up to 60 minutes. In this time they completed the NEO-PI-R and the PEPQ, either by themselves or with other participants in a quiet room at their treatment facility. The participants were then individually interviewed by the researcher about their present circumstances and previous experiences with the polygraph. The NART-2 was administered during this interview. At the completion of this interview participants were again informed that they would be asked a question about drug use on their next polygraph and would also be asked to supply a hair sample for drug testing. They were again reminded that the result of this hair test was confidential, however, any disclosure made to the polygraph examiner was not confidential and would be reported to the relevant professionals. Because the participant's regular polygraph test was used in the research, the times when participants were tested varied depending upon when their test had been scheduled. This meant that some participants were tested shortly after being interviewed (i.e. days or weeks), whilst others were not seen for several months.

## Results

Data entry was verified by a random check across all instruments in the study of 10% of the data. Less than .01% error was found. SPSS was used for data entry and analysis. The sample size for suitable power (0.8) was determined with consideration of a sensitivity and specificity methodology. The following equation was used (Cohen, 1968):

$$\{100^2 \times 4 \times P (1-p)\} / A^2 = n$$

P = proportion of false positives and negatives

A = margin of error

With consideration of previous research a conservative false positive and negative rate of 15% was adopted with a confidence interval of 10%, giving an overall sample size of 96, with 48 participants in each condition (i.e. truth tellers and liars). Considering that subjects would be less likely to be engaging in drug abuse whilst on probation, it was expected that more subjects would need to be approached to gain sufficient numbers in the 'lying' condition (i.e. the drug takers). The data collection was stopped at 80 completed drug tests, because only 5 participants had returned 'positive' drug tests. At a drug-taking base rate of 5% it could be expected that only an additional 4 of the remaining 96 participants would have been taking drugs. Volunteers were still given a rebate on their regular polygraph even if they were not asked to supply a hair sample or had a question about drug use on their exam.



Of the 80 tests that were conducted, only 70 charts could be re-scored by blind examiners. Three charts were excluded because of an inconclusive drug test outcome. Another volunteer was drug tested, however, the examiner forgot to include a question about drug use on the polygraph exam. Six charts were irretrievable due to computer malfunction. Thus, the 70 remaining charts represented 40% of the volunteers, and 22% of the overall sample approached.

The results section is broadly divided into three sections. The initial section provides descriptive data on measures used in the study. The second section presents information on the sensitivity and specificity of the polygraph for detecting drug use. Univariate analyses were used to identify predictors of true negatives and false positives. The third section presents data from the questionnaire component of the study.

### *Characteristics of sample*

According to the Static-99, 93 (58%) participants were assessed as being a low risk of a sexual or violent re-offence, 46 (29%) were evaluated as medium-low, 19 (11.8%) as medium-high and only 3 (2%) as high risk. Table 13 displays the means for IQ, the number of previous post-conviction polygraph tests and also the outcome of the DASS-42 scales. Thirteen participants did not complete the NART-2 because they were illiterate or did not have their reading glasses with them, and three participants left the interview room prior to being interviewed about their demographic circumstances.



Table 13

*Sample characteristics: intelligence, previous polygraph tests & DASS-42.*

Variable	Mean (SD)	Range
IQ (n =163)	102 (11.9)	75 – 128
Number of post conviction polygraph tests (n =175)	1.5 (2)	1 – 20
Dass-42 (n=70)		
Stress	7.4 (7)	0 – 33
Anxiety	4.2 (4.9)	0 – 21
Depression	5.4 (8.1)	0 – 36

Twenty-four participants either did not successfully complete the NEO-PI-R or produced an invalid profile. These were excluded from any analyses. Table 14 displays the means, standard deviations and the qualitative range of the scores for the overall sample. Like the participants in study 1 the FFM profile for these sex offenders was characterised by high Neuroticism and low Conscientiousness. Indicating that this sample of sex offenders are characterised by emotional maladjustment and a lackadaisical approach to work and other commitments. In addition, this sample also scored low on Openness suggesting that these individuals are conventional in behaviour and conservative in outlook.

Table 14

*NEO mean scores, standard deviations and qualitative descriptions for overall sample study 2 (n=142)*

Domain & Facet Scales	Mean (SD)	Qualitative range
<b>NEUROTICISM</b>	<b>87 (21)</b>	<b>High</b>
Anxiety	15 (5)	Average
Anger Hostility	13 (5)	Average
Depression	16 (5)	High
Self-consciousness	16 (4)	Average
Impulsiveness	15 (4)	Average
Vulnerability	11 (4)	Average
<b>EXTRAVERSION</b>	<b>101 (16)</b>	<b>Average</b>
Warmth	21 (4)	Average
Gregariousness	14 (4)	Average
Assertiveness	15 (4)	Average
Activity	17 (4)	Average
Excitement Seeking	16 (5)	Average
Positive Emotions	18 (4)	Average
<b>OPENNESS</b>	<b>100 (15)</b>	<b>Low</b>
Fantasy	14 (4)	Average
Aesthetics	16 (4)	Average
Feelings	18 (3)	Average
Actions	15 (3)	Average
Ideas	18 (5)	Average
Values	18 (4)	Low
<b>AGREEABLENESS</b>	<b>120 (15)</b>	<b>Average</b>
Trust	18 (4)	Low
Straightforwardness	21 (4)	Average
Altruism	23 (3)	Average
Compliance	18 (5)	Average
Modesty	20 (4)	Average
Tender-mindedness	21 (3)	Average
<b>CONSCIENTIOUSNESS</b>	<b>114 (17)</b>	<b>Low</b>
Competence	20 (3)	Low
Order	17 (4)	Average
Dutifulness	21 (4)	Low
Achievement Striving	19 (4)	Average
Self-discipline	21 (4)	Average
Deliberation	17 (4)	Average

## Reliability

### *Numerical scores*

To assess the consistency of the numerical scoring of the polygraph charts, Pearson correlation was computed between the blind re-scorers. The correlations for the individual question scores ranged between .92 to .97, with the mean being .94. The alpha for the drug issue was .97. It would have been ideal to correlate the numerical scores of the blind examiners with those of the original examiners. However, the original examiners did not provide these scores.

### *Classifications*

To assess the reliability of classifications, the agreement between the original examiner's opinion (truthful, deception or inconclusive) for each test and the classification based on the mean of the blind re-scorers were calculated. The three re-scorers for each test were averaged to provide a single, more stable blind score. For the purposes of this analysis, subjects were classified as deceptive if their numerical score was  $-3$  or lower. Subjects were classified as truthful if they scored  $+3$  or higher. Cases that failed to meet the criteria for deceptive or truthful were classified as inconclusive. The kappa coefficient (Cohen, 1960) between the original examiners classifications and the blind scorers was very poor at  $-.04$ . The inter-rater reliability between the blind re-scorers, however, was reasonable at  $.72$ .

The original examiners scored 54 (77%) charts as truthful, 12 (17%) as deceptive and only 4 (6%) as inconclusive. The blind examiners classified 37 (53%) charts as truthful, 14 (20%) as deceptive and 19 (27%) as inconclusive. A chi squared test revealed that the original examiners were more likely to score charts as truthful compared to the blind examiners,  $\chi^2(1, 70) = 5.09, p < .05$ . This suggests that the original examiners maintained a positive bias in their decision making (i.e. tendency to evaluate charts as non-deceptive). Post hoc power calculations revealed that power was .70 for this analysis.

Table 15 compares the classifications for the blind and original examiners. There was agreement in 41 (59%) cases (6 deceptive, 2 inconclusive and 33 truthful). The examiners arrived at opposite conclusions (truthful versus deceptive) in 10 (14%) cases, while for the remaining 19 (27%) tests one of the outcomes for either the original or re-scorers was inconclusive.

Table 15  
*Original Examiner Classifications Versus Blind Examiners Classifications*

Original Examiner's Classifications				
	Truthful	Deception	Inconclusive	Total (%)
Blind Re-scorers				
Truthful	33	2	2	37 (53)
Deceptive	8	6	0	14 (20)
Inconclusive	13	4	2	19 (27)
Total (%)	54 (77)	12 (17)	4 (6)	70



## Accuracy

The accuracy of the outcomes was computed by comparing classifications based on original examiner opinions and mean blind rescores with drug test outcome. Score cut-offs were the same as described in the preceding section. Table 16 and 17 displays the results for the original and blind examiners. Excluding the inconclusive outcomes the original examiners correctly identified truth tellers (i.e. no drug use) 84% of the time (specificity) and wrongly accused participants of taking drugs on 16% of occasions (false positives). Five participants were identified to have been taking drugs through the hair testing. The examiners correctly identified only two participants (40%). When considering the accuracy of the two test outcomes (Passed versus Failed) for the original examiner: 95% of the 'passed' outcomes were correct (Negative predictive value), compared with only 15% of 'failed' outcomes (Positive predictive value). The blind examiners had a specificity rate of 79% and a false-positive rate of 21%. The sensitivity was 100% (one 'liar', that is drug taker, was classified as inconclusive). Failed test outcomes were correct only 29% of the time, while 100% of the passed outcomes were correct.

Table 16

*Accuracy rates for Original & Blind examiners*

Criterion		
Classification	Deceptive (Drugs)	Truthful (No drugs)
Original opinion		
Truthful	3	51
Deceptive	2	10
Inconclusive	0	4
Blind re-score		
Truthful (Pass)	0	37
Deceptive (Fail)	4	10
Inconclusive	1	18

Table 17

*Sensitivity, Specificity, False negative & Positive probability, Positive & Negative Predictive values for Original & blind examiners.*

	Original	Blind
	(%)	(%)
Sensitivity	40	100
False negative probability	60	0
Specificity	84	79
False positive probability	16	21
Positive predictive value	15	29
Negative predictive value	95	100

### *Receiver operating characteristic (ROC)*

A receiver operating characteristic (ROC) curve was generated for the original examiners and the blind re-scorers. The ROC curves were generated by comparing the drug test results with the polygraph classifications. The area under each of the ROC curves was computed. This area is a statistic that assumes values between 0 and 1, with an area of 0.5 indicating that the two distributions are undifferentiated, and an area of 1 indicating that there is no overlap between the two distributions. Bamber (1975) showed that the area under a ROC curve has an asymptotic normal distribution. He described a method for estimating the variance of the area statistic and computing confidence intervals for the true area. Using Bamber's method, a 95% confidence interval was computed for each area. The areas under the ROC curves, as well as the corresponding 95% confidence intervals are presented in Table 18.

Table 18 reveals that both the original and blind examiners were reasonably good at distinguishing truthfulness from deception (the area in both are significantly better than a chance area of .50). The blind re-scorers yielded a significantly larger area under the ROC curve than the original examiners (.88 versus .68).

Table 18

*Area under the ROC Curve and a 95% Confidence Interval computed for the Original examiners and Blind re-scorers.*

Examiner	Area	95% confidence interval
Original	.68	.37 - .99
Blind	.88	.78 - .98

*Note.* ROC = receiver operating characteristic

*False positives*

Both the original examiners and blind rescorers wrongly accused 10 volunteers of taking drugs. Four participants were wrongly identified as taking drugs by both sets of examiners. An effort was made to identify factors associated with false positive errors (i.e. innocent participants misclassified as deceptive). One question of interest was whether there were unique subject characteristics associated with innocent subjects who produced deceptive test results as compared with subjects who produced truthful results. Three broad categories of variables were investigated: ‘psychological’ characteristics, ‘demographic’ and ‘self-report’ features.

The psychological characteristics included the five-factor model domains (neuroticism, extraversion, openness to experience, agreeableness and conscientiousness), psychological distress (as measured by the DASS) and intelligence. Four demographic variables were considered: age, number of previous polygraph tests, ethnic origin (Caucasian versus non-Caucasians) and previous psychological treatment. Finally the self-report addressed: whether the



offender had ever experienced a sanction for a polygraph exam, whether they reported it helpful, claimed to be a false positive or negative and also how accurate they believed the polygraph to be. These variables were examined separately for original and blind examiners. Table 19 and 21 displays the means, standard deviations and other figures for these variables.

### *False positives (blind examiners)*

To examine whether the psychological variables differed between the false positive (n=10) and accurately identified truthful individuals (n=37), a multivariate analysis of variance (MANOVA) was performed on the seven variables. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There was no significant main effect,  $F(7, 34) = 1.34, p > .05$ . Table 19 displays the mean scale scores for the two groups. Only one of the variables (Intelligence) showed a significant difference between the false positives and accurates:  $F(1, 41) = 5.3, p < .05$ . Inspection of the mean scores showed that false positives (IQ = 92) were less intelligent than the participants who were accurately identified. This finding needs to be regarded with a degree of caution as power was calculated to be only .36 for this analysis.

Separate chi squared and MANOVAs were used to examine the demographic and self report variables. With regards to the demographic variables there were no significant results for Ethnic origin or previous psychological

treatment. When considering the MANOVA there was no significant main effect,  $F(2, 43) = 1.74, p > .05$ , neither was there results when the dependent variables were considered separately. With regards to the self-report variables, the false positives were more likely to have experienced a sanction (40% compared to 10%) than volunteers who were scored correctly. MANOVA did not yield a significant outcome for the remaining variables (see table 19).

A direct logistic regression was carried out to examine the utility of the variables that distinguished the two groups in the prediction of group membership with respect to false positive outcome (i.e. false positive versus accurate). Intelligence and having ever experienced a sanction were entered together. There was a good model fit (i.e. discrimination between the groups) as measured by the Hosmer and Lemeshow Test:  $\chi^2(8, n = 51) = 9.37, p > .05$ . A test of the full model with both predictors against a constant only model was statistically significant,  $\chi^2(2, n = 51) = 7.6, p < .05$ , indicating that the predictors, as a set, reliability distinguished between false positives and subjects accurately identified by the polygraph. Correct classification of the cases excluding the predictor variables was high at 81.4% (100% accurate and 0% for false positives), when the predictor variables are included prediction success remains at 81.4% (97.1% accurates and 12.5 false positives), the model displayed slightly better predictive success for the false positives (0% versus 12.5%). Nonetheless the variance accounted for is small at 18%. Table 20 displays the Wald statistics and odds ratios. According to the Wald criterion, when the predictor variables were considered separately neither reliably predicted status.

Fourteen (20%) of the 70 examinees disclosed having experienced a sanction. Of these 9 (64%) were either evaluated as being inconclusive ( $n = 5$ , 34%) or a false positive ( $n = 4$ , 30%). In total, examinees that had been sanctioned were significantly more likely to be incorrectly scored, either as a false positive or as an inconclusive ( $\chi^2(1) = 3.1, p < .05$ ). When considering the types of sanctions that the false positives had experienced: five volunteers (50%) had to address additional issues in treatment; two (20%) had experienced an unspecified sanction; two (20%) had their supervision times increased, and another had curfews imposed. The remainder did not specify the type of sanction they had experienced.



Table 19

*MANOVA Means, standard deviations and numbers for variables used to examine differences between false positives and accurates for the blind scorers.*

Blind Examiners	False positives (n=10)	Accurate (n=37)	
Psychological (SD)			
Neuroticism (FFM)	99.9 (21)	87.9 (18.5)	F =1.04, ns
Extraversion (FFM)	97.7 (12.5)	102.6 (16.7)	F = .62, ns
Openness to experience (FFM)	99.2 (12.5)	101 (14.1)	F = .01, ns
Agreeableness (FFM)	120.4 (15.6)	119.8 (11.3)	F = .72, ns
Conscientiousness (FFM)	113.6 (17)	114.2 (14.6)	F = .07, ns
Intelligence (SD)	92.4 (10.6)*	102.2 (11)*	F = 5.3, p < .05
Psychological distress (SD)	19.3 (14)	13.2 (16.4)	F = .21, ns
Demographic			
Age (SD)	38.3 yrs (9.6)	38.8 yrs (12)	F = .23, ns
Number of previous polygraph tests (SD)	1 (0)	.72 (.42)	F = 1.74, ns
Ethnic origin (% ,Caucasian)	100	92	$\chi$ (1) = .86, ns
Previous psychological treatment (%)	60	43	$\chi$ (1) = .88, ns
Self report			
Sanction (%)	40*	10*	$\chi$ (1) = 4.79, P < .05
Previously inaccurate (%)	10	14	$\chi$ (1) = .08, ns
Helpfulness (SD)	10.8 (4)	9.9 (3.9)	F = .30, ns
Accuracy (SD)	3.8 (1.3)	3.2 (1.1)	F = 1.2, ns

Table 20.

*Logistic regression analysis of polygraph accuracy as a function of intelligence and having previously experienced a sanction (blind examiners)*

Variable	B	Wald	Sig.	Odds ratio
Sanction	-1.62	2.59	Non-sign	.19
Intelligence (NART IQ)	-.09	3.81	Non-sign	.91
Constant	8.52	3.41	Non-sign	



### *False positives (original examiners)*

A series of MANOVA and chi squared were used to examine the difference between the false positives and accurates as evaluated by the original examiners on the three variable categories. These results are displayed in table 21.

The first MANOVA examine the psychological variables. No significant main effect was found for status (i.e. false positive or accurate),  $F(7, 48) = 1.34, p > .05$ . When the results for the dependent variables were considered separately, the only difference to reach statistical significance was the NEO scale of Extraversion:  $F(1, 50) = 9.87, p < .05$ . Inspection of the mean scores indicates that the subjects that were accurately scored were more extraverted than the false positives (102 versus 89). Power was, however, poor at only 0.49.

The second MANOVA used the demographic variables (age and number of previous polygraph tests). There was no significant main effect,  $F(2, 59) = .54, p > .05$ . When the dependent variables were considered separately there were no significant results: Age,  $F(1, 61) = .35, p > .05$  & Number of previous polygraph tests,  $F(1, 61) = .99, p > .05$ . There were no significant outcomes for the chi squared analyses.

The final MANOVA examined the self-report variables (Helpfulness of polygraph and Accuracy of the polygraph). There was no significant main effect,  $F(2, 56) = .26, p > .05$ . When the dependent variables were considered separately,

there also were no significant outcomes: Helpfulness,  $F(1, 58) = .09, p > .05$ ; and Accuracy,  $F(1, 58) = 1.02, p > .05$ .

A logistic regression was carried out to examine the utility of the NEO scale Extraversion in prediction of group membership with respect to status (i.e. false positive or accurate). Using Extraversion, there was a good model fit as measured by the Hosmer and Lemeshow,  $\chi^2(1, n = 61) = 4.47, p > .05$ . This model was shown to be significantly better than a constant-only model containing only the intercept and not the predictor variable:  $\chi^2(1, n = 61) = 4.57, p < .05$ . Correct classification of cases using Extraversion was 84% (accurates 100% & False positives 10%), this was only slightly better than the constant only model 83% (accurates 100% & false positives 0%). The Wald statistics and odds ratios are shown in table 22. These indicate that for Extraversion, a one-unit increase in this predictor variable will decrease the likelihood of a false positive outcome by 4%.

Table 21

*MANOVA Means, standard deviations and numbers for variables used to examine differences between false positives and accurates for the original examiners.*

Original Examiners	False positives (n=10)	Accurate (n = 51)	
Psychological (SD)			
Neuroticism (10 versus 47, FFM)	95 (20)	88.5 (19)	F = .6, ns
Extraversion (FFM)	88.8 (19.6)	101.7 (16.3)	F = 9.87, P < .05
Openness to experience (FFM)	101.4 (16.4)	100.1 (13.0)	F = .01, ns
Agreeableness (FFM)	119.1 (14.7)	120.8 (11.4)	F = .12, ns
Conscientiousness (FFM)	112.1 (16.7)	114.3 (15.2)	F = .01, ns
Intelligence, (10 versus 47)	105.5 (15)	100.7 (11)	F = .37, ns
Psychological distress (10 versus 53)	15.4 (15)	15.4 (20)	F = .05, ns
Demographic			
Age (SD)	38.5 yrs ( 9.9)	41.1 yrs (12.3)	F = .35, ns
Ethnic origin (% ,Caucasian)	100	93	$\chi$ (1) = .49, ns
Number of previous polygraph tests (10 versus 52, SD)	1 (.8)	1.9 (3)	F = .99, ns
Previous psychological treatment (%)	60	34	$\chi$ (1) = 2.4, ns
Self report			
Sanction (%)	21	30	$\chi$ (1) = .41, ns
Helpfulness (SD)	9.7 (5.1)	10.2 (4.2)	F = .09, ns
Previously inaccurate (%)	20	15	$\chi$ (1) = .15, ns
Accuracy (SD)	3.8 (1.0)	3.5 (1.2)	F = -1.1, ns

Table 22.

*Logistic regression analysis of polygraph accuracy as a function of extraversion (original examiners)*

Variable	B	Wald	Sig.	Odds Ratio
Extraversion	-.043	4.206	Sign	.958
Constant	2.547	1.667		



Self-report

Accuracy

One hundred and twenty-one participants (69%) reported having completed a total of 263 polygraph tests whilst on probation. Table 23 displays the various classifications for these tests. Going by the participants self-report, therefore, the specificity of these tests was 85% and the false positive rate 15%. Sensitivity was 84% and the false negative rate 16%. Ninety-seven percent of ‘passed’ test outcomes were correct in comparison to only 47% of ‘failed’ outcomes. A receiver operating characteristic (ROC) curve was also generated for the self-report data. The ROC statistic was .84 (.77 - .91, 95% confidence interval).

Considering only subject numbers shows that 27 participants claimed to have previously been a false positive, 6 reported to have been a false negative, 69 a true negative (i.e. a truth teller and correctly identified by polygraph) and 19 were true positives (i.e. a liar and correctly identified by the polygraph).

Table 23

*Self-reported accuracy rates for post-conviction tests*

Polygraph Test Result	True Condition		Total
	Deceptive	Truthful	
Deception Indicated	32	34	66
No Deceptive Indicated	6	191	197
Total	38	225	263



*False positives (self-report)*

Individuals who reported telling the truth but were wrongly labelled as deceptive (false positives; n = 27) were compared with those who said they had been correctly classified as telling the truth (true negatives; n = 64). Relevant variables were grouped into two broad categories: *historical* (age, ethnic origin (Caucasian versus other) and previous psychological treatment) and *psychological* (personality and intelligence). There were no significant findings (see table 24).

Table 24  
*Means, standard deviations and numbers for variables used to examine differences between self-reported false positives and true negatives for the original examiners.*

	False positives (n=27)	True negatives (n = 64)	
Psychological (SD)			
Neuroticism (19 versus 55, FFM)	89.7 (20.4)	85.4 (19.6)	F = .11, ns
Extraversion (FFM)	101.9 (19.1)	97.9 (15.8)	F = .99, ns
Openness to experience (FFM)	94.6 (11.2)	99.3 (12.9)	F = .35, ns
Agreeableness (FFM)	116.2 (15.7)	121.1 (14.4)	F = 2.33, ns
Conscientiousness (FFM)	110.4 (24.3)	114.7 (13.3)	F = 1.17, ns
Intelligence (24 versus 60)	101.6 (12)	101.2 (11.9)	F = .83, ns
Historical			
Age (SD)	40.6 yrs (9.6)	41.1 yrs (12.6)	Z = -.04, ns
Ethnic origin (% , Caucasian)	85	80	$\chi$ (1) = .37, ns
Previous psychological treatment (%)	33	37	$\chi$ (1) = .7, ns

*False negatives (self-report)*

Individuals who claimed that they had been deceptive but were classified as ‘truthful’ (false negatives, n = 6) were compared with those who reported being deceptive and were accurately labelled as such (true positives, n = 29), and with those who said they had been correctly labelled as non-deceptive (true negative; n = 64). Univariate non-parametric analyses did not yield any significant results (see tables 25 and 26).

Table 25

*Univariate analyses for false negatives and true positives*

	False negatives (n=6)	True positives (n = 29)	
Psychological (SD)			
Neuroticism (5 versus 23, FFM)	91.2 (9)	81.6 (20.5)	Z = -1.02, ns
Extraversion (FFM)	102 (13)	105 (12.7)	Z = -1.1, ns
Openness to experience (FFM)	101.6 (13.1)	102.3 (14.7)	Z = -.21, ns
Agreeableness (FFM)	108 (24.1)	124.1 (12.3)	Z = -1.53, ns
Conscientiousness (FFM)	118 (14.5)	119.6 (15.5)	Z = -.06, ns
Intelligence (6 versus 27)	100.6 (17.7)	100.2 (10.9)	Z = -.16, ns
Historical			
Age (SD)	43.1 yrs (17.4)	44.5 yrs (12.9)	Z = -.15, ns
Ethnic origin (% , Caucasian)	83	83	$\chi$ (1) = .01, ns
Previous psychological treatment (%)	50	48	$\chi$ (1) = .01, ns

Table 26

*Univariate analyses for false negatives and true negatives*

	False negatives (n=6)	True negatives (n = 64)	
Psychological (SD)			
Neuroticism (5 versus 52, FFM)	91.2 (9)	85.2 (18.6)	Z = -.88, ns
Extraversion (FFM)	102 (13)	98.5 (15.6)	Z = -.18, ns
Openness to experience (FFM)	101.6 (13.1)	99.7 (13)	Z = -.60, ns
Agreeableness (FFM)	108 (24.1)	121.4 (14.7)	Z = -1.2, ns
Conscientiousness (FFM)	118 (14.5)	114.8 (13.7)	Z = -.52, ns
Intelligence (6 versus 27)	100.6 (17.7)	101.8 (12.2)	Z = -.4, ns
Historical			
Age (SD)	43.1 yrs (17.4)	41.8 yrs (12.5)	Z = -.33, ns
Ethnic origin (% , Caucasian)	83	79	$\chi$ (1) = .05, ns
Previous psychological treatment (%)	50	40	$\chi$ (1) = .24, ns

*Utility*

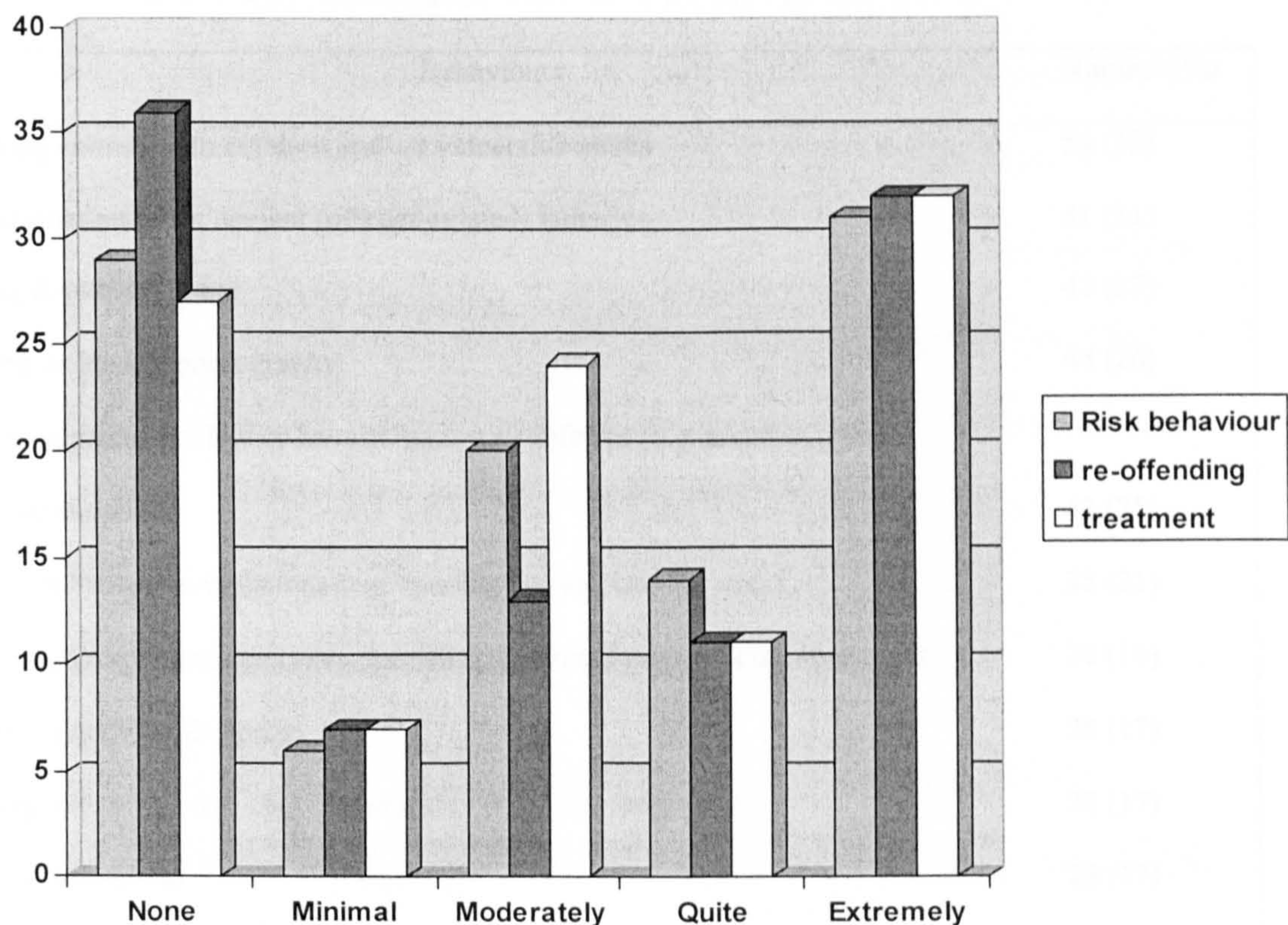
One hundred and fourteen of the 126 offenders who had been polygraph tested fully completed the PEPQ. Of these, 50 (44%) reported that they were more truthful with their probation officers and treatment providers than they otherwise would have been because of their experience of the polygraph; 39 subjects (34%) reported that it assisted them in being more truthful about their behaviours to families and friends. These findings are similar to those in the 45 subjects who had not yet been tested and fully completed the PEPQ, with 20 (44%) and 16 (36%) indicating that the expectation of a polygraph test increased their disclosures to probation officers and family and friends respectively.

Regarding behaviours associated with offending, 71 individuals (56%) who had previously been polygraphed reported that the polygraph was moderately to extremely helpful assisting them to avoid reoffending, 81 (64%) that it was useful in assisting them to avoid engaging in risk behaviours, and 84 (67%) that it was generally helpful in respect of treatment. Similar responses were given by those awaiting their first examinations. Responses for the whole sample in respect of the impact of polygraph on engaging in risk behaviours, actual re-offending, and engagement in treatment are shown in Figure 5.



Figure 5.

*Helpfulness of polygraph with avoiding risk behaviours and re-offences and with overall treatment? (%)*



Considering the group as a whole in regards to specific risk behaviours, 57 (33%) individuals reported they were less likely to masturbate to deviant (offence-related) fantasies, 53 (31%) that they were less likely to have contact with children and/or potential victims because of the polygraph, 47 (27%) that the polygraph contributed to less drug and alcohol use, and 44 (26%) that they were less likely to use or buy pornography. However, a significantly greater proportion of those who had undergone polygraph testing than those still awaiting their first test reported that they were less likely to visit places to view children (37 versus 5,  $\chi^2 = 5.9$ ,  $df = 1$ ,  $p < .05$ ) and to engage in other more general risk behaviours (18 versus 1,  $\chi^2 =$



4.2,  $df = 1$ ,  $p < .05$ ). Table 27 displays the risk behaviours that participants reported being less likely to engage in due to the polygraph.

Table 27

*Behaviours less likely to be engaged in due to polygraph (n = 168)*

Behaviours	Number (%)
Having contact with children and/ or vulnerable adults	54 (33)
Masturbation using deviant (offence-related) fantasies	51 (31)
Drug & alcohol usage	47 (27)
Using or buying pornography	44 (26)
Visiting places to view or have contact with children (e.g. arcades, gyms, swimming pools)	42 (25)
General Probation violations (e.g. missing curfew, leaving state)	35 (21)
Missing group meetings / not completing treatment related homework tasks	30 (18)
Visiting adult book stores	28 (17)
Using internet to visit chat rooms and/or access pornography	28 (17)
Use of prostitutes	28 (17)
Other identified risk behaviours	26 (16)
Collecting pictures of children for masturbation purposes	23 (14)
Engaging in extra-marital affairs or sexual relationship outside primarily relationship	23 (14)
Use of telephone sex lines	22 (13)

There was no difference in the perception of the polygraph’s accuracy between the subjects who had previously been polygraphed and the ones who had not. Overall, 16 subjects (10%) considered it to be no more accurate than chance, 15 (9%) slightly accurate, 73 (44%) ‘moderately’ accurate and 63 (38%) rated it as being ‘quite’ to ‘extremely’ accurate.

### *Sanctions*

Twenty-seven (22%) out of 121 men who had completed a post-conviction polygraph exam reported experiencing a direct sanction due to a polygraph result or a disclosure made during a test, the most common of which involved having to address additional issues in treatment or supervision (78%). Four of these individuals (15%) claimed that there was a change to their supervision conditions: two that their treatment was terminated and two a change in living circumstances (e.g. reduced contact with family). There was no relationship between having experienced a sanction and claiming to have been a false positive ( $\chi^2 = 3.07$ ;  $df = 1$ ;  $p > .05$ ).

To test whether having been sanctioned or erroneously classified (false-positive or negative) effected the participant's perception of its utility, an overall 'helpfulness' variable was created by combining the scores for the three utility scales. There was, however, no difference in perceptions of utility between those participants who were sanctioned and those who were not ( $t(111) = .38$ ,  $p > .05$ ), nor was there a difference between those who reported being false-positives and true negatives, or between the false-negatives and true positives.

### *Countermeasures and false admissions*

Only two participants (1%) claimed to have used drugs to beat the polygraph. Both also claimed to have previously been deceptive without being detected.

Twelve participants (9%) reported making false admissions regarding their behaviour at some stage during a post-conviction polygraph test, only five of whom claimed to have been wrongly labelled as being deceptive. The main reasons given for making a false admission were in three cases (25%) the fear of getting in trouble with their probation officers, and in another three cases (25%) feeling pressured by the polygraph examiner. In the remaining cases one man said he wanted to 'make a good impression', another was 'confused', one wanted to ensure that he passed the polygraph test and another wanted to demonstrate commitment to therapy. Two participants did not provide a reason.

A significant difference was found when a one-way between groups MANOVA was performed with the five NEO-PI-R domain scores as the dependent variables and 'having made a false admission' as the independent variable ( $F(5, 96) = 2.46, p < .01$ ). When results for the dependent variables were considered separately, two reached statistical significance using a Bonferroni adjusted alpha level of .01: Neuroticism [ $F(1, 102) = 10.08, p < .01$ ] and Conscientiousness [ $F(1, 102) = 7.85, p < .01$ ], with the false confessors having higher levels of Neuroticism (104 compared to 84) and lower levels of



Conscientiousness (101 compared with 116). Post hoc calculation revealed that this analysis had sufficient power at 0.88.

### *Main findings*

The polygraph was reasonably accurate with identifying truth telling (79%), while 21% were wrongly accused of drug taking when their drug test indicated otherwise. Only a small number of offenders ( $n = 5$ ) were found to be taking drugs and lying about having done so. The blind scorers correctly identified four of these individuals, and scored the fifth examinee as inconclusive. The Area under the curve index for these examinees was .88. False positives were more likely to have experienced a sanction due to a polygraph test and were less intelligent than others. The inter-rater reliability between the blind scorers and the original examiners was poor. It appears that the original examiners mistrusted the numerical scoring procedures and utilised extra-polygraphic information when making their evaluations. The original examiners were less accurate than the blind scorers (Area under the curve index = .68) and only correctly identified two of the five liars (40%). For the original examiners false positives were less extraverted than correctly identified examinees. When considering only the self-report data sensitivity was 84% and specificity 85%, whilst false positive and negative rates were 16 and 15% respectively. There was no difference in regards to personality or demographic characteristics between the offenders reporting accurate (i.e. true positives and negatives) and inaccurate outcomes (i.e. false positives and negatives). The majority of participants found the polygraph helpful with treatment and also with avoiding both re-offending and engaging in general risk

behaviours (e.g. having contact with children, masturbating to deviant fantasies). A significant number of participants also claimed to be more truthful with their treatment provider and probation officer because of polygraph testing.

A limited number of subjects had experienced a sanction because of a polygraph outcome. The most commonly reported sanction included having to address additional issues in treatment or supervision. Only a small number of participants claimed to have used countermeasures or made false admissions during a polygraph exam. Notably false confessors were more emotionally disturbed and negligent than the other participants.

## CHAPTER 7

### DISCUSSION

The discussion section is divided into three broad sections. The initial section will address the results related to utility (from study 1 and 2), the second section will focus upon accuracy, and the third section will discuss the overall implications of the findings.

#### The issue of utility

The first study was designed to address many of the problems of previous research by using a prospective design with comparison groups to investigate the impact of polygraph testing on the engagement or avoidance of risk behaviours. The results of the first study showed that there was no significant difference in either the number or seriousness of behaviours reported by the two conditions. This would appear to indicate that ‘knowing’ about the polygraph had no impact upon the behaviour of participants. This finding does not support the initial hypothesis. When considering this result, however, it is important to bear in mind that the research was being conducted in the United Kingdom, where post-conviction polygraph testing is not used. The majority of probationers had, therefore, never experienced a polygraph test. This unfamiliarity with the procedure may have been a contributing factor in there being no difference between the conditions.



The results at the time of the second polygraph provide some support for this conclusion. At this time, the individuals reported significantly less behaviours when compared to their disclosure at the initial polygraph. In addition the reported behaviours were less serious, although the questionnaire responses at the final polygraph indicated that only half of the participants felt that knowledge of the impending polygraph test assisted them with avoiding risk behaviour. The less serious behaviour reported at the final polygraph, therefore, may also be a reflection of the progress that the men were making in their treatment groups.

An alternative explanation for the observed behaviour change is that participants may simply have reported less serious risk behaviour when they failed the polygraph on the second occasion. Since no follow-up polygraphs could be conducted it was impossible to confirm whether the participants had disclosed the full extent of their risk behaviour. As discussed earlier, Ahlmeyer et al. (2000) reported that offence related admissions (i.e. previous victims, acts and risk behaviour) increased with subsequent exams, it could be assumed, therefore, that the scope of disclosed risk behaviour underestimates the actual activity. The participants were, however, more likely to pass the second polygraph (72%) than the first (23%), this suggests that the majority of men had made complete disclosures at the final polygraph. This indicates, therefore, that the reduction in both the number and the seriousness of behaviours reported between the two occasions reflects a genuine change. Considering the high failure rate at the first polygraph it could indeed be speculated that the men would have been more likely to attempt to minimise their disclosures at this time because they were asked to explain the deceptive result. It is also not clear why an offender who is still



engaging in risk behaviours would run the risk of being found out, particularly when considering the voluntary nature of the research.

An important finding was that the majority of participants (97%) reported actively engaging in risk behaviours. At time 1, a comparatively large portion of men disclosed active victim acquisition behaviours (26%). Noteworthy examples included one offender that was actually re-offending (frottage), another who was cruising public toilets in search of children and a third who was having unsupervised contact with his previous 11-year-old victim. Knowledge of these behaviours provided professionals with the opportunity to intervene therapeutically, and in some cases actually altered the way in which supervision was carried out—one man, for example, was recalled to a probation hostel. These findings appear consistent, overall, with other research suggesting that sex offenders engage in a large amount of deception and un-known offending behaviours (e.g Abel & Rouleau, 1990; Barbaree, 1991). Considering that only 21 men out of 116 approached completed the study, these results prompt the somewhat unsettling concern – what behaviours were the non-volunteers and dropouts engaging in during the same time?

It is unclear whether this information would have emerged without the use of the polygraph, although it would seem unlikely. Only one participant had disclosed his behaviour to his supervisor at time 1. While almost half the men disclosed some risk behaviour in the interview to the researcher prior to the polygraph, nearly all the men (94%) made ‘new’ admissions during the polygraph process. This finding is similar to other studies demonstrating that the polygraph

improves the quality and quantity of information available to supervisors and treatment providers. For instance, studies have found that sex offenders report larger numbers of previously unknown offences and victims, younger ages at which they began to engage in deviant sexual behaviour, and a reduction in claims of their own sexual victimisation (Ahlmeyer *et al*, 2000; Emerick & Dutton, 1993; Hindman & Peters, 2001).

Comparison of the disclosure times on the two occasions suggests that periodic polygraph testing may also prompt offenders to disclose behaviours earlier in the interview process. At the time of the first polygraph, only 37% disclosed *all* their risk behaviour during the initial or pre-test interviews, compared with 80% that did this at the second. At this latter time the men were more likely to have disclosed risk behaviours to their treatment providers (60% compared with 3%). This finding would appear consistent with Harrison and Kirkpatrick's (2000) conclusion that periodic polygraph testing affected a sex offender's motivation in a positive direction. The survey results in the second study likewise indicate that the polygraph assisted some offenders with being more open and honest. Here a large portion also reported that the polygraph was helpful both with preventing re-offending and also with avoiding risk behaviours. Almost a third claimed that they were less likely after the polygraph test to have deliberate contact with potential victims. A quarter of offenders also felt that the threat of the polygraph assisted with reducing drug and alcohol use. These findings provide complementary evidence for the claims made by advocates that the polygraph acts as a type of 'artificial conscience' (e.g. Blasingham, 1998; Wilcox, 2001).



A degree of caution should, nonetheless, be exercised before drawing such a conclusion from the initial study. It is conceivable that offenders simply felt more comfortable with disclosing various behaviours due to the progress they had made in their therapy. When considering the self-report data, it must be kept in mind that these findings are based on the reports of only the 55% of programme participants who agreed to take part in the study, and it is possible that the other 45% may have had very different views on the value of polygraphy.

In a different vein, the high-rate of disclosure provides some insight into why many polygraph examiners believe the instrument to be infallible in spite of various studies showing significant error rates. In real-life contexts, post-test confessions are likely to constitute the only major source of feedback that the examiners receive about their tests. As noted earlier, ‘confessions’ are highly dependent on the failed outcome of a test. Nearly all the feedback that examiners receive would thus tend to be confirmatory.

### *Risk & Personality*

The Static-99 did not assist with predicting number of behaviours reported. It could be expected that high-risk offenders would report a greater number of behaviours. While such a finding may appear to cast doubt on the value of a well-validated risk assessment tool, it is important to remember that the instrument was actually developed to predict re-offences, and not the engagement in risk behaviours. In addition, the low amount of ‘high-risk’ offenders in the sample is also likely to have been a contributing factor in this non-significant result.

Antisocial personality characteristics were, unsurprisingly, associated with *reporting* a greater number of risk behaviours. This finding appears consistent with other research demonstrating that an antisocial orientation and lifestyle instability (crime-prone personality) are important factors in both sexual and general recidivism (Bonta, 1996; Gendreau, Little & Groggin, 1996; Hanson & Bussière, 1998).

Personality disorder, more generally, and normal personality traits (neuroticism, extraversion and agreeableness) as measured by the five-factor model (FFM) were not associated with reporting a greater number of behaviours. Whilst this finding may appear to be inconsistent with other research suggesting that such factors are related to recidivism (Hanson & Harrison, 1998), it is important to recognise that such findings have been established through examining actual recidivists. It is worthwhile briefly discussing the findings regarding the relationship between personality disorder and normal personality traits as measured by the NEO-PI-R as this is the first attempt to examine this relationship in a sample of sex offenders.

Personality disorder in this sample is characterised by high neuroticism and low agreeableness. This finding is consistent other studies (Saulsman & Page, 2004), and further supports the view that these domains may represent the core of personality pathology. With consideration of the facet scores the results appear broadly consistent with common theoretical conceptualizations of PD as longstanding patterns of emotional maladjustment (high vulnerability), poor



impulse control (high impulsivity) and poor interpersonal functioning (high anger & hostility, low trust, straightforwardness and compliance) (e.g. Millon, 1996). These results support findings from elsewhere (e.g. Miller, Reynolds & Pilkonis, 2004), and suggest that the FFM domains can be used to differentiate individuals with PD from those without. Due to the similarity of this profile, however, the FFM showed little ability to differentiate among specific PD categories. Again, this finding is consistent with results from other studies (e.g. Morey, et al., 2002), and implies that the FFM may not be a complete substitute for current PD diagnostic methods. The implications of these findings are discussed further in Appendix G.

#### *Reasons for additional admission*

It is interesting to speculate on the reasons why sex offenders might be more likely to disclosure information during a polygraph examination than in other contexts. As highlighted in chapter 4, Gudjonsson (2003) suggested that three factors are important for why suspects confess to crimes they have committed. These include: the perception of the strength of the evidence against them (i.e. the greater the evidence against an individual the more likely he is to confess), external pressure (e.g. fear of custody, doing a 'deal'), and internal pressure (e.g. wanting to 'clear their conscience', guilt). Usually more than one of these factors are present when suspects confess, although the empirical evidence indicates that it is the *perception of evidence against them* that is the most powerful reason for suspects confessing to an offence (Gudjonsson, 2001; Gudjonsson & Sigurdsson, 1999).

With consideration of post-conviction polygraphy therefore it seems likely that the polygraph may enhance the perception in offenders that the ‘evidence is stacked against him’, and thus prompt him to disclose information. A limited number of studies provide support for this hypothesis. Jones and Sigall (1971), for example, found that if an individual believes an instrument can detect deception (even if it cannot) then he or she will be more likely to disclose undesirable attitudes, presumable to avoid being ‘caught out’. Post-conviction polygraph testing may have a similar placebo effect on offenders with regard to disclosing and avoiding risk behaviours.

### *False confessions*

It is possible that because the majority of risk behaviours could not be corroborated, some men may have disclosed bogus behaviour (i.e. risk behaviours that they had not engaged in) believing that it might benefit them in some way. Published research, unfortunately, does not shed much light on this issue. As noted earlier Koskish and Blasingham (2002) found that only 5% of their sample claimed to have made erroneous disclosures during a polygraph examination. Nine per cent of the offenders in this study claimed to have done so, suggesting that while the incidence of this is not high, it is of relevance. High neuroticism and low conscientiousness scores characterised those who reported making false admissions; ‘high neuroticism’ is associated with pervasive feelings of guilt, fear, disgust, anger and embarrassment as well as high impulsivity, and ‘low conscientiousness’ with being less scrupulous and reliable. Neuroticism was also highly associated with personality disorder (these findings are discussed in



appendix G). Taken together this finding suggests that offenders who disclose false information in polygraph tests are more emotionally disturbed and impulsive when compared to other offenders.

This finding provides complementary evidence to other studies suggesting that false confessors in general criminal justice contexts are more likely to display abnormal personality characteristics (Gudjonsson, et al., 2004; Sigurdsson & Gudjonsson, 2001). It appears that these individuals in difficult interview situations may cope by ‘confessing’ or disclosing information. Thus ending the interview situation and removing the source of stress. Examining the reasons the 12 participants gave for making false disclosures provides some support for this, with 6 of the 12 citing either a fear of getting into trouble with their probation officers or feeling pressured by the polygraph examiner. Interestingly, none of the 12 said they had received a sanction because of a polygraph test, indicating perhaps that the ‘false’ disclosures were either not particularly significant, or were viewed with scepticism. Regardless, it appears that some individuals may be more prone to making false disclosures, and polygraph examiners as well as those providing treatment and supervision need to be cautious when interpreting polygraph examination disclosures.

With consideration of the initial study all the men were aware that they could withdraw at any stage without providing a reason and that the specific result of their polygraph test would not be shared with others. Other than appearing co-operative, there would seem, therefore, to be little to be gained from making a false admission. Because risk behaviours would be reported to the relevant

professionals, it could be argued that there would a greater reason to not disclose information. The high drop out rate in this study between time 1 and 2 perhaps provides some support for this conclusion, and further suggests that the volunteers were not attempting to please examiners or to gain rewards for disclosure. Furthermore, because polygraph tests are not used in the treatment programme the offender's treatment provider had little invested in the polygraph outcome.

The literature, as highlighted earlier, also suggests that false confessions in criminal justice contexts are reasonably rare. Sigurdsson and Gudjonsson (1996), in a sample of inmates, found that 12% claimed to have made false confessions to the police at sometime in their lives. In another study with university students only 4% who had previously been interviewed by police reported making a false confession (Gudjonsson, et al., 2004). Taken together it would seem that the likelihood of false admissions in the initial study was small.

### *Limitations of study 1*

It is important to recognise that the results from this study only suggest that the polygraph may assist with gaining information about high-risk behaviour, in addition to acting as a deterrent for engaging in such behaviours. The findings are not evidence that the polygraph reduces recidivism. Additional research needs to be conducted that examines this issue.

A possible weakness of the initial study was that the number of times the offender engaged in each of the specific behaviours was not recorded. Not



gathering this information might have masked some of the therapeutic benefit that the men may have gained. It is possible that the polygraph contributed to a reduction in the *incidence* of risk behaviour rather than prevented it completely. For example, by reducing the number of times an offender masturbated to deviant fantasy or by reducing the amount of alcohol he consumed over a week. Such a behavioural change could represent important progress, and could provide further support for the therapeutic value of the polygraph in this context. The difficulty with gathering such data is, of course, the reliance on an offender's recollection of sometimes relatively innocuous behaviours over a long period of time.

The attrition rate in this study is particularly noteworthy, and represents a serious limitation to this study. Only 42% of the participants who volunteered completed the research. This represents 18% of the total sample approached. While two participants were returned to custody, one can only speculate on the reasons for the other 'drop-outs'. It was observed that there was considerable variability in the follow-up to the polygraph results by the various treatment providers. While a number of participants stated that they had been reprimanded and that their disclosures had been actively incorporated into their treatment, others reported that they had not discussed the results with anyone, despite, in some cases, having disclosed serious risk behaviour. The inconsistency in the application of consequences, or indeed acknowledgement, of the disclosure of serious risk behaviour is likely to have affected the impact of the polygraph and possibly also the participant's motivation to complete the research. One man did report, for example, that he was unprepared to continue because of the consequences that he had experienced due to his admissions at the initial polygraph. Another person

reported that he had found the polygraph experience 'very stressful' and since stress was a risk factor, he thought it sensible to withdraw.

An alternative explanation for the high attrition rate is that the polygraph was too effective. The men that stayed away from the final polygraph may have been wishing to hide behaviours that they feared the polygraph could detect. This explanation illuminates an inherent 'double bind' often associated with conducting research in this area. Offenders are unlikely to volunteer and complete research that could potentially uncover behaviour that might lead to adverse consequences for themselves. An attempt was made to overcome this problem by monitoring only 'legal' risk behaviours. Disclosure of such behaviours is, nonetheless, still likely to have affected aspects of their supervision and treatment in manner that they might consider undesirable. The alternative of providing offenders with absolute confidentiality is both problematic and ethically questionable. Actually having disclosed the risk behaviours in the initial study appears to have prevented at least two sexual offences.

In a related vein, the methodology used highlights an important ethical issue in relation to the volunteers. Is it fair to have used the offenders as was done here? It is unfortunately very difficult to adequately examine these issues in a controlled experiment without resorting to some form of deception. The need to conduct research in this area gains a particular significance in light of the emerging enthusiasm for the introduction of post-conviction polygraphy. In this regard it is perhaps wise to bear in mind that historically the polygraph has been rapidly introduced into various contexts as a type of 'magic lasso' with little apparent

forethought. Such fervour has led to widespread abuses (e.g. Coor's brewery employment vetting), poor regulation and generally inadequate research agendas within the scientific community. The same could happen to post-conviction polygraphy, so to avoid repeating the mistakes of the past, the polygraph should be regarded with a degree of scepticism and a more cautious approach adopted to its implementation. Further research needs to be conducted that not only replicate the present research but also explores how best to use the instrument. In addition, standards of practice and regulatory mechanisms need to be established. These latter issues will be discussed in greater detail further on in this chapter.

Finally, it is important to briefly highlight that the power was low. This is partly due to the low subject numbers in the initial study. Unfortunately, as highlighted above, it is difficult to get subjects to volunteer for this type of research.



## The issue of accuracy

While the emphasis on utility in post conviction settings is understandable, polygraph accuracy cannot be ignored. If subjects do not believe that polygraphy ‘works’ they will be less likely to disclose relevant information during a test, a knowledge of accuracy rates is also required to make sense of test results in cases where there is an indication of deception in the absence of disclosure. Thus, those tested as well as those who rely on test results must have confidence in the validity of the technique if it is to be clinically viable.

In terms of polygraph accuracy in general, the literature contains conflicting accounts, with many studies criticised for their methodological weaknesses (Furedy, 1996; Lykken, 1998; Cross & Saxe, 2001). A major reason why scientific debate over polygraph validity yields conflicting conclusions is that the validity of such a complex procedure is very difficult to assess and may vary widely from one application to another (i.e. pre-employment screening versus criminal investigations versus post-conviction sex offender testing). This study is the first experimental attempt to investigate the accuracy of the polygraph in a post-conviction context with community based adult sex offenders. With this in mind, it is useful to begin by reviewing the ways in which the utilised methodology overcomes many of the problems of other studies related to criterion development and external validity.

Firstly, the testing situation was authentic for both the examiner and examinee. Secondly, the result of the polygraph was compared with an objective

and independent criterion, a drug test. Only the researcher was aware of the drug test outcome. Thirdly, polygraph outcome (pass or fail) was determined by blind examiners only evaluating the charts, with no access to other information, such as the examinee's background, disclosures or behaviour during the test.

In spite of these conditions, it could perhaps be argued that because an additional 'research question' was included in the participant's regular polygraph test it was not a real situation for either the examiners or examinees. This is unlikely to be the case. All the volunteers could have experienced sanctions if they reported illicit drug use. The 'drug question' therefore represented an authentic issue with potentially serious consequences for all the examinees. Similarly the polygraph examiners evaluations have significance for themselves, as their accuracy was being directly tested. It could be expected therefore that they would have been motivated to do their best. Of course this could have had the effect of making examiner's judgements more cautious than in the typical examination where their performance cannot be readily evaluated.

### *Blind examiners*

The blind examiners scored almost a third of the charts as inconclusive (27%). When considering the remaining charts, the data showed that the polygraph is highly accurate in identifying both deception and truth telling (truth tellers, 79% & liars, 100%; ROC = .88). This finding is similar to accuracy rates reported elsewhere (e.g. Raskin & Honts, 2001; Saxe, Dougherty & Cross, 1983),

and further supports the view that the polygraph is highly accurate in detecting deceit, though less accurate with detecting truth telling.

In this study individuals were more likely to be classified as a false positive than a false negative (21% versus 0%). This is not the first study to report a large error rate for innocent subjects: Kleinmuntz and Szucko (1984), Patrick and Iacono (1991) and Forman and McCauley (1986) reported 19%, 24% and 47% respectively. However, in each of these studies methodological weaknesses have been identified that could account for their error rates. The results from the present study suggest that when these methodological problems are eliminated the observed hit rates may still contain significant error rates.

The false positive problem in Comparison Questioning Test (CQT) formats may be due to a number of factors. Firstly, the scoring procedure utilised in the CQT may predispose the testing format to false positive outcomes. To briefly reiterate, to be considered to have passed a polygraph test, an examinee must score at least +3 on *every* relevant question. If an examinee has a score of -3 on *any* of the relevant questions it is considered to be a 'failed' (deception indicated) test. If an examinee does not achieve +3 or higher on every R question, and does not score a -3 on any of the R questions, then the test is considered inconclusive (Matte, 1996). It seems easier for an examinee to achieve a failed outcome, because they only have to react strongly to one of the relevant questions, in contrast, to having to strongly react to every comparison question to pass. This inherent bias is likely to inflate positive rates.



In a related vein, the false positive issue also may reflect the underlying problems with the theory of the Comparison Question Test (CQT). Briefly, the CQT compares physiological responses to relevant questions that are considered relevant (R) to the investigation at hand, which evoke a lie from someone who is being deceptive, with responses to comparison (C) questions to which the person responds in a presumably known way (i.e. is deceptive). The responses are compared only for one individual because it is recognised that there are individual differences in physiological reactivity. According to this theory, a guilty person lies to both the relevant and comparison questions, while the innocent person lies only to the comparison questions and not the relevant questions. The theory is that the innocent person will show equal or less physiological reactivity to relevant than comparison questions ( $R < C$ ), whilst the opposite is expected to occur for the guilty examinees ( $R > C$ ). There is, however, no unique physiological lie response, thus, an individual's stronger response to the relevant questions may not necessarily be indicative of deception. Some innocent individuals, for instance, may perceive the relevant questions as more threatening than the comparison questions for a variety of reasons.

#### *False positives & false negatives (blind examiners)*

Examining the characteristics of the false positives showed that they were more likely to have experienced a sanction due to a polygraph test. The sanctions reported by examinees, including curfews, increased supervision and having to address additional issues in treatment. Only 4 out of 14 (28%) individuals who had experienced a sanction were actually a false positive. This would initially

appear to suggest that this 'threat-effect' was reasonably minor, however, examining the accuracy of the outcomes for all these subjects revealed that only 36% were correctly classified. This would seem to suggest that for some individuals, having been punished due to a polygraph result might predispose them to react to the relevant issues at a later test. Because the polygraph only measures physiological arousal it is possible that these examinees may have become conditioned to respond to issues they had previously failed. Some examinees may also have been fearful that an additional 'failed' polygraph outcome may result in a more severe punishment, and thus further priming them to physiologically respond to the relevant issues.

This finding appears consistent with results from Patrick and Iacono (1991) and Forman and McCauley's (1986) studies, and suggests that in situations where there is genuine concern about the outcome of a polygraph test, a substantial proportion of innocent examinees may react more strongly to the relevant issues, producing deceptive test results. Clearly this finding has implications for how treatment providers should respond to a different polygraph test outcomes.

Numerous proponents have argued for applying rewards and sanctions (Cooley-Towell, et al., 2000). The sex offender treatment program at the Colorado Department of Corrections has developed a 'sanctions grid' (Heil & Simmons, 2003). This grid links sanctions to when the new information was disclosed. The severity of the consequences is related to when the information is disclosed and what type of information is revealed. An offender who discloses information during group therapy would thus receive less of a sanction than one who fails a



polygraph test, and continues to deny the issues under investigation. Both Abrams (1991) and Matte (1996) contend that situations where offenders experience no consequences for negative test outcomes would eventually undermine both the utility and accuracy of the procedure. This line of argument endorses the 'infallibility' view of the polygraph, and consequently, fails to consider (or recognise) the impact of unjustified sanctions. The results of this, and other, studies indicate that sanctions may predispose truthful individuals to physiologically react to the relevant issues. It would be difficult for professionals using polygraphy to have no response one way or another to an offender's test result. It could be expected that if an offender has nothing to lose by lying or not disclosing information, it is likely that this would reduce the utility of the polygraph.

In light of the findings from this study a failed polygraph test is perhaps best viewed as a type of 'red flag'. For example, a deception –indicated result would prompt professionals to further investigate further by other means, such as, additional interviews with the offender's family and associates, home visits or surveillance. Indeed it may be more suitable to view the polygraph as an additional assessment and treatment tool; in this regard professionals would consider polygraphy in manner as they would regard other commonly used tools such as psychometric testing and phallometric assessment. In other words, taking into account the short-comings and limitations of each of these tools.

In this study, intelligence was found to differentiate false positives from true negatives. This result may suggest that cognitive or perhaps attention deficits



have an important affect on the accuracy of polygraph outcomes. This result might support concerns raised by others (e.g. Cross & Saxe, 2001; Lykken, 1998) regarding the theoretical reasoning of the CQT. That is, examiners are instructed to create emotional conditions designed to lead to differential levels of arousal and physiological responsiveness in innocent and guilty examinees. It may be that this process is made more difficult with individuals who have lower cognitive abilities. This conclusion should, however, be regarded with a degree of caution as the difference in IQ observed in this study was not great (90 versus 100).

Normal personality characteristic as measured by the NEO-PI-R did not affect the accuracy of outcomes in this study. The strong relationship between the NEO domains (neuroticism and agreeableness) and personality disorder (PD), also provided the opportunity to also examine the affect that PD may have on the accuracy, again there was no significant result. Overall, these findings are consistent with other research that have found personality does not affect polygraph outcome (e.g. Hammond, 1980; Honts, Raskin & Kircher, 1985; Patrick & Iacono, 1986, 1989; Raskin, Barland & Podlesny, 1976; Raskin & Hare, 1978).

### *Base rates*

Examining the accuracy of the different test outcomes revealed that 'passed' (NDI) outcomes were highly accurate (100%), whilst failed (DI) outcomes were much less reliable (27%). Together these findings provide some support for the concerns raised by critics, such as Lykken (1998), that the polygraph may be biased against innocent individuals. This result however needs

to be regarded with caution, as it may be an artefact of the low base rate of deception found in the sample (i.e. 7%). In circumstances where base rates of deception are low, even a highly accurate test will produce more false than true positives (Murphy, 1987). For example, if only 10% out of a sample of 1000 individuals are deceptive, a test that is 90% accurate will correctly identify 90 of the 100 'liars', but it will also falsely label as 'liars' 90 of the 900 truth tellers, resulting in 50% of failed outcomes being incorrect. This phenomenon is illustrated with varying base rates and accuracy levels in table 28. Thus, in the present study if the deception rate had been 50% and everything else being equal, the positive predictive rates would have been 79% (equal to the specificity).

Table 28.

*The impact of base rates on the Positive & Negative Predictive Values at 90% & 95% accuracy*

Sensitivity & Specificity 90%		
Deception Base Rates	Positive Predictive Value (Correct Deception Indicated)	Negative predictive Value (Correct No-deception indicated)
10%	50%	99%
50%	90%	90%
90%	99%	50%
Sensitivity & Specificity 95%		
Deception Base Rates	Positive Predictive Value (Correct Deception Indicated)	Negative predictive Value (Correct No-deception indicated)
10%	68%	99%
50%	95%	95%
90%	99%	68%

The importance of base rates was also highlighted by the National Academies of Science review (National Research Council, 2002), and is one of the primary reasons why the review did not support the use of the polygraph in security contexts. They argued that in a security context the base rate of deception is likely to be extremely low (the assumption being that there are very few spies in federal agencies), therefore the vast majority of ‘failed’ polygraph results would be false positives. The review suggested that polygraphy only becomes viable when the base rate of deception exceeds 10%. Those arguing for its use in the intelligence community, however, point out that while the negative predictive rate



may be low in this setting, polygraphy nonetheless serves as a useful screening tool to identify a small group of individuals who need to be looked at more closely.

Without going into the pros and cons of the various arguments regarding polygraphy in security settings, it should be recognised that the issues are not identical in post-conviction contexts as there are a number of important differences between the two applications. Firstly, the base rate of deception in sex offender samples is likely to exceed the 10% minimum recommended by the National Academies Review. Indeed, as highlighted in chapter 4, research has suggested that up to 90% of offenders are dishonest about their offending behaviour and offence history during treatment (e.g. Barbaree, 1991; Maletzky, 1991). Of course this does give rise to a somewhat paradoxical situation that the more honest the examinees are the less confident one can be with a 'deception indicated' result. Secondly, in post conviction contexts the emphasis is less on 'passing' or 'failing' the polygraph, and more on gaining information relevant to the offender's supervision and treatment. Finally, the consequences associated with inaccurate outcomes in pre-conviction contexts are different to those in post-conviction settings. In a pre-conviction context, for example, innocent examinees could have their career terminated or interrupted, whilst in a post-conviction context consequences are comparatively minor (e.g. addressing additional issues in treatment, additional supervision sessions, etc).

It is valuable to consider the accuracy of the original examiners, as these findings may provide a 'real-world' picture of the polygraph's accuracy. Notably, there was poor reliability between these examiners and the blind scorers, suggesting that these original examiners were using additional information when making their determinations. The issue of reliability will be discussed further on in this chapter.

The original examiners were much less likely to evaluate examinees as inconclusive when compared to the blind examiners (6% versus 27%). They were however much less accurate with the sensitivity being only 40% and the specificity 84%. When considering the accuracy of the different test outcomes, similar to the blind examiners, failed (15%) polygraph outcomes were much less likely to be accurate than passed (95%) polygraph outcomes. Again these results need to be considered in light of the low base rate for deception.

The only feature that distinguished the false positives from the examinees accurately identified was the personality trait of Extraversion. High scorers on this trait tend to be sociable and talkative, whilst low scorers are inclined to be introverted and reserved (Costa & McCrae, 1992). In this study, the false positives were much less extraverted than the true negatives. This finding perhaps indicates how the original examiners may have made their determinations. The implication is that the original examiners made their determinations based on their evaluation of the examinees interpersonal behaviour during the testing process. For example,

examinees that were less talkative or forthcoming during the pre-test or post-test interviews might have been regarded with suspicion by examiner and subsequently failed.

It is notable that the relationship between intelligence and 'having been sanctioned for a polygraph outcome' did not emerge when examining the results from the original examiners. It may be that these examiners were able to recognise when examinees were overly 'stressed' and cognitively struggling with differentiating the relevant and comparison issues.

### *Self-report*

Some complementary evidence for the accuracy of the polygraph can be seen in the responses to the self-report measures. In terms of the validity of the polygraph, it must be emphasised that my findings are based on the uncorroborated self-reports of participants, and it was not possible to compare their recollections with actual test outcomes. Nonetheless, the self-report of the offenders suggesting accuracy of 85% in detecting truth-telling and 84% in detecting deception is consistent with the blindly re-scored accuracy rates. These results are also consistent with the findings from Kokish and Blasingham's (2002) study. The offenders themselves also perceived the accuracy of the polygraph to fall within this range, with the majority rating it as 'moderately' to 'highly' accurate.

Again examining the accuracy of the different test outcomes showed that whilst passed outcomes were highly accurate (97%), failed outcomes were only



correct half the time (48%). This finding indicates that offenders are more likely to erroneously fail a polygraph test (i.e. be a false positive) than to beat it (i.e. false negative). This outcome may of course be due to self-presentation biases (i.e. deceptive offenders may have been more likely to claim that the polygraph got it wrong when caught out however be less likely to disclose having 'beaten it').

## Reliability

For the 70 tests that were blindly re-scored numerical scores for the three blind examiners showed substantial agreement with each other. Comparison of the classifications (DI, NDI and IC) between the original and the blind examiners, however, revealed poor inter-rater reliability. This result supports the concerns raised elsewhere regarding the issue of standardisation (Lykken, 1998; National Research Council, 2003). The reliability between the blind examiners was high, which suggests that when examiners adhered to scoring procedures they are quite reliable. It is perhaps noteworthy that one of the original examiners had been trained in the John Reid School that endorses the use of behavioural characteristics when 'making a diagnosis'. As highlighted in chapter 2, little research has supported the accuracy of non-verbal behaviours as indicative of deception. It is unclear, however, whether these examiners used behavioural cues or other additional information when 'scoring' the polygraph charts. It is also conceivable that the expectations of the original examiner regarding drug use affected their decision-making.

Only five participants (7%) returned positive drug tests. Whilst the original examiners would not have known the actual base rate for the sample, it is possible that they would have been aware that drug-using offenders would have been unlikely to volunteer for the research and thus be inclined to 'pass' the volunteers. Comparison of the pass rates between the re-scorers and original examiners does indicate that this factor may have contributed to some of the original examiner's decisions (77% vs. 53%). In addition, it is likely that the original

examiners would have been aware of the examinee's drug use history, and may also have used this knowledge when evaluating the charts.

### *Limitations of study 2*

A notable limitation of this study was that because the polygraph was used to detect drug use, it might not generalise to actual sexual offending behaviours. It is possible that an offender would experience a greater level of stress, when they are being examined about actual sexual offending when compared to drug use.

Several potential examiner variables that could not be controlled may also have affected the results in this study. Because the polygraph tests were conducted in real life circumstances it was not possible to ensure that examiners adhered to standardised procedures. Thus, it is likely that the examiners' behaviour (e.g. time spent discussing drug issue, presentation of questions) varied in important ways that may have affected the results. In a similar vein the expectations of the original examiner may have introduced an inherent bias by affected the actual physiological reactions. For example, an examiner that believes that the examinee is unlikely to be taking drugs because they have no drug abuse history or convictions may be motivated (either unconsciously or consciously) to produce a non-deceptive result. In such a situation the examiner might over emphasise the comparison questions and spend very little time discussing and reviewing the relevant drug use question. This would conceivably decrease the probability that that the examinee will react to the relevant drug use question. Such a bias, because of the very low base rate, would inflate the accuracy of the results in this study.



The alternative scenario may also occur where the examiner suspects that the examinee is taking (or has taken) drugs and ‘produces’ a failed outcome. In such a situation the examiner might ask an over-general or provocative drug use question, and spend a great deal of time in reviewing and presenting this question. Clearly, future research should ensure that the polygraph exams are standardised, or have a method of monitoring the tests. For example, by video taping the tests and then randomly selecting a number to evaluate would provide a method of ensuring reliability.

Another limitation is the use of hair drug tests. This test cannot evaluate drug use over the previous seven days, thus it is possible that some offenders may have used drugs in this period and not been identified. For example, some offenders who had used drugs over the last seven days may have failed the polygraph test, however, returned a ‘clean’ hair sample. Similarly it is possible that error on hair tests may have undermined the accuracy of the polygraph.

Only 55% of the individuals approached agreed to take part in the study. All of the offenders approached for the study were required to complete, and pay for, a minimum of two polygraph tests per year (\$225 per test). To encourage participation I arranged to use the participant’s regular exam in the research and also offered a small discount (\$25). Two of the three treatment facilities also allowed volunteers to complete the questionnaires and interview whilst attending their weekly session. It is therefore surprising that despite these incentives a large portion of offenders declined to participate. Clearly it is possible that some of the

offenders were suspicious of my motives and/or did not believe my assurance that the drug test result would remain confidential. Others may simply have had no interest in participating or found the collection of a hair sample to be too intrusive.

Of course some individuals may have been fearful that the polygraph would detect illicit drug-use. In the present methodology any disclosures regarding illicit drug use as well as the outcome of the actual polygraph test was shared with the subject's therapist and probation officer. Thus if a subject failed the drug question or the polygraph examiner became suspicious that the offender had used drugs they could report this to the offender's probation officer. The risk for drug using offenders therefore is significant. A small number of offenders did report to the researcher in confidence that they were fearful that the polygraph would detect their regular illicit drug use. Again this highlights the significant obstacle faced by researchers attempting to evaluate the polygraph's accuracy. That is, it is difficult manufacturing voluntary situations where the consequences are real and meaningful for volunteers.

This thesis is among the first empirical attempts to investigate the polygraph's use in the treatment and supervision of community based sex offenders. The first study attempted to examine its utility in this context, whilst the second investigated its accuracy. Overall, whilst the findings were promising, the studies have suffered from low power that has limited the value of the findings, and highlights the need for replication with larger numbers. In addition this research has brought to light a number of important issues of both theoretical and clinical interest suggesting some key lines of enquiry.

Firstly, systematic research on the bogus pipeline phenomenon can help with deterring and detecting sexual offences in more than one way. It can clarify the extent to which the practical value of the polygraph for eliciting admissions results from test accuracy, or merely from the offenders' beliefs and concerns. Such research would also help the professionals to better interpret the information they get from using the polygraph. It may also help improve interrogation and interviewing techniques. A possible research methodology may be could be to compare the information gained between an actual polygraph examination and a bogus examination with neither the examiner nor examinee aware that the test output is bogus. Another methodology could be to have offenders randomly allocated to polygraph versus non-polygraph groups, however, only after the examiner and examinee have completed the initial pre-test interview.



Secondly, in a related vein, it would be valuable to compare the recidivism and programme violation rates for polygraph versus non-polygraph offenders in identical treatment programmes. No research to date has actually shown that polygraphy reduces sexual recidivism rates, or that it improves the supervision of such individuals.

Thirdly, it would be valuable to explore the impact of periodic polygraph testing has upon offenders. For example, does it lower their motivation for engaging in treatment and adhering to their risk management plan? Does it reduce their capacity to engage effectively in treatment? Would such an outcome and possible sanctions increase an offenders psychological distress, and thus possibly, their risk of a re-offence?

Fourthly, in parallel to the issues highlighted above it would be valuable to examine the issue of sex offender risk assessment using polygraphy. The polygraph studies to date suggest that offenders have a much larger and diverse offending history. It would be useful to examine how this information may assist with improving current risk assessment measures, as many of these tools (e.g. STATIC 99) have been developed using only recorded (known) offending history. Such research could potentially have significant implications for both the assessment and treatment of sexual offenders, and perhaps also may contribute to the theoretical understanding of sex abuse.

One method of investigating the areas mentioned above, would be to ensure that services that use polygraphy maintain documentation of the tests

administered. Amongst other things such records could include: information on exactly which question or questions produced responses indicating deception; precisely when in polygraph examination admissions were made (i.e. before, during or after); and, whether admissions were made in response to examiner's claim of deception supported by polygraph chart, or to other factors. This information, combined with individual background information and reports on subsequent outcomes, would provide a valuable body of data that could provide clarification on such issues.

Finally, there must be empirical evidence of the polygraph's accuracy in a post-conviction context. The broader the range of examinees, examiners, and social contexts in which accuracy is demonstrated the greater the confidence that a technique will perform well. As highlighted throughout this thesis the research evidence for the accuracy of polygraph testing is problematic and far from being desirable. Future research could perhaps use a similar methodology as used in this research however with actual drug offenders. In this scenario the issue of drug use would be a much greater issue for participants, than it perhaps was with the subjects in this research. It is clear that more research is also needed to examine the various issues related to validity. Some research questions worthy of investigation are offered below:

- Are the mechanisms purported to link deception to physiological states, universal for all people who might be examined, or do they differ for individuals in different situations or for different tests?

- How does the social context and social interactions that constitute the testing procedure affect the reliability and validity of the outcome?
- What mechanisms might a truthful response produce a false positive result?  
What do examiners do to counteract or correct for such mechanisms? Is this response to the possibility of false positives reasonable considering the mechanisms involve?
- By what means could a deceptive response produce a false negative response? That is, what is the potential for effective countermeasures?  
What do examiners do to counteract or correct for such behaviours? Is this response to the possibility of countermeasures reasonable considering the mechanisms involved?
- Do specific types of countermeasures (e.g. visualisation; inducing pain or discomfort) have specific physiological profiles? If so are these different from deceptive profiles?

Future studies should ensure that they incorporate a number of features, such as randomisation, adequate sample size, blind administration and evaluation of the charts, and standardisation of procedures, to ensure the quality of research.



## Conclusions & Implications

In summary, the findings in relation to utility support the view that the polygraph is likely to be a useful adjunct to both the treatment and supervision of sex offenders. The amount of information gained through the use of the polygraph highlights its potential value to professionals, although the high attrition rate and small sample size make the applicability of the present findings unknown. Replication of this research is, therefore required before the polygraph can confidently be incorporated into post-conviction contexts. As research accumulates in this regard, it will be possible to further examine how best to utilise the polygraph, and also to examine the effect that personality, risk and other variables may have on utility. With regards to accuracy, the second study demonstrated that the polygraph is reasonably accurate in detecting drug use, however, only when the charts were scored blindly. Nonetheless, examinees were more likely to be false positive, and failed polygraph outcomes were less reliable. In addition, imposing penalties for a specific test outcome may create a de facto involuntary condition that increases the chances of invalid or inconclusive test results on a later test. Thus applying sanctions for different test outcomes would mean that some offenders would be unjustly sanctioned for wrong outcomes. To reduce the probability of such a situation the polygraph could be viewed as a type of 'red flag' or warning mechanism. In such a situation a failed outcome should prompt further investigation by some other means, such as follow up interviews or surveillance. Finally, it is important to recognise that the results from these studies

have been troubled by low power, and thus highlights the importance of replication.

### *Post-conviction polygraphy in the United Kingdom?*

As mentioned earlier the polygraph has not been used in an official capacity in the UK. This situation is partly due to a review by the British Psychological Society in the mid eighties (see chapter 3). This review raised concerns about theory underlying polygraph, standardisation and monitoring of examiners, inducing anxiety in examinees, and countermeasures. Whilst these issues clearly are problematic and need to be addressed it is important to reiterate that the review in the 80's did not consider the polygraph's use in a post-conviction with sex offenders. For this reason it is debatable whether all the criticisms equally apply. As highlighted earlier pre-conviction polygraphy in criminal investigations is primarily focused upon determining a suspect's guilt or innocence in relation to a specific crime. In contrast, the focus of post-conviction polygraphy is to gain information about an offender's current and previous behaviour. The main issues raised in the BPS report are discussed in turn below.

#### *The issue of theory*

The issue of theory remains an on-going concern. As highlighted in chapter 5 little research has developed and tested theories of the underlying factors that might produce the observed responses in a polygraph test. The majority of published research has mainly addressed so-called 'technical issues' related to



accuracy, such as the type and positioning of questions asked, and thus can be considered essentially atheoretical in nature. This narrow research scope appears to have largely been due to the context in which polygraphy has tended to be used, that is, in criminal investigations by police departments (National Academy of Sciences, 2002). Thus, it has mainly been law enforcement agencies, whose primary concern is with the practical detection of deception, which have funded and conducted research on polygraphy. These circumstances, therefore, appear to have determined the nature and direction of scientific inquiry and consequently have encumbered the development of a more robust theory base.

To briefly reiterate, the most commonly accepted theory at present is that, when the person being examined fears detection, that fear produces a measurable physiological reaction when the person responds deceptively. Thus, in this theory, the polygraph instrument is measuring the fear of detection rather than deception per se. The examiner infers deception when the physiological response to questions about the crime is greater than the response to other questions. Thus, the two critical theoretical issues are: 1) can the physiological phenomena, as measured by the polygraph, be used to identify and distinguish between specific internal psychological states, such as lying, and 2) is the assumption justified that reactions to comparison and relevant questions constitute separable constructs.

It is possible, of course, that if intensive research efforts had been made at an earlier stage these issues would be resolved. Clearly future research needs to investigate the variety of variables and mechanisms that link deception or other phenomena to the physiological responses measured in the polygraph test. It is



difficult, however, to identify a methodology that could provide unequivocal findings, in the short-term. Theory development and testing, perhaps, needs to occur as it does in other areas of science, by repetition and through the use of prospective imperfect methodologies. Conceivably such a research agenda would eventually lead to innovation and improvement in both the methodologies used in research and also the polygraph. Indeed, it is of course also possible, that the polygraph as a 'lie detecting machine' would be abandoned.

### *Standardisation & monitoring*

Since the 1980's the American Polygraph Association (APA) has set national standards for sex offender testing and ATSA have also include a section on polygraph in its practice standards and guidelines (ATSA, 1997). The various documents are essentially consistent with one another and generally address the following issues:

1. Maximising accuracy and reliability by adhering to explicit standards for:

- a) examiner training and experience;
- b) instrumentation and instrument calibration;
- c) pre- and post-test interview procedures and formats;
- d) chart scoring and interpretation;
- e) examinee fitness; and,
- f) test format, environment, and frequency.

2. Appropriate and inappropriate application for making case decisions.

### 3. Protecting the treatment setting and examinee by:

- a) obtaining informed consents;
- b) maintaining accurate record keeping and secure record storage;
- c) ensuring confidentiality; and
- d) careful management of disclosures related to criminal acts committed before entering treatment.

Yet, in spite of these guidelines, the poor inter-rater reliability that was observed in this study raises the concern of whether field examiners actually adhere to 'best practice' guidelines. This concern has been raised by numerous opponents of polygraph (e.g. Cross & Saxe, 2001; Furedy, 1996a). A review of testing practice in post-conviction sex offender testing is beyond the scope of this thesis. Nonetheless it is important to emphasise two issues regarding reliable test administration and interpretation. Firstly, reliable test administration and interpretation are both desirable, and essential, if a post-conviction sex offender testing program is to have scientific standing. Secondly, it is critical to remember that reliability, no matter how well ensured, does not confer validity on a sex offender testing programme. Thus, if the polygraph is to be considered a valid psychological test, a higher level of consistency and standardisation is needed. It is, therefore, incumbent on examiners to provide sufficient standardisation, validation and reliability data that are empirically based and define more accurately the degree to which the polygraph results can be relied on in a decision-making process. Unfortunately, there are a number of factors related to the training of polygraph examiners that are likely to make this task a difficult one for

the polygraph profession. These factors are also likely to have contributed to the poor reliability found in this study.

As discussed in chapter 2, to become a polygraph examiner in the US, an individual requires no specific qualification and only completes an eight-week course. Current APA accredited programmes provide limited teaching of psychology and physiological, and no statistics or psychometric training. The wider implications of this lack of grounding in these areas means that many examiners will have little understanding of the importance of standardisation, and the other issues related to validity and reliability. Furthermore, the training does not address many of the concerns raised by the scientific community with regards to both the theory and accuracy. Thus, it is debatable whether the current polygraph training programmes would foster a scientific mindset in their students. In other words, current empirical literature and research is unlikely to be accessed by examiners to guide and develop practice. The problems of poor training are exacerbated by a general lack of standardisation of polygraph techniques. There is a plethora of different testing formats, scoring procedures and techniques, the majority of which have never undergone scientific scrutiny, available to the examiner. It is conceivable, therefore, that some polygraph examiners in their day-to-day practice will simply *drift* from best practice APA guidelines.

Many of these problems are structural in nature and therefore can be remedied. In the US there is little regulation of polygraph examiners, training or the techniques that can be used. The American Polygraph Association (APA) whilst having made efforts to develop standards of practice, in reality, have little



power to sanction poor practice. In the UK because the polygraph has not been used, it is possible to address many of the problems experienced in the US. A regulating body, for example, could be established to develop and conduct training and also monitor examiners. In this regard, it is interesting to consider the American Psychological Association's policy on polygraph testing, which states:

“Polygraph tests used in all applied settings should be based on adequate psychological and psychophysiological training and sophistication. Their use by psychologists must be consistent with the American Psychological Association's Standards for Educational and Psychological Testing and the *Ethical Principles of Psychologists*. They should be used only when such use is justified by the existence of sufficient data on their reliability and validity for the particular population, context and specific purpose”  
(American Psychological Association, 1986).

This policy provides some guidance on how training and standards of practice could be developed for polygraph examiners in the UK. The *Ethical Principles of Psychologists*, for example, states that psychologists who are responsible for decisions involving individuals based on test results need to have an understanding of psychological measurement, validation problems and test research (American Psychological Association, 1981).

### *Inducing stress*

The BPS report also raised concerns regarding the ‘anxiety’ induced and the inherent coercive nature of obtaining compliance with the polygraph assessment. It is difficult to see how such a criticism is valid when considering the polygraph’s use with sex offenders. In the UK the majority of community based sex offenders are required to engage in treatment, a process which is arguably much more stressful than a polygraph test and in the majority of cases not voluntary. Indeed these programmes utilise a variety of therapeutic techniques that are very stressful for the offenders, for example, having to write ‘victim’ letters (to themselves) and being challenged by other group members and the facilitators about their justifications (i.e. cognitive distortions). Some offenders are occasionally also asked to complete a penile phallometric assessment, an arguably much more intrusive procedure than the polygraph.

More generally, stress and anxiety are deliberately induced in a variety of psychological interventions, such as flooding or systematic desensitisation. The value of such interventions (when suitable) are generally not questioned, similarly, neither is the value of sex offender treatment programmes.

### *The concern of countermeasures*

Countermeasures pose a significant threat to the value of the polygraph in its use with sex offenders. Countermeasures are anything that an examinee might do in an effort to defeat or distort a polygraph test. Drugs are commonly brought

up in this context, but no drugs have been found to be an effective countermeasure against the CQT. It would appear that drugs are unlikely to be an effective measure against the comparison question test since the comparison question requires differential reactivity between two stimuli. Only 2 subjects claimed to have used countermeasures, both reported using drugs. Notably both also reported being a false negative. Of course this finding should be treated with some degree of caution, as examinees successfully utilising countermeasures may be less likely to disclose doing so.

Nevertheless, basic science gives reason for concern that polygraph test accuracy may be degraded by countermeasures. All the physiological indicators measured by the polygraph can be altered by conscious efforts through cognitive or physical means, and there is enough empirical research to justify concern that successful countermeasures may be learnable. This research does not clarify, however, whether users of countermeasures can be detected in contexts in which systematic efforts are made to detect or deter them. The possible effects of countermeasures are particularly significant to the extent that the polygraph is used and relied on for monitoring sex offences, since even a small false negative rate could have serious consequences. Similarly there also remains the danger that professionals may develop an unjustified confidence in the instrument's abilities. This could create a false sense of security amongst professionals and also lead to inappropriate relaxation of other means of supervision or treatment.



### *Some recommendations for use of polygraphy with sexual offenders*

The following guidelines are suggestions for how the polygraph may be effectively utilised in a post-conviction capacity with community-based sex offenders.

1. The focus of the polygraph test should be to gather information about the offenders previous and current behaviour, it should not be used to determine guilt or innocence regarding a specific offence.
2. It is unreasonable to expect offenders to make disclosures that could lead to additional prosecutions for self-disclosed crimes made prior to treatment or supervision. Measures should be taken to ensure that offenders have the guidelines with regards to what they should disclose.
3. Professionals (i.e. probation officers and clinicians) involved with polygraph should be aware of its strengths and weaknesses; specifically they should have a detailed understanding of false-positive and –negative rates.
4. The polygraph should be used in conjunction with other more conventional tools used in treatment and supervision, such as, clinical interviews, psychometric and drug testing.

5. Professionals should avoid being over reliant on the polygraph and seek to corroborate their information gained with other sources.
6. A polygraph result alone should not be considered sufficient evidence to determine facts or terminate treatment.
7. Polygraphers should use numerical scoring procedures.
8. Examiners should assist in developing empirically based guidelines for the use of polygraph as a treatment tool. They should participate as a treatment member, attending training that are specific to sexual offender treatment.
9. Polygraph examiners should be encouraged to participate in local, national and international organisations such as National Organisation for the Treatment of Abusers (NOTA) and Association for the Treatment of Sexual Abusers (ATSA).
10. Finally, it will be absolutely critical to establish a registration board in the UK to monitor the development and use of polygraphy in a post-conviction context. Such a body should be responsible for setting standards, promoting training and guarding against abusive practice. Adequate controls would enable the accumulation of clinical polygraph examination data

on sex offenders. This information would have research applications, and enable practitioners to gain a better understanding of offenders and offending behaviour.



The essence of any innovation involves a type of ‘turning the corner’ and advancing current practices in a new direction. Innovation however rarely develops within a vacuum though emerges through the creative application of theories and practices from other fields. The use of the polygraph with sex offenders is both innovative and is leading to a change of approach in both the supervision and treatment of such offenders. The findings reported in this thesis are consistent with the other research conducted in the area and indicate that the balance of evidence presently available, weighs in favour of using the polygraph in a treatment and supervision capacity with sex offenders. Supervision and treatment programmes that endeavour to incorporate polygraphy are likely gain significant amounts of pertinent information about their sex offenders. Arguably, this information can make treatment more effective, possibly assisting with lowering recidivism rates and perhaps even contributing to an improved quality of life for the offenders and their families. Polygraphy may also increase programme compliance and serve as a deterrent to engaging in risk behaviours. There is however no empirical evidence thus far that demonstrate that polygraph testing actually reduces recidivism, and programmes that employ the instrument may risk generating false admissions and wrongfully sanctioning substantial numbers of offenders. Programmes that reject polygraphy will avoid these ethical risks, but will almost certainly have to function with less information about their offenders. Consequently, they are likely to have a higher risk of undetected programme violations and perhaps even higher rates of sexual recidivism.

In reality, the use of the polygraph in this capacity is in its infancy, and many issues need to be resolved. Guidelines have been suggested for the responsible use of the polygraph in treatment programmes, though without replication of the present findings and additional research exploring the polygraph's application in a post-conviction capacity, it will continue to be viewed as a controversial tool. In the short-term, however, as long as those who use polygraphy do not fall into the trap of believing it to be the equivalent of Wonder Woman's magic lasso, there would seem to be a role for it in the management of sex offenders.

**PAGE**

**NUMBERING**

**AS ORIGINAL**



## CHAPTER 8

### References

Abel, G. & Rouleau, J.L. (1990). The nature and extent of sexual assault. In W.L. Marshall, D.R. Laws, & H.E. Barbaree (Eds.), *Handbook of Sexual Assault: Issues, Theories, and Treatment of the Offender*, New York: Plenum Press.

Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Sciences*, 17, 313 – 327.

Abrams, S. (1989). Probation polygraph surveillance of child abusers. *The Prosecutor*, 22, 29 – 38.

Abrams, S. (1991). The use of the polygraph with sex offenders. *Annals of Sex Research*, 4, 239 – 263.

Abrams, S. (1999). A response to Honts on the issue of the discussion of questions between charts. *Polygraph*, 28, 223 – 228.

Abrams, S. & Abrams, J.B. (1993). *Polygraph Testing of the Pedophile*. Portland, OR: Ryan Gwinner Press.

Abrams, S. & Ogard, E. (1986). Polygraph surveillance of probationers.

*Polygraph*, 15, 174 – 182.

Abrams, S. & Simmons, G. (2000). Post-conviction polygraph testing: Then and

now. *Polygraph*, 29 (1), 63 – 67.

Ahlmeyer, S., Heil, P., McKee, B. & English, K. (2000). The impact of polygraphy

on admissions of victims and offenses in adult sexual offenders. *Sexual*

*Abuse: A Journal of Research and Treatment*. 12 (2), 123 – 139.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of*

*Mental Disorders* (4<sup>th</sup> Ed.). Washington DC : American Psychiatric

Association.

Anolli, L. & Ciceri, R. (1997). The voice of deception: vocal strategies of naïve

and able liars. *Journal of Nonverbal Behavior*, 21 (4), 259 – 283.

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W. & Swinson, R.P. (1998).

Psychometric properties of the 42-item and 21-item versions of the

Depression Anxiety Stress Scales (DASS) in clinical groups and a

community sample. *Psychological Assessment*, 10, 176-181.

Armstrong, J.G. (1999). False memories and true lies: the psychology of a

recanter. *The Journal of Psychiatry and Law*, 27, 519 - 547.

Ashton, M.C. & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, 15, 327 – 353.

Association for the Treatment of Sexual Abusers. (1997). *Ethical standards and principles for the management of sexual abusers*. Beaverton, OR: Author.

Axelrod, S.R., Widiger, T.A., Trull, T.J. & Corbitt, E.M. (1997). Relations of five-factor model antagonism facets with personality disorder symptomatology. *Journal of Personality Assessment*, 69 (2), 297 – 313.

Backster, C. (1963a). Polygraph professionalism through technique standardization. *Law & Order*, 11 (4), 63 – 64.

Backster, C. (1963b). New standards in polygraph chart interpretation – do the charts speak for themselves? *Law & Order*, 11 (6), 67 – 68.

Bagby, R.M., Costa, P.T., McCrae, R.R., Livesley, W.J., Kennedy, S.H., Levitan, R.D., Levitt, A.J., Joffe, R.T. & Young, L.T. (1999). Replicating the five-factor model of personality in a psychiatric sample. *Personality and Individual Differences*, 27, 1135 – 1139.

Bagby, R.M., Russell, J., Parker, J., Kalembo, V. & Harkness, K. (1995). Major depression and the five-factor model of personality. *Journal of Personality Disorder*, 9, 224 – 234.



- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12, 387 – 415.
- Barbaree, H.E. (1991). Denial and minimization among sex offenders: Assessment and treatment outcome. *Forum on Corrections research*, 3, 30 – 33.
- Barland, G.H. (1999). A message. *American Polygraph Association Newsletter*, 32 (3), 16 – 17.
- Barland, G.H. & Raskin, D.C. (1975). Psychopathy and detection of deception in criminal suspects. *Psychophysiology*, 12, 224.
- Barratt, E.S. (1965). Factor analysis of some psychometric measures of impulsiveness and anxiety. *Psychological Reports*, 16, 547 – 554.
- Barrick, M.R. & Mount, M.K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81 (3), 261 – 272.
- Bashore, T.R. & Rapp, P.E. (1993). Are there alternatives to traditional polygraph procedures? *Psychological Bulletin*, 113 (1), 3 – 22.
- Beardsley, T. (1999). Truth or Consequences: A polygraph screening program

raises questions about the science of lie detection, *Scientific American*, 140, 21-24.

Beech, A.R. & Fisher, D.D. (2002). The rehabilitation of child sex offenders. *Australian Psychologist*, 37, 206 –214.

Ben-Shakhar, G. (2002). A critical review of the control question test (CQT). In M. Kleiner (Ed.). *Handbook of Polygraph Testing* (pp. 103 – 126). London: Academic Press.

Ben-Shakhar, G., Bar-Hillel, M. & Lieblich, I. (1986). Trial by polygraph: Scientific and juridical issues in lie detection. *Behavioral Sciences & the Law*, 4 (4), 459 – 479.

Ben-Shakhar, G. & Dolev, K. (1996). Psychophysiological detection through guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, 81 (3), 273 – 281.

Ben-Shakhar, G. & Lieblich, I. (1984). On statistical detection of deception: comment on Szucko and Kleinmuntz. *American Psychologist*, 79 – 80.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1982). An evaluation of polygraphers' judgements: A review from a decision theoretic perspective. *Journal of Applied Psychology*, 67 (6), 701 – 713.

Bersh, P.L. (1969). A validation study of polygraph examiner judgements. *Journal of Applied Psychology*, 53, 399 – 403.

Birgisson, G.H. (1996). Differences of personality, defensiveness, and compliance between admitting and denying male sex offenders. *Journal of Interpersonal Violence*, 11 (1), 118 – 125.

Blackburn, R. (1989). Psychopathy and personality disorder in relation to violence. In K.Howells & C.R. Hollin (Eds). *Clinical Approaches to Violence* (61 – 87). London: Wiley.

Blackburn, R. (2000). Treatment or incapacitation? Implications of research on personality disorders for the management of dangerous offenders. *Legal and Criminological Psychology*, 5, 1 – 21.

Blais, M.A. (1997). Clinician ratings of the Five-Factor Model of personality and the DSM-IV personality disorders. *Journal of Nervous and Mental Disease*, 185, 388 – 393.

Blascovich, J., Mendes, W.B., Hunter, S. & Lickel, B. (2000). Challenge, threat, stigma. In T.Heatherton, R. Kleck, M. Hebl, and J.G. Hull (eds.) *The Social Psychology of Stigma* (pp. 307 – 333). New York: Guilford Press.

Blascovich, J., Spencer, S., Quinn, D. & Steele, C. (2001a). African Americans



and high blood pressure: the role of stereotype threat. *Psychological Science*, 12, 225 – 229.

Blascovich, J., Mendes, W., Hunter, S., Lickel, B. & Kowai-Bell, N. (2001b). Perceiver threat in social interactions with stigmatised others. *Journal of Personality and Social Psychology*, 80, 253 – 267.

Blasingame, G.D. (1998). Suggested clinical uses of polygraphy in community-based sexual offender treatment programs. *Sexual Abuse: A journal of Research and Treatment*, 10 (1), 37 – 45.

Boaz, T.L., Perry, N.W., Raney, G., Fischler, I.S. & Shuman, D. (1991). Detection of guilty knowledge with event-related potentials. *Journal of Applied Psychology*, 76 (6), 788 - 795.

Bond, C.F. & Atoum, A.O. (2000). International deception. *Personality and Social Psychology Bulletin*, 26 (3), 385 – 395.

Bond, C.F., Omar, A., Pitre, U., Lashley, B.R., Skaggs, L.M., & Kirk, C.T. (1992). Fishy-looking liars: deception judgement from expectancy violation. *Journal of Personality and Social Psychology*, 63 (6), 969 – 977.

Bradley, M.T. (1988). Choice and the detection of deception. *Perceptual and Motor Skills*, 66, 43 – 48.

- Bradley, M.T. & Ainsworth, D. (1984). Alcohol and the psychophysiological detection of deception. *Psychophysiology*, 21 (1), 63 – 71.
- Bradley, M.T. & Black, M.E. (1998). A control question test orientated towards students. *Perceptual and Motor Skills*, 87, 691 – 700.
- Bradley, M.T. & Cullen, M.C. (1993). Polygraph lie detection on real events in a laboratory setting. *Perceptual and Motor Skills*, 76, 1051 – 1058.
- Bradley, M.T. & Janisse, M.P. (1981a). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and papillary measures. *Psychophysiology*, 18 (3), 397 – 315.
- Bradley, M.T. & Janisse, M.P. (1981b). Extraversion and the detection of deception. *Personality and Individual Differences*, 2, 99 – 103.
- Bradley, M.T. & Kohn, K.I. (1987). Machiavellianism, the control question test and the detection of deception. *Perceptual and Motor Skills*, 64, 747 – 757.
- Bradley, M.T., MacLaren, V.V. & Black, M.E. (1996). The control question test in polygraphic examinations with actual controls for truth. *Perceptual and Motor Skills*, 83, 755 – 762.
- Bradley, M.T., MacLaren, V.V. & Carle, S.B. (1996). Deception and nondeception

in guilty knowledge and guilty actions polygraph tests. *Journal of Applied Psychology*, 81 (2), 153 – 160.

Bradley, M.T. & Rettinger, J. (1992). Awareness of crime-relevant information and the guilty knowledge test. *Journal of Applied Psychology*. 77 (1), 55 – 59.

British Psychological Society. (1986). Report of the working group on the use of the polygraph in criminal investigation and personnel screening. *Bulletin of the British Psychological Society*, 39, 81 – 94.

Brooks, J. (1985). Polygraph testing thoughts of a sceptical legislator. *American Psychologist*, 348 – 354.

Brown, T.A., Korotitsch, W., Chorpita, B.F. & Barlow, D.H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, 35, 79-89.

Browne, K.D. & Lynch, M.A. (1998). Protecting children from sex offenders. *Child Abuse Review*, 7, 369 - 378.

Buckley, J.P. & Senese, L.C. (1991). The influence of race and gender on blind polygraph analyses. *Polygraph*, 20 (4), 247 – 258.

Buller, D.B., Comstock, J., Aune, R.K. & Strzyzewski, K.D. (1989). The effect of



probing on deceivers and truth-tellers. *Journal of Nonverbal Behavior*, 13 (3), 155 –170.

Bumby, K.M. & Hansen, D.J. (1997). Intimacy deficits, fear of intimacy, and loneliness among sexual offenders. *Criminal Justice and Behavior*, 24 (3), 315 – 331.

Cacioppo, J.T. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, 45, 16 – 28.

Campbell, T.W. (2000). Sexual predator evaluations and phrenology: considering issues of evidentiary reliability. *Behavioral Sciences and the Law*, 18, 111 – 130.

Carpenter, D.R., Peed, S.F. & Eastman, B. (1995). Personality characteristics of adolescent sexual offenders: A pilot study. *Sexual Abuse: A Journal of Research and Treatment*, 7 (3), 195 – 203.

Center for Sex Offender Management. (2000a). *Community Supervision of the Sex Offender: An overview of current and promising practices*. Silver Spring, MD: Author.

Center for Sex Offender Management. (2000b). *Public opinion and the Criminal Justice System: Building support for Sex Offender Management Programs*. Silver Spring, MD: Author.

Center for Sex Offender Management. (2001). *Recidivism of Sex Offenders*. Silver Spring, MD: Author.

Chambers, H. (1994) 'Snohomish county juvenile court sex offender treatment programme policy statement on the use of polygraph in treatment of juvenile sex offenders'. *Interchange*. November, 1- 6.

Clark, L.A., Livesley, W.J. & Morey, L. (1997). Personality disorder assessment: the challenge of construct validity. *Journal of Personality Disorders*. 11 (3), 232 – 241.

Clarkin, J.F., Hull, J.W., Cantor, J. and Sanderson, C. (1993). Borderline personality disorder and personality traits: A comparison of SCID-II BPD and NEO-PI. *Psychological Assessment*. 5(4), 472 – 476.

Clark, L.A., Vorhies, L. & McEwan, J.L. (1994). Personality disorder symptomatology from five-factor model perspective. In P.T.Costa & T.A.Widiger (Eds.). *Personality Disorders and the five-factor model of personality* (pp 95 – 116). Washington: American Psychological Association.

Clarkin, J.F., Hull, J.W., Cantor, J. and Sanderson, C. (1993). Borderline personality disorder and personality traits: A comparison of SCID-II BPD and NEO-PI. *Psychological Assessment*. 5(4), 472 – 476.

Clower, C.E. & Bothwell, R.K. (2001). An exploratory study of the relationship between the big five and inmate recidivism. *Journal of Research in Personality*. 35, 231 – 237.

Colorado Sex Offender Management Board. (2000). *Research supporting restricted contact with children*. Colorado: Author.

Consigli, J. (2002). Post-conviction sex offender testing and the American Polygraph Association. In Kleiner, M. (Ed.), *Handbook of Polygraph Testing* (pp.237 – 249). London: Academic Press.

Cooley-Towell, S., Pasini-Hill, D. & Patrick, D. (2000) ‘The value of the post-conviction polygraph: The importance of sanctions’. *Polygraph*, 29(1), 6-19.

Cooper, A., Delmonico, D.L., & Burg, R. (2000). Cybersex users, abusers, and compulsives: New findings and implications. *Sexual Addiction & Compulsivity*, 7, 5 – 29.

Corruble, E., Damy, C., & Guelfi, J.D. (1999). Impulsivity: a relevant dimension in depression regarding suicide attempts? *Journal of Affective Disorders*, 53, 211 – 215.

Costa, P.T. & McCrae, R.R. (1992a). *Revised NEO Personality Inventory (NEO-*



*PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.*

Odessa, FL: Psychological Assessment Resources.

Costa, P.T. & McCrae, R.P. (1992b). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4 (1), 5 – 13.

Costa, P.T. & McCrae, R.R. (1997). Stability and change in personality assessment: The revised NEO personality inventory in the year 2000. *Journal of Personality Assessment*, 68 (1), 86 – 94.

Costa, P.T. & Widiger, T. (1994). Summary and unresolved issues. In P. Costa & T. Widiger (Eds.). *Personality disorders and the five-factor model of personality* (pp. 319 – 327). Washington DC: American Psychological Association.

Crawford, J.R., Besson, J.A.O., Parker, D.M., Sutherland, K.M., & Keen, P.L. (1987). Estimation of premorbid intellectual status in depression. *British Journal of Clinical Psychology*, 26, 313 – 314.

Crawford, J.R., Stewart, L.E., Cochrane, B., Parker, D.M. & Besson, A.O. (1989). Construct validity of the National Adult Reading Test: a factor analytic study. *Personality and Individual Differences*, 10 (5), 585 – 587.

Cross, T.P. & Saxe, L. (2001). Polygraph testing and sexual abuse: The lure of the

magic lasso. *Child Maltreatment*, 6 (3), 195 – 206.

Cumming, G.H. & Buell, M.M. (1996). Relapse prevention as a supervision strategy for sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 8 (3), 231 - 241.

Daniels, C.W. (2002). Legal aspects of polygraph admissibility in the United States. In M.Kleiner (Ed.). *The Handbook of Polygraph Testing* (pp. 327 – 338). London: Academic Press.

Daubert v. Merrell Dow, Inc., 509 US 579. (1993)

Davis, R.C. (1961). Physiological responses as a means of evaluating information. p. 142 – 168 in *Manipulation of Human Behaviour*, A.Biderman & H.Zimmer, eds. New York: Wiley.

Davidson, P.O. (1968). Validity of the guilty-knowledge technique: the effects of motivation. *Journal of Applied Psychology*. 52 (1), 62 – 65.

Dawson, M. E. (1980). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers, *Psychophysiology* 17:8-17.

DePaulo, B.M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111 (2), 203 – 243.

DePaulo, B.M., Lindsay, J.L., Malone, B.E., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74 – 118.

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

DePaulo, B.M. & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology*, 37 (10), 1713 – 1722.

Dennison, S.M., Stough, C. & Birgden, A. (2001). The big 5 dimensional personality approach to understanding sex offenders. *Psychology, Crime and Law*, 7, 243 – 261.

Deu, N. & Edelmann, R.J. (1997). The role of criminal fantasy in predatory and opportunist sex offending. *Journal of Interpersonal Violence*, 12 (1), 18 - 29.

Diven, K. (1937). Certain determinants in the conditioning of anxiety reactions. *Journal of Psychology*, 3: 291 – 308.

Dohm, T. E., & Iacono, W. G. (1993). Design and pilot of a polygraph filed validation study with appendices (DoDPI93-R- 0006). Fort McClellan, AL: Department of Defense Polygraph Institute.



- Dollins, A.B., Cestaro, V.L., & Pettit, D.J. (1998). Efficacy of repeated psychophysiological detection of deception testing. *Journal of Forensic Science*, 43 (5), 1016 - 1023.
- Dollins, A.B., Kraphol, D.J. & Dutton, D.W. (2000). A comparison of computer programs designed to evaluate psychophysiological detection of deception examinations: Bakeoff 1. *Polygraph*, 29 (3), 237 – 257.
- Doren, D. (1998). Recidivism base rates, predictions of sex offender recidivism, and the "sexual predator" commitment laws. *Behavioral Sciences and the Law*, 16, 97 - 114.
- Dutton, D. (2000). Post-conviction sex offender testing: an introduction. *Polygraph*, 29 (1), 1 – 6.
- Dripps, D. A. (1997). Polygraph Evidence: Winds of Change? *Trial*, 33, 74-77.
- Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66, 183-201.
- Edson, C. (1991). *Sex offender treatment*. Medford, OR: Department of Corrections.
- Ekman, P. (1992). *Telling Lies*. London: W.W.Norton & Company.

- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913-920.
- Ekselius, L., Lindstrom, E., von Knorring, L., Bodlund, O. & Kullgren, G. (1994). SCID-II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III-R. *Psychiatric Journal of Scandinavia*, 90 (2), 120 –123.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, 75 (5), 521 - 529.
- Elaad, E., Ginton, A. & Ben-Shakhar, G. (1998). The role of prior expectations in polygraph examiners decisions. *Psychology, Crime & Law*, 4, 1 – 16.
- Elaad, E., Ginton, A. & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, 77 (5), 757 – 767.
- Emerick, R.L. & Dutton, W.A. (1993). The effect of polygraphy on the self report of adolescent sex offenders: implications for risk assessment. *Annals of Sex Research*, 6, 83 – 103.
- English, K. (1998). The containment approach: An aggressive strategy for he

community management of adult sex offenders. *Psychology, Public Policy, & Law Special Issue: Sex Offenders; Scientific, Legal, and Policy Perspectives*, 4, 218 – 235.

English, K., Pullen, S. & Jones, L. (1996). *Managing adult sex offenders: A containment approach*. Lexington, KY: American Probation and Parole Association.

English, K., Jones, L., Pasini-Hill, D. & Cooley-Towell, S. (2000). *The value of polygraph testing in sex offender management*. Research report submitted to the National Institute of Justice. Colorado: Colorado Department of Public Safety.

English, K., Jones, L., Patrick, D., Pasini-Hill, D., & Gonzalez, S. (2000). We need you to become experts in the post-conviction polygraph. *Polygraph*, 29 (1), 44 – 62.

Fagon, P.J., Wise, T.N., Schmidt, C.W., Ponticas, Y., Marshall, R.D. & Costa, P.T. (1991). A comparison of five-factor personality dimensions in males with sexual dysfunction and males with paraphilias. *Journal of Personality Assessment*, 57 (3), 434 – 448.

Falkenhain, M.A., Duckro, P.N., Hughes, H.M., Rossetti, S.J. & Gfeller, J.D.



(1999). Cluster analysis of child sex offenders: A validation with Roman Catholic priests and brothers. *Sexual Addiction & Compulsivity*, 6, 317 – 336.

Faller, K. (1997). The polygraph, its uses in cases of alleged sexual abuse: An exploratory study. *Child Abuse & Neglect*, 24, 1215 – 1225.

Fazel, S., Hope, T., O'Donnell, I. & Jacoby, R. (2002). Psychiatric, demographic and personality characteristics of elderly sex offenders. *Psychological Medicine*. 32(2), 219-226

Fedoroff, J.P. & Moran, B. (1997). Myths and misconceptions about sex offenders. *The Canadian Journal of Human Sexuality*, 6 (4), 263 - 275.

Fieldler, K., Schmid, J. & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology*, 24 (4), 313 – 324.

Finkelhor, D. & Araji, S. (1986). Explanations of pedophilia: A four-factor model. *The Journal of Sex Research*, 22(2), 145 – 161.

Fisher, D., Beech, A. & Browne, K. (1999). Comparison of sex offenders to nonoffenders on selected psychological measures. *International Journal of Offender Therapy and Comparative Criminology*, 43 (4), 473 – 491.

Ford, C. (1996). *Lies! Lies! Lies! The Psychology of Deceit*. American Psychiatric Press: Washington, DC.

Ford, C., King, B. & Hollender, M. (1988). Lies and liars: Psychiatric aspects of prevarication. *American Journal of Psychology*, 145, 554 – 562.

Forensic Research. (1997). Validity and reliability of polygraph testing. *Polygraph*, 26, 215 – 239.

Forman, R.F. & McCauley, C. (1986). Validity of the positive control polygraph test using the field practice model. *Journal of Applied Psychology*, 71 (4), 691 – 698.

Fraedrich, S. & Pfaefflin, F. (2000). The prevalence of personality disorders in a prison population/Zur Praevalenz von Persoenlichkeitsstoerungen bei Strafgefangenen. *Recht und Psychiatrie*. 18(3), 95-104

Francis, B. & Soothill, K. (2000). Does sex offending lead to homicide? *The Journal of Forensic Psychiatry*, 11 (1), 49 - 61.

Furby, L., Weinrott, M.R. & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin*, 105 (1), 3 – 30.

Furedy, J.J. (1993). The 'control' question 'test' (CQT) polygrapher's dilemma:

- logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 15, 263 - 267.
- Furedy, J.J. (1996a). Some elementary distinctions among, and comments concerning the 'control' question 'test' (CQT) polygrapher's many problems: A reply to Honts, Kircher, and Raskin. *International Journal of Psychophysiology*, 1 – 10.
- Furedy, J.J. (1996b). The North American polygraph and psychophysiology: disinterested, uninterested and interested perspectives. *International Journal of Psychophysiology*, 21, 97 - 105.
- Geer, J.H., Estupinan, L.A., & Manguno-Mire, G.M. (2000). Empathy, social skills, and other relevant cognitive processes in rapists and child molesters. *Aggression and Violent Behavior*, 5(1), 99 – 126.
- Ginton, A., Daie, N., Elaad, E. & Ben-Shakhar, G. (1982). A method of evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67 (2), 131 - 137.
- Greenberg, E. (1989). Workplace Testing: Who's Testing Whom? *Personnel*, 66, 39 – 46.
- Grime, L.G. & Yarhold, P.R. (Eds.).(2000). *Reading and Understanding more Multivariate Statistics*. American Psychological Association: Washington,



DC.

- Gudjonsson, G.H. (1990). Self-deception and other-deception in forensic assessment. *Personality and Individual Differences*, 11 (3), 219 - 225.
- Gudjonsson, G.H. (1999). The making of a serial false confessor: the confessions of Henry Lee Lucas. *The Journal of Forensic Psychiatry*, 10 (2), 416 - 426.
- Gudjonsson, G.H. (2001). False confession. *The Psychologist*, 14, 588 – 591.
- Gudjonsson, G.H. (2003). *The Psychology of Interrogations and confessions: A handbook*. Chichester: Wiley.
- Gudjonsson, G.H. & Haward, L.R. (1982). Detection of deception: Consistency in responding and personality. *Perceptual and Motor Skills*, 54, 1189 – 1190.
- Gudjonsson, G.H. & Moore, E. (2001). Self-deception and other-deception among admissions to a maximum security hospital and a medium secure unit. *Psychology, Crime and Law*, 7, 25 -31.
- Gudjonsson, G.H. & Petursson, H. (1991). Custodial interrogation: Why do suspects confess and how does it related to their crime, attitude and personality? *Personality and Individual Differences*, 12 (3), 295 – 306.
- Gudjonsson, G.H. & Sigurdsson, J.F. (1994). How frequently do false confessions

occur? An empirical study among prison inmates. *Psychology, Crime and Law*, 1, 21 – 26.

Gudjonsson, G.H. & Sigurdsson, J.F. (2004). Motivation for offending and personality. *Legal and Criminological Psychology*, 9, 69 – 81.

Gudjonsson, G.H. & Sigurdsson, J.F. (1999). The Gudjonsson Confession Questionnaire Revised (GCQ-R): Factor structure and its relationship to personality. *Personality and Individual Differences*, 27, 953 – 968.

Gudjonsson, G.H. & Sigurdsson, J.F. (2000). Differences and similarities between violent offenders and sex offenders. *Child Abuse & Neglect*, 24 (3), 363 – 372.

Gudjonsson, G.H., Sigurdsson, J.F., Bragason, O.O., Einarsson, E. & Valdimarsdottir, E.B. (2004). Confessions and denials and the relationship with personality. *Legal and Criminological Psychology*, 9, 121 – 133.

Guertin, W.H. & Wilhelm, P.L. (1954). A statistical analysis of the electrodermal response employed in lie detection. *The Journal of General Psychology*, 51, 153 – 160.

Gustafson, L.A. & Orne, M.T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology*, 47 (6): 408 – 411.

- Hall, G.C.N. (1995). Sexual offender recidivism: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, 63 (5), 802 - 809.
- Hanson, R.K. & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66 (2), 348 – 362.
- Hanson, R.K. & Harris, A.J.R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behaviour*, 27 (1), 6 – 35.
- Hanson, R.K. & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24 (1), 119 - 136.
- Hare, R.D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, 23(1), 25 – 54.
- Harrelson, L.H., Gerow, N. & Gerow, J. (1998). *Lie Test: deception, truth and the polygraph*. Fort Wayne, USA: Jonas Publishing.
- Harrison, J.S. & Kirkpatrick, B. (2000). Polygraph testing and behavioural change with sex offenders in an outpatient setting: an exploratory study. *Polygraph*, 29 (1), 20 – 26.



Heersink, N. & Strassberg, D.S. (1995). A normative and descriptive study of the MMPI with acknowledged child molesters. *Journal of Psychopathology and Behavioral Assessment*, 17 (4), 377 – 391.

Heil, P., Ahlmeyer, S., Simons, D. & English, K. (1999). *The impact of polygraph admissions of crossover offending behavior in adult sexual offenders.*

Research presented at the annual Association for the Treatment of Sexual Abusers Research and Treatment Conference, Lake Buena Vista, Florida.

Heil, P., Ahlmeyer, S., McCullar, B., & mc Kee, B. (2000). Integration of polygraph testing with sexual offenders in the Colorado Department of Corrections. *Polygraph*, 29, 26 – 35.

Herkov, M.J., Gynther, M.D., Thomas, S. & Myers, W.C. (1996). MMPI differences among adolescent inpatients, rapists, sodomists, and sexual abusers. *Journal of Personality Assessment*, 66 (1), 81 – 90.

Hindman, J. & Peters, J.M. (2001). Polygraph testing leads to better understanding adult and juvenile sex offenders. *Federal Probation*, 65 (1), 8 – 15.

Holden, E.J. (2000). Pre- and post-conviction polygraph: building blocks for the future – procedures, principles, and practices. *Polygraph*, 29 (1), 69 – 97.

Holt, S.E., Meloy, J.R. & Strack, S. (1999). Sadism and psychopathy in violent and

sexually violent offenders. *Journal of the American Academy of Psychiatry Law*, 27 (1), 23 – 32.

Home Office (1995). *Information on the criminal justice system in England and Wales: Digest 3*. London: Home Office Research and Statistics Department.

Honts, C.R. (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports*, 4, 91 – 116.

Honts, C.R. (1996). Criterion development and validity of the control question test in field application. *The Journal of General Psychology*, 123, 309 – 324.

Honts, C.R., Devitt, M.K., Winbush, M. & Kircher, J.C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, 33, 84 – 92.

Honts, C., Kircher, J.C. & Raskin, D.C. (1995). Polygrapher's dilemma or psychologist's chimaera: A reply to Furedy's logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 20, 199 – 207.

Honts, C.R. & Perry, M.V. (1992). Polygraph admissibility. *Law and Human Behavior*, 16 (3), 357 - 379.

Honts, C.R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social

Scientists as Amicus Curiae. *United States v. Scheffer*, in the Supreme Court of the United States. Retrieved from [www.truth.boisestate.edu/polygraph/polylaw.html](http://www.truth.boisestate.edu/polygraph/polylaw.html)

Honts, C. R. & Quick, B.D. (1995). The polygraph in 1995: progress in science and the law. *North Dakota Law Review*, 71, 987 – 1020.

Honts, C.R. & Raskin, D.C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16 (1), 56 – 61.

Honts, C.R., Raskin, D.C. & Kircher, J.C. (1985). Effects of socialization on the physiological detection of deception. *Journal of Research in Personality*, 19, 373 – 385.

Honts, C.R., Raskin, D. & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79 (2), 252 – 259.

Horowitz, S.W., Kircher, J.C., Honts, C.R. & Raskin, D.C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108 – 115.

Horvath, F. (1976). Detection of deception: A review of field and laboratory research. *Polygraph*, 5: 107 – 145.



- Horvath, F. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 62 (2), 127 – 136.
- Hunter, F.L. & Ash, P. (1973). The accuracy and consistency of polygraph examiners' diagnoses. *Journal of Police Science and Administration*, 1, 370 – 375.
- Iacono, W.G., Boisvenu, G.A., & Fleming, J.A.E. (1984). Effects of diazepam and methylphenidate on electrodermal detection of guilty knowledge, *Journal of Applied Psychology*, 69 (2), 289 – 299.
- Iacono, W.G., Cerri, A.M., Patrick, C.J. & Fleming, J.A. (1992). Use of antianxiety drugs as countermeasures in the detection of guilty knowledge. *Journal of Applied Psychology*, 77 (1), 60 – 64.
- Iacono, W.G. & Lykken, D.T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology*, 82 (3), 426 - 433.
- Jacobsberg, L., Perry, S. & Frances, A. (1995). Diagnostic agreement between the SCID-II screening questionnaire and the personality disorder examination. *Journal of Personality Assessment*, 65 (3), 428 – 433.
- Jannisse, M.P. & Bradley, M.T. (1980). Deception, information, and the papillary response. *Perceptual and Motor Skills*, 50, 748 – 750.

- Jones, E.E. & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76 (5), 349 – 364.
- Kashy, D.A. & DePaulo, B.M. (1996). Who lies? *Journal of Personality and Social Psychology*, 70, 1037 – 1051.
- Kircher, J.C., Horowitz, S.W. & Raskin, D.C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behaviour*, 12, 79 – 90.
- Kicher, J. C., Packard, T.; Bell, B. G. & Bernhardt, P.C. (2001) Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Probable-Lie and Directed-Lie Polygraph Tests. Report No. DoDPI02-R-0002. Ft. Jackson: U.S. Department of Defense Polygraph Institute.
- Kircher, J.C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73 (2), 291 - 302.
- Kircher, J.C. & Raskin, D.C. (2002). Computer methods for the psychophysiological detection of deception. In M.Kleiner (Ed.). *The Handbook of Polygraph Testing* (pp.287 - 326). London: Academic Press.
- Kleinmuntz, B. & Szucko, J.J. (1984). Lie detection in ancient and modern times.

*American Psychologist*, 766 - 776.

Knox D, Schact C. & Holt J. (1993). Sexual lies among university students.

*College Student Journal*, 27, 269-272.

Kohnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, 2, 1-17.

Kokish, R. (2003). The current role of post-conviction sex offenders polygraph testing in sex offender treatment. *Journal of Child Sexual Abuse*, 12 (3/4), 175 – 194.

Kokish, R. & Blasingham, G.D. (2002). An Exploration of the Accuracy, Benefits and Detriments of Post-conviction Sex-offender Polygraph Testing Using Offender Self-report. Unpublished study.

Kraut, R.E. and Poe, D. (1980). On the line: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology*, 39: 784-98.

Krapohl, D.J., Shull, K.W., & Ryan, A.A. (2002). Does the confession criterion in case selection inflate polygraph accuracy estimates? *Forensic Science Communications*, 4 (3), 3 – 14.

Krueger, R.F., Caspi, A., Moffitt, T.E., Silva, P.A., & McGee, R. (1996).



Personality traits are differentially linked to mental disorders: A multi-trait-multidiagnosis study of an adolescent birth cohort. *Journal of Abnormal Psychology*, 105 (3), 299 – 312.

Larson, J.A. (1932). *Lying and its Detection*. Chicago: University of Chicago Press.

Larson, J.A. (1938). The lie detector polygraph: Its history and development. *Journal of the Michigan State Medical Society*, 37, 893 – 897.

Lamb, M. (2001). Who was Wonder Woman? *Bostonia* [Online] Available: [www.bu.edu/alumni/bostonia/fall2001/ww/index.html/](http://www.bu.edu/alumni/bostonia/fall2001/ww/index.html/)

Langevin, R., Lang, R.A. & Curnoe, S. (1998). The prevalence of sex offenders with deviant fantasies. *Journal of Interpersonal Violence*, 13 (3), 315 - 327.

Lanyon, R.I. (1996). Assessment of specific deception strategies used by personality inventory respondents. *American Journal of Forensic Psychology*, 14 (3), 37 – 53.

Lanyon, R.I. (1997). Detecting deception: Current models and directions. *Clinical Psychology: Science and Practice*. 4 (4), 377 – 387.

Launay, G. (2001). Relapse prevention with sex offenders: practice, theory and research. *Criminal Behaviour and Mental Health*, 11, 38 – 54.

Laws, D.R. (1999). Relapse prevention. *Journal of Interpersonal Violence*, 14 (3), 285 - 302.

Lawson, A. (1988). *An Analysis of Love and Betrayal*. Basic Books: New York

Leary, M.R., Patton, K.M., Orlando, A.E. & Funk, W.W. (2000). The impostor phenomenon: Self-perceptions, reflected appraisals, and interpersonal strategies. *Journal of Personality*, 68 (4), 725 – 756.

Lehrer, E. (1998). Lies, Damned Lies and Polygraph Tests. *Insight on the News*, 14, 44-45.

Lehne, G.K. (1994). The NEO-PI and the MCMI in the forensic evaluation of sex offenders. In P.T.Costa & T.A.Widiger (Eds.). *Personality Disorders and the five-factor model of personality* (pp 175 – 188). Washington: American Psychological Association.

LeDoux, J.E. (1995). Emotion: Clues from the brain. *Annual Review of Psychology*, 46: 209 – 235.

Levin, S.M. & Stava, L. (1987). Personality characteristics of sex offenders: A review. *Archives of Sexual Behavior*, 16 (1), 57 – 77.

Lezak, M.D. (1995). *Neuropsychological Assessment* (3<sup>rd</sup> Ed.). Oxford University

Press: Oxford.

Lieblich, I., Ben-Shakhar, G., Kugelmass, S. & Cohen, Y. (1978). Decision theory approach to the problem of polygraph interrogation. *Journal of Applied Psychology*, 63 (4), 489 – 498.

Linehan, J. (1996). Polygraph research study of fear in a field situation review of a 1951 study. *Polygraph*, 25, 152 – 158.

Lovibond, S.H. & Lovibond, P.F. (1995a). Manual for the Depression Anxiety Stress Scales. (2<sup>nd</sup>. Ed.) Sydney: Psychology Foundation.

Lovibond, P.F. & Lovibond, S.H. (1995b). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335-343.

Lund, C.A. (2000). Predictors of sexual recidivism: Did meta-analysis clarify the role and relevance of denial? *Sexual Abuse: A Journal of Research and Treatment*. 12 (4), 275 - 287.

Lundell, R.G. (2000). Sex offender testing: Still basic polygraph. *Polygraph*, 29 (1), 40 – 43.

Lykken, D.T. (1974). Psychology and the lie detector industry. *American*



*Psychologist*, 725 - 739.

Lykken, D.T. (1979). The detection of deception. *Psychological Bulletin*, 86 (1), 47 – 53.

Lykken, D.T. (1988). Detection of guilty knowledge: A comment on Forman and McCauley, *Journal of Applied Psychology*, 73 (2), 303 - 304.

Lykken, D.T. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. New York: Plenum Press.

Lynn, R. (1966). *Attention, Arousal, and the Orientation Reaction*. Oxford: Pergamon.

Maier, R.A. & Lavrakas, P.J. (1976). Lying behavior and evaluation of lies. *Perceptual and Motor Skills*, 42, 575 – 581.

Maletzky, B.M. (1991). *Treating the sexual offender*. Newbury Park, CA: Sage Publications.

Maletzky, B.M. (1998). The paraphilias: Research and treatment, in P.E. Nathan & J.M. Gorman (1998). *A Guide to Treatments that Work* (pp. 472 – 500). Oxford University Press: New York.

Mandeville-Norden, R. & Beech, A. (2004). Community-based treatment of sex

offenders. *Journal of Sexual Aggression*, 10 (2), 193 – 214.

Mann, R.E.(2004). Innovations in sex offender treatment. *Journal of Sexual Aggression*, 10 (2), 141 – 152.

Maring, J. (1998). Using the Millon Adolescent Personality Inventory to differentiate sexually offending from nonsexually offending conduct-disorder adolescents. *Dissertation Abstracts International*. 59 (2 – B), 879.

Marques, J.K., Day, D.M., Nelson, C., West, M. (1994). Effects of cognitive-behavioral treatment on sex offender recidivism preliminary results of a longitudinal study. *Criminal Justice and Behavior*, 21 (1), 28 – 54.

Marsh, R.L. & Walsh, A. (1995). Physiological and psychosocial assessment and treatment of sex offenders: A comprehensive victim-orientated program. *Journal of Offender Rehabilitation*, 22 (1/2), 77 – 96.

Marshall, W.L., Anderson, D. & Champagne, F. (1997). Self-esteem and its relationship to sexual offending. *Psychology, Crime and Law*, 3, 161 - 186.

Marshall, W.L., Anderson, D. & Fernandez, Y. (1999). *Cognitive Behavioural Treatment of Sexual Offenders*. New York: Wiley.

Marston, W.M. (1917). Systolic blood pressure changes in deception. *Journal of Experimental Psychology*, 2: 117 – 163.

Marston, W.M. (1938). *The Lie Detector Test*. New York: Richard R. Smith.

Maschke, G.W. & Scalabrini, G.J. (2000) *The Lie Behind the Lie Detector*. Digital edition: Antipolygraph.org.

Matte, J.A. (1996). *Forensic Psychophysiology using the Polygraph: Scientific truth verification and lie detection*. Williamshire: J.A.M.

McCann, J.T. (1998). *Malingering and deception in adolescents: Assessing credibility in clinical and forensic settings*. Washington: American Psychological Association.

McCrae, R.R. & Costa, P.T. (1985). Comparison of the EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, 6 (5), 587 – 597.

McCrae, R.R. & Costa, P.T. (2003). *Personality in adulthood: A five-factor theory perspective*. 2<sup>nd</sup> Ed. New York: Guilford.

McCrae, R.R., Costa, P.T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M.D, Sanz, J., Sánchez-Bernardos, M.L., Kusdil, M.E., Woodfield, R.,  
Saunders, P.R. & Smith, P.B. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78 (1), 173 – 186.



McCrae, R.R., Yang, J., Costa, P.T., Dai, X., Yao, S., Cai, T. & Gao, B. (2001).

Personality profiles and the prediction of categorical personality disorders.

*Journal of Personality*, 69 (2), 155 – 174.

McElroy, S.L., Soutullo, C.A., Taylor, P., Nelson, E.B., Beckman,

D. A., Brusman, L. A., Ombaba, J. M., Strakowski, S. M., Keck, P. E.

(1999). Psychiatric features of 36 men convicted of sexual offenses.

*Journal of Clinical Psychiatry*. 60(6), 414 – 420.

McEwan, A.W. & Knowles, C. (1984). Delinquent personality types and the

situational contexts of their crimes. *Personality and Individual Differences*,

5(3), 339 – 344.

McGrath, R.J. (1991). Sex-offender risk assessment and dispositional planning: A

review of empirical and clinical findings. *International Journal of Offender*

*Therapy and Comparative Criminology*, 35 (4), 328 - 350.

McGuire, T. .J. (2000). Correctional institution based sex offender treatment: A

lapse behaviour study. *Behavioral Sciences and the Law*, 18, 57 - 71.

McGuire, W.J. (1969). The nature of attitudes and attitude changes. In *The*

*Handbook of Social Psychology*, 2<sup>nd</sup> ed., Vol. 3. G. Lindzey & E.Aronson,

eds. Reading, MA: Addison-Wesley.

- McKay, B. (2000). The state of polygraph testing on sex offenders under community supervision in Texas. *Polygraph*, 29 (1), 36 – 40.
- Melonas, I.A. (1998). A comparison of psychopathology in male adolescent child molesters and other male adolescent sex offenders. *Dissertations Abstracts International*. 58 (9 – B), 5130.
- Meloy, J.R. (2000). The nature and dynamics of sexual homicide: An integrative review. *Aggressive and Violent Behavior*, 5 (1), 1 - 22.
- Meloy, J.R. & Gacono, C.B. (1992). A psychotic (sexual) psychopath: “I just had a violent thought...”. *Journal of Personality*, 58 (3), 480 – 493.
- Miller, T.R. (1991). The psychotherapeutic utility of the five-factor model of personality: A clinician’s experience. *Journal of Personality Assessment*, 57 (3), 415 – 433.
- Miller, J.D., Lynam, D.R., Widiger, T.A., & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the five-factor model adequately represent psychopathy? *Journal of Personality*, 69 (2), 253 – 276.
- Miller, J.D., Pilkonis, P.A., & Morse, J.G. (2004). Five-factor prototypes for personality disorders: the utility of self-reports and observer ratings. *Assessment*, 11, 127 – 138.

- Monts, J.K., Zurcher, L.A. & Nydegger, R.V. (1977). Interpersonal self-deception and personality correlates. *The Journal of Social Psychology*, 103, 91 - 99.
- Morey, L.C., Gunderson, J., Quigley, B.D., & Lyons, M. (2002). Dimensions and categories: The “Big Five” factors and the DSM personality disorders. *Assessment*, 7, 203 – 216.
- Moston, S., Stephenson, G. & Williamson, T.M. (1992). The effects of case characteristics on suspect behaviour during questioning. *British Journal of Criminology*, 32, 23 – 40.
- Murphy, K.R. (1987). Detecting infrequent deception. *Journal of Applied Psychology*, 72 (4), 611 – 614.
- Munsterberg, H. (1908). *On the Witness Stand*, New York: McClure.
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigations in Japan. In M. Kleiner (Ed.). *Handbook of Polygraph Testing* (p. 49 – 86). Academic Press: London.
- National Research Council. (1994). *Under the Influence: Drugs and the American Workforce, Committee on Drug Use in the Workplace*, J. Normand, R. Lempert, & C. O’Brien, eds. Washington, DC: National Academy Press.



National Research Council (2003). *The Polygraph and Lie Detection*. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nelson, H.E. & Willison, J. (1991). *National Adult Reading Test (NART): Test Manual*. 2<sup>nd</sup> Edition. Windsor, UK: NFER Nelson.

Nichols, D.S. & Greene, R.L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment*, 68 (2), 251 – 266.

Nussbaum, D. & Rogers, R. (1992). Screening psychiatric patients for Axis II disorders. *Canadian Journal of Psychiatry*, 37, 658 – 660.

Oaksford, K.L. & Frude, N. (2001). The prevalence of child sexual abuse: Evidence from a female university sample in the UK. *Child Abuse Review*, 10, 49 – 59.

Office of Technology Assessment (1983). *Scientific validity of Polygraph Testing: A Research Review and Evaluation*. Washington, DC: US Government Printing Office.

Okami, P. & Goldberg, A. (1992). Personality correlates of pedophilia: Are they reliable indicators? *The Journal of Sex Research*, 29 (3), 297 – 328.

Oksenberg, L. (1970). Machiavellianism and emotionality. In R. Christie & F.L. Geis (Eds.), *Studies in Machiavellianism* (p.396). New York: Academic Press.

O'Connell, M.A. (1998) 'Using polygraph testing to assess deviant sexual history of sex offenders. *Dissertation Abstracts International*, 58, 3023

O'Connor, B.P. (2002). A quantitative study of the dimensions and facets of the five-factor model in the diagnosis of personality disorder. *Assessment*, 9: 188 – 203.

O'Toole, D.J., Yuille, J.C., Patrick, C.J. & Iacono, W.G. (1994). Alcohol and the physiological detection of deception: Arousal and memory influences. *Psychophysiology*, 31 : 253 – 263.

Oxford University Press. (1996). *The New Shorter Oxford English Dictionary*. University Press: Oxford.

Patterson, J. & Kim, P. (1991) *The Day America Told the Truth*. Prentice Hall: New York.

Pathé, M., Mullen, P.E., & Purcell, R. (1999). Stalking: false claims of victimisation. *British Journal of Psychiatry*, 174, 170 – 172.

Patrick, C.J. & Iacono, W.G. (1986). The validity of lie detection with criminal Psychopaths. *Psychophysiology*, 23, 452 – 453.

Patrick, C.J. & Iacono, W.G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology*, 74 (2), 347 - 355.

Patrick, C.J. & Iacono, W.G. (1991a). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology*, 28 (6), 632 – 638.

Patrick, C.J. & Iacono, W.G. (1991b). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology*, 76 (2), 229 – 238.

Paulhus, D.L. & John, O.P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66 (6), 1025 – 1059.

Piper, A. (1993). "Truth serum" and "recovered memories" of sexual abuse: a review of the evidence. *The Journal of Psychiatry and Law*, 447 - 471.

Podlesny, J.A.& Truslow, C.M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distribution-crime-roles context. *Journal of Applied Psychology*, 78 (5), 788 – 797.



- Polizzi, D.M. & MacKenzie, D.L. (1999). What works in adult sex offender treatment? A review of prison- and non-prison-based treatment programs. *International Journal of Offender Therapy and Comparative Criminology*, 43 (3), 357 – 374.
- Pollina, D.A., Dollins, A.B., Senter, S.M., Krapohl, D.J. & Ryan, A.H. (2004). Comparison of polygraph data obtained from individuals involved in mock and actual criminal investigations. *Journal of Applied Psychology*, 89 (6), 1099 – 1105.
- Porter, S., Fairweather, D., Drugge, J., Hervé, H., Birt, A. & Boer, D.P. (2000). Profiles of psychopathy in incarcerated sexual offenders. *Criminal Justice and Behavior*, 27 (2), 216 – 233.
- Prentky, R.A., Knight, R.A., & Lee, A.F. (1997). Risk factors associated with recidivism among extrafamilial child molesters. *Journal of Consulting and Clinical Psychology*, 65(1), 141 – 149.
- Prentky, R.A., Lee, A.F.S., Knight, R.A. & Cerce, D. (1997). Recidivism rates among child molesters and rapists: A methodological analysis. *Law and Human Behavior*, 21 (6), 635 - 659.
- Quigley-Fernandez, B. & Tedeschi, J.T. (1978). The bogus pipeline as lie detector: Two validity studies. *Journal of Personality and Social Psychology*, 36, 247 – 256.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15(2), 126-136.

Raskin, D.C. & Honts, C.R. (2002). The comparison question test. In M.Kleiner (Ed.). *The Handbook of Polygraph Testing* (pp. 1 – 47). London: Academic Press.

Raskin, D.C., Kircher, J.C., Honts, C.R. & Horowitz, S.W. (1988). A study of the validity of polygraph examinations in criminal investigations. Grant No. 85-IJ-CX-0040. Salt Lake City: Department of Psychology, University of Utah.

Raskin, D.C. & Podlesny, J.A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin*, 86 (1), 54 - 59.

Raymond, N.C., Coleman, E., Ohlerking, F., Christenson, G.A. & Miner, M. (1999). Psychiatric comorbidity in pedophilic sex offenders. *American Journal of Psychiatry*, 156 (5), 786 – 788.

Reed, S. D. (1993). Subcultural Report: Effects of examiner's and examinee's race on Psychophysiological detection of deception outcome accuracy. Report No. DoDP194-R-0012. Ft. McClellan, AL. U.S. Department of Defense Polygraph Institute.

Reed, S.D. (1993). Effect of demographic variables on psychophysiological detection of deception outcome accuracies. Report No. DoDPI93-R-0007. Ft. McClellan, AL: U.S. Department of Defense Polygraph Institute.

Reid, J. (1947). A revised questioning technique in lie-detection tests, *Journal of Criminal Law and Criminology*, 37, 542 – 547.

Reid, J. & Arther, R.O. (1953). Behaviour symptoms of lie detector subjects. *Journal of Criminal law and Criminology*, 44, 104 – 108.

Reid, J.E. & Inbau, F.E. (1977). *Truth and Deception*, 2<sup>nd</sup> Ed. Baltimore: Williams & Wilkins.

Reid, C., Lachs, M.S. & Feinstein, A.R. (1995). Use of methodological standards in diagnostic test research. *Journal of the American Medical Academy*. 274 (8), 645 – 651.

Riggio, R.E. & Friedman, H.S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology*, 45 (4), 899 – 915.

Roese, N.J. & Jamieson, D.W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, 114, 363 – 375.

Rogers, R. & Dickey, R. (1991). Denial and minimization among sex offenders: A



review of competing models of deception. *Annals of Sex Research*, 4, 49 – 63.

Rosenfeld, J.P. (1995). Alternative views of Bashore and Rapp's (1993) alternatives to traditional polygraphy: A critique. *Psychological Bulletin*, 117 (1), 159 - 166.

Rosenfeld, J.P., Ellwanger, J.W., Nolan, K., Wu, S. & Bermann, R. (1999). P300 scalp amplitude distribution as an index of deception in simulated cognitive deficit model. *International Journal of Psychophysiology*, 33 (1), 3 – 20.

Rosenfeld, J.P., Reinhart, A.M., Bhatt, M., Willwanger, J., Gora, K., Sekera, M. & Sweet, J. (1998). P300 correlates of simulated malingered amnesia in a matching-to-sample task: topographic analyses of deceptive vs. truth-telling responses. *International Journal of Psychophysiology*, 28 (3), 233 – 249.

Ryan, J.J. & Paolo, A.M. (1992). A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist*, 6, 53 – 62.

Salters, A. C. (1995). *Transforming Trauma. A Guide to Understanding and Treating Adult Survivors of Child Abuse*. California: Sage Publications.

Salters, A.C. (1997). 'Special issues in Sexual Offender and Victim Treatment', (conference presentation). *Sex Offenders – Assessment, Risk Management and Treatment*. San Diego, California: Specialised Training Services.

Samuels, J., Nestadt, G., Beinvenu, O.J., Costa, P.T., Riddle, M.A., Liang, K., Hoehn-Saric, R., Grados, M.A., & Cullen, B.N. (2000). Personality disorders and normal personality dimensions in obsessive-compulsive disorder. *British Journal of Psychiatry*, 177, 457 – 462.

Sanderson, C. & Clarkin, J.F. (1994). Use of the NEO-PI personality dimensions in differential treatment planning. In P.T.Costa & T.A.Widiger (Eds.). *Personality Disorders and the five-factor model of personality* (pp 219 – 235). Washington: American Psychological Association.

Santtila, P., Alkiora, P., Ekholm, M. & Niemi, P. (1999). False confession to robbery: the roles of suggestibility, anxiety, memory disturbance and withdrawal symptoms. *The Journal of Forensic Psychiatry*, 10 (2), 399 - 415.

Saulsman, L.M. & Page, A.C. (2003). Can trait measures diagnose personality disorder? *Current Opinions in Psychiatry*, 16, 83 – 88.

Saulsman, L.M. & Page, A.C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, 23, 1055 – 1085.

Saxe, L. & Ben-Shakhar, G. (1999). Admissibility of polygraph tests : The

application of scientific standards post-Daubert. *Psychology, Public Policy and Law*, 5, 203 – 223.

Saxe, L., Dougherty, D. & Cross, T. (1985). The validity of polygraph testing. *American Psychologist*, March, 355 – 366.

Schlank, A.M. & Shaw, T. (1996). Treating sexual offenders who deny their guilt: A pilot study. *Sexual Abuse: A Journal of Research and Treatment*, 8 (1), 17 -23.

Schroeder, M.L., Wormworth, J.A., & Livesley, W.J. (1994). Dimensions of personality disorder and the five-factor model of personality. In P.T.Costa & T.A.Widiger (Eds.). *Personality Disorders and the five-factor model of personality* (pp 117 – 127). Washington: American Psychological Association.

Segrave, K. (2004). *Lie Detectors: A social history*. McFarland & Company: North Carolina.

Seymour, T.L., Seifert, C., Shafto, M.G. & Mossman, A.L. (2000). Using response time measures to assess “guilty knowledge”. *Journal of Applied Psychology*, 85 (1), 30 – 37.

Shea, W.M. (1996). Personality characteristics of child molesters, non-sex



offending criminal child abuse controls, and normals as differentiated by the Million Clinical Multiaxial Inventory – II. *Dissertations Abstracts International*, 56 (9-B), 5184.

Sigurdsson, J.F. & Gudjonsson, G.H. (1996a). The psychological characteristics of ‘false confessors’. A study among Icelandic prison inmates and juvenile offenders. *Personality and Individual Differences*, 20 (3), 321 – 329.

Sigurdsson, J.F. & Gudjonsson, G.H. (1996b). The relationship between types of claimed false confession and the reasons why suspects confess to the police according to Gudjonsson Confession Questionnaire (GCQ). *Legal and Criminological Psychology*, 1, 259 – 269.

Sigurdsson, J.F. & Gudjonsson, G.H. (2001). False confessions: the relative importance of psychological, criminological and substance abuse variables. *Psychology, Crime and Law*, 7, 275 - 289.

Simmons, G.R. (1998). Survey of polygraph testing of convicted sexual offenders for monitoring purposes in the United States. Unpublished study.

Sivley, J.O. (1998). A comparison of personality characteristics between juvenile sex offenders and delinquent non-sex offending adolescents. *Dissertation Abstracts International*. 58 (11 – B), 6247.

Slowik, S.M. & Buckley, J.P. (1975). Relative accuracy of polygraph examiner

diagnosis of respiration, blood pressure, and GSR recordings. *Journal of Police Science and Administration*, 3, 305 – 309.

Smith, A.D. & Taylor, P.J. (1999). Serious sex offending against women by men with schizophrenia. *British Journal of Psychiatry*, 174, 233 - 237.

Sokolov, A.N. (1963). *Perception and the Condition Reflex*. Oxford: Pergamon Press.

Spencer, H. (1998). Lie testing upheld for sex offenders, retrieved from [www.seattletimes.com](http://www.seattletimes.com), June 5<sup>th</sup>.

Spren, O. & Strauss, E. (1998). *A Compendium of Neuropsychological Tests* (2<sup>nd</sup> Ed.). Oxford University: Oxford.

Steinbrook, R. (1992). The polygraph test - A flawed diagnostic method. *The New England Journal of Medicine*, 327 (2), 122 - 123.

Stern, R.M., Breen, J.P., Watanabe, B.T. & Perry, B.S. (1981). Effects of feedback of physiological information on responses to innocent associations and guilty knowledge. *Journal of Applied Psychology*. 66 (6), 677 – 681.

Stough, C., Donaldson, C., Scarlata, B. & Ciorciari, J. (2001). Psychophysiological

correlates of the NEO PI-R Openness, Agreeableness and  
Conscientiousness: preliminary results. *International Journal of  
Psychophysiology*, 41, 87 – 91.

Sullivan, E. (2001). *The Concise Book of Lying*. Picador: New York.

Swets, J.A. (1996). *Signal Detection Theory and the ROC Analysis in Psychology  
and Diagnostics*. Mahwah, NJ: Erlbaum.

Swets, J.A., Dawes, R.M. & Monahan, J. (2000). Better decisions through science.  
*Scientific American*, October, 70 – 75.

Szucko, J.J. & Kleinmuntz, B. (1984). Statistical versus clinical lie detection  
reconsidered. *American Psychologist*, 80.

Tabachnick, B.G. & Fidell, L.S. (2000). *Using Multivariate Statistics* 4<sup>th</sup> ed. Allyn  
& Bacon: London.

Trost, K.K., Herbst, J.H., Masters, H.L., & Costa, P. (2001). Personality  
pathways to unsafe sex: personality, condom Use, and HIV risk behaviors.  
*Journal of Research in Personality*, 36 (2), 117-133.

Teuscher, T. (1978). The polygraph and probation. *Polygraph*, 7, 1 – 4.

Trovillo, P.V. (1939). A history of lie detection. *American Journal of Police*



*Science*, 29, 848- 881.

Trull, T.J. (1992). DSM-III-R personality disorders and the five-factor model of personality: An empirical comparison. *Journal of Abnormal Psychology*, 101, 553- 560.

Toris, C. & DePaulo, B.M. (1985). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology*, 47 (5), 1063 – 1073.

United States Congress, House Committee on Government Operations,  
Subcommittee on Government Information and Foreign Operations.  
(1965). *Use of polygraphs as "lie detectors" by the federal government*  
(house report No. 198). Washington, DC: U.S. Government Printing  
Office.

US Congress. (1983). *Scientific Validity of Polygraph Testing: A Research and  
Evaluation – A Technical Memorandum*. Washington, DC: US Government  
Printing Office.

Vendemia, J.M.C. (2003). Hobson's choice: the relationship between  
consequences and the comparison question. Unpublished paper.

Vrij, A. (2000). *Detecting Lies and Deceit*. Chichester, England: Wiley.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve.

*Legal and Criminological*, 9, 159 –181.

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual movements and hand movements during deception. *Journal of Nonverbal Behavior*, 21 (2), 87 – 101.

Vrij, A. & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: truth bias, lie bias and individual differences. *Expert Evidence*, 7: 25 – 36.

Waid, W.M. & Orne, M.T. (1982). The physiological detection of deception. *American Scientist*, 70, 402 – 409.

Waid, W.M., Orne, E.C. & Orne, M.T. (1981). Selective memory for social information, alertness, and physiological arousal in the detection of deception. *Journal of Applied Psychology*, 56 (2), 224 – 232.

Waid, W.M., Orne, M.T. & Wilson, S.K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology*, 16, 15 – 22.

Ward, T., Hudson, S.M., & Marshall, W.L. (1995). Cognitive distortions and affective deficits in sex offenders: A cognitive deconstructionist interpretation. *Sexual Abuse: A Journal of Research and Treatment*, 7 (1), 67 – 83.

- West, D. (2000). Paedophilia: plague or panic? *The Journal of Forensic Psychiatry*, 11 (3), 511 - 531.
- Whiten, A. & Bryne, R. (1988). *Machiavellian Intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Clarendon Press: Cambridge, UK.
- Wicklander, D. E., and Hunter, F. L. (1975). The Influence of Auxiliary Sources of Information in Polygraph Diagnoses. *Journal of Police Science and Administration*, 208 (3), 405-409.
- Widiger, T.A. (2005). A dimensional model of personality disorder. *Current Opinions in Psychiatry*, 18, 41 – 43.
- Widiger, T.A. & Costa, P.T. Jr. (1994). Personality and personality disorders. *Journal of Abnormal Psychology*, 103 (1), 78 –91.
- Widiger, T.A. & Costa, P.T. (2002). Five-factor model personality disorder research. In: P.T. Costa & T.A. Widiger (Eds.), *Personality disorders and the five-factor model of personality*. (2<sup>nd</sup> ed., pp. 59 – 87), Washington, DC: American Psychological Association.
- Widiger, T.A., Trull, T.J., Clarkin, J.F., Sanderson, C.J. & Costa, P.T. (2002). A



description of the DSM-IV personality disorder with the five-factor model of personality. In P.T. Costa & T.A. Widiger (Eds.). *Personality disorders and the five-factor model of personality* (pp. 89 – 102). Washington, DC: American Psychological Association.

Widiger, T.A. & Trull, T.J. (1992). Personality and psychopathology: An application of the five-factor model. *Journal of Personality*, 60 (2), 363 – 393.

Wilberg, T., Urnes, Ø., Friis, S., Pedersen, G. & Karterud, S. (1999). Borderline and avoidant personality disorders and the five-factor model of personality: A comparison between DSM-IV diagnoses and the NEO-PI-R. *Journal of Personality Disorders*, 13 (3), 226 – 240.

Wilcox, D. (2000). Application of the clinical polygraph examination to the assessment, treatment and monitoring of sex offenders. *The Journal of Sexual Aggression*, 5 (2), 134 – 152.

Wilcox, D. & Sosnowski, D. (2001). Polygraph examinations of British sexual offenders: A pilot study on sexual history disclosure testing. Unpublished study.

Wiley, S.D. (1998). Deception and detection in psychiatric diagnosis. *The Psychiatric Clinics of North America*, 21 (4), 869 – 893.

- Williams, V.L. (1995). Response to Cross and Saxe's "A critique of the validity of polygraph testing in children sexual abuse cases". *Journal of Child Sexual Abuse*, 4 (3), 55 – 71.
- Wilson, G. & Cox, D.H. (1983). Personality of paedophile club members. *Personality and Individual Differences*. 4(3), 323 – 329.
- Wise, T.N., Fagan, P.J., Schmidt, C.W., Ponticas, Y. & Costa, P.T. (1991). Personality and sexual functioning of transvestitic fetishists and other paraphilics. *The Journal of Nervous and Mental Disease*, 179 (11), 694 – 698.
- Wiseman, R. (1996). Towards a psychology of deception. *The Psychologist*, 61 – 64.
- Zuckerman, M., Spiegel, N.H., DePaulo, B.M. & Rosenthal, R. (1982). Nonverbal strategies for decoding deception. *Journal of Nonverbal Behavior*, 6 (3), 171 – 187.

## APPENDICES



Appendix A

Comparison Question Techniques

### *The Reid Comparison Question Test*

The Reid CQT was the earliest form of comparison question test (Reid, 1947). This testing protocol typically includes an extensive pre-test interview about the issues under investigation, which provides a basis for introducing and reviewing the relevant questions. Four relevant questions are used which encompass different levels or aspects of the event being investigated. This is followed by an in-depth discussion of the background and moral values of the examinee. The latter is designed to establish the context for introducing the comparison questions and arousing concerns that will lead the subject to deny any of the acts embodied in the comparison questions. Two comparison questions are used in the Reid technique and are formulated from the same category as the relevant issue, for instance in a theft case a comparison question might be “Have you ever stolen from someone that trusted you?” Table 27 provides an example of a Reid CQT question format. In the evaluation of charts, the Reid examiner performs a global evaluation of the recordings (Office for Technology Assessment, 1983). This includes considering other information such as the subject behaviour, case reports or other sources of information about the individual or case (Matte, 1996; Raskin & Honts, 2002).

Table 29.

An example of the Reid CQT question sequence

Question	Question Type
1. Is today Wednesday?	Irrelevant
2. Are you sitting down?	Irrelevant
3. Did you rob the supermarket last night?	Relevant
4. Do you sometimes watch TV?	Irrelevant
5. Did you use a gun to rob the TV?	Relevant
6. Have you ever stolen anything from someone that trusted you?	Comparison
7. Is your name John?	Irrelevant
8. Did you take money from the supermarket?	Relevant
9. Did you drive the getaway car from the supermarket?	Relevant
10. Have you ever cheated anyone?	Comparison

*The Backster Zone Comparison Test*

The Backster Zone Comparison Test is so named because of the three “zones” or blocks of time during the test. These are: the relevant questions (red zone); the probable-lie comparison questions (green zone); and other questions (black zone). Black zone questions are included to reveal issues not covered by the green or red zones (Backster, 1963a; 1963b). Backster added these ‘outside issue’ questions because he was concerned that such issues could dominate the subject’s concern and, thereby, reduce reactivity to the relevant questions. Backster described this effect as the ‘outside super-dampening effect’ (Matte, 1996). Two relevant and two comparison questions are used in Backster’s test. Backster also added a ‘sacrificial relevant question’. This is a weak overall question such as, “regarding whether you robbed the supermarket last night, do



you intend to answer truthfully?” This question is placed first in the question sequence and is expected to absorb the general impact of being confronted with the relevant issue. The examinees reaction to this question is not included in the evaluation of the test outcome (Raskin & Honts, 2002). Backster’s comparison questions are formulated from the same category of the offence as the relevant issue (Matte, 1996). In the evaluation of the charts Backster developed the previously described numerical scoring system in the evaluation of the charts.

## Appendix B

### Study 1 Information Sheets & Consent Form

[This is the information sheet for the control subjects, i.e. those who are not expecting the polygraph]

## **Information Sheet**

You have been invited to take part in research being carried out by the Department of Psychiatry in the University of Newcastle. The aim of the study is to help improve the way in which men with convictions for sex offences are supervised in the community in order to reduce the risks of their re-offending.

We are interested in finding out how easy you find it to stick to the conditions of your Order and to your relapse prevention plan. If you agree to take part in the research you will be seen on 3 occasions over 6 months, during which you will be asked about your risk behaviours and your relapse prevention plan. There are also a number of questionnaires you will need to fill in. At some stage during the research you may be asked to take a polygraph examination. If you are asked, this test will only focus upon risk behaviours and NOT actual offences.

The information you provide about yourself will be confidential, except where it relates directly to your supervision. We will also have to disclose details of any offences that you tell us about which are not already known.

**Participation in this research is completely voluntary, and you may withdraw from it at any time without consequence.**

Ethical approval for this study has been gained through the Northumberland, Tyne and Wear Local Research Ethics Committee. The National Probation Service has accepted this committee's decision for the research to be conducted in the probation service areas outside Northumberland.

If you have any questions about this research, you can contact:



## Information Sheet

You have been invited to take part in research being carried out by the Department of Psychiatry in the University of Newcastle. The aim of the study is to help improve the way in which men with convictions for sex offences are supervised in the community in order to reduce the risks of their re-offending.

We are interested in finding out how easy you find it to avoid risk behaviours. We want to test whether the use of the polygraph can help with this.

If you agree to take part in the research you will be seen on 3 occasions over 6 months, during which you will be asked about your identified risk behaviours. There are also a number of questionnaires you will need to fill in.

The polygraph will be used in the second interview, and possibly at the third. The polygraph examinations will relate only to your risk behaviours. You will NOT be asked anything about actual current or previous sex offences.

The information you provide about yourself will be confidential, except where it relates directly to your supervision. So if you disclose engaging in risk behaviour, then this will be reported to your probation officer or treatment provider.

We will also have to disclose details of any offences that you tell us about which are not already known. WE WILL NOT ASK YOU ABOUT ANY SEXUAL OFFENCES CURRENT OR HISTORICAL. Results of the polygraph examinations may be shared with your probation officer to assist in your treatment programme.

Participation in this research is completely voluntary, and you may withdraw from it at any time without consequence.

Ethical approval for this study has been gained through the Northumberland, Tyne and Wear Local Research Ethics Committee. The National Probation Service has accepted this committee's decision for the research to be conducted in the probation service areas outside Northumberland.

If you have any questions about this research, you can contact:

CONSENT FORM

I have read the information sheet that describes the nature of this research, and I have had an opportunity to ask questions about the study.

I understand that participation in this research is completely voluntary, that I do not have to participate in it, and that I can withdraw from it at any time without consequence.

I have been told that the information I provide about myself will be confidential, except where it relates directly to my supervision. I understand that any information about specific offences not already known will be disclosed.

Name: \_\_\_\_\_

Location: \_\_\_\_\_

\_\_\_\_\_/\_\_\_\_\_/2001  
(Signature – Participant)

\_\_\_\_\_  
(Witness)

## Appendix C

### Study 1 Sample Polygraph Report



## POLYGRAPH TEST REPORT (Example)

Subject X

John

### SPECIFIC BEHAVIORS MONITORED

- Try to set up situation in order to be around children under the age of 16 years.
- Being alone and unsupervised with children under the age of 16 years.
- Masturbate to fantasies that involve children under the age of 16 years.
- Working around children under the age of 16 years.

### PRETEST ADMISSIONS

- Fantasies & Masturbation:
  - Reported masturbating to fantasies 10-12 times since initial interview – fantasies involved females and males between the ages of 10 and 14 years.
- Situations with children: Applied for job at local fast food pizza place. Also reported applying for work at a Rest Home and a Day Care Centre.
- Unsupervised contact with minors: Only contact with children under the age of 16 years was on train.

### FAILED TEST (DECEPTION INDICATED) – POST TEST ADMISSIONS

- Reported deliberately and for sexual purposes brushing against female and male children whilst travelling on train. Reported that he masturbating to these memories 40-50 times.

## APPENDIX D

### Study 1 Questionnaire

THE INFORMATION THAT YOU PROVIDE IS COMPLETELY CONFIDENTIAL. YOUR RESPONSES TO THESE QUESTIONS WILL NOT BE SHARED WITH YOUR PROBATION OFFICER OR TREATMENT

Please answer the following questions:

1. Do you think that the polygraph could be helpful in assisting you with remaining offence-free in the community?

Not at all	Very Little	Probably	Very Much	Yes, absolutely
------------	-------------	----------	-----------	-----------------

2. Did you report any information during the research about risk behaviours that you had not reported elsewhere?

YES	NO
-----	----

If yes, was this because?

I failed the polygraph	Interviewers asked me	I knew that I would fail polygraph so I told
------------------------	-----------------------	--

Are there any other reasons why you reported this information?


3. Did knowing that you were going to be polygraphed during the research assist you with avoiding behaviours or situations that are considered high risk for you?

No did not think about it	Very Little	Sometimes	Most of the time	All the time
---------------------------	-------------	-----------	------------------	--------------

4. Did knowing you were going to be polygraphed lead you to report more about high-risk behaviours to your supervisor or therapist that you otherwise may not have?

No	Very Little	Sometimes	Mostly	Yes, very much
----	-------------	-----------	--------	----------------



5. Did you try to lie or hide information about high-risk behaviours during any of the polygraph tests?

YES	NO
-----	----

5.B. Did the polygraph accurately identify this?

YES	NO
-----	----

## APPENDIX E

### Study 2 Information Sheet and Consent Form

## **INFORMATION SHEET & CONSENT FORM**

### **Study title**

The accuracy of the polygraph in maintenance examinations.

### **Invitation**

You have been invited to take part in research being carried out by the University of Newcastle in England. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information sheet carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

### **What is the purpose of the study?**

The aim of the study is to help improve the way in which men with convictions for sex offences are supervised in the community in order to reduce the risks of their re-offending. The study investigates personality characteristics associated with sexual offending and also with successful participating in community supervision. The study is also investigating the accuracy of the polygraph instrument, and how personality characteristics can affect this.

### **Why have I been chosen?**

Approximately 300 individuals will take part in the study throughout Georgia. You have been chosen because you have been convicted of a sexual offence and are required to take periodic polygraph examinations.

### **Do I have to take part?**

It is entirely up to you to decide whether or not you decide to take part. If you do decide to take part you will be asked to sign a consent form. The consent sheet will outline the details of the study and the conditions of confidentiality. If you choose not to take part this will have no effect on your current treatment.

If you do decide to take part you can withdraw from the study at any time and without giving a reason. Again, this will not affect your treatment in any way.

### **What will happen to me if I take part?**

If you agree to take part in the research you will be seen on a single occasion by the researcher for up to 30 minutes. During this meeting you will complete a short 5 minute test (National Adult Reading Test; NART-2) and there will be a questionnaire (NEO – Personality Inventory revised; NEO-PI-R) for you to fill out which can take up to 25 minutes. We will also gather some information from your professional supervision file about the details of your previous convictions.



You may have an additional question about drug use during your polygraph examination. You may also be asked to supply a hair sample for analysis after your regular polygraph test. However, the results of any hair analysis would be confidential and under no circumstances will be shared with your probation officers or treatment providers.

In some cases we will have the polygraph charts scored by other examiners, if so, all identifying information will be removed from them.

### **What do I have to do?**

If you agree to take part all you need to do is arrange a convenient time to meet with the researcher and complete the questionnaire.

### **What are the possible disadvantages and risks of taking part?**

The possible disadvantages for you of taking part in this research will be the loss of your time to complete the questionnaire and interview. Another disadvantage may be that you have a drug question on your polygraph test.

### **What are the possible benefits of taking part?**

The information we get from this study may help us to improve the treatment and supervision of individuals convicted of sexual offending.

When you complete the psychometric testing and provide a hair sample, if requested, you will receive a \$ 25 discount on the polygraph examination, which is used in the research.

### **Will my taking part in this study be kept confidential?**

All the information you provide about yourself will be confidential and stored anonymously. All the results of the questionnaires and possible hair analysis will also be completely confidential. That is, this information will NOT be shared with your probation officers or treatment providers.

### **What will happen to the results of the research study?**

The study is expected to be completed in October 2003. The results of the research will be used in a doctoral study and will be published in various psychological journals. It will not be possible to identify anyone who takes part in any of the publications.

**Who is organising and funding the research?**

The research is being conducted by the Department of Psychiatry at the University of Newcastle in England. The study is being funded by the Forensic Mental Health Research and Development Programme Fellowship in England.

**Contact for Further Information**

THANK YOU FOR READING THIS

# CONSENT FORM

**Title of Project:**

The accuracy of the polygraph in maintenance examinations.

**Name of Investigators:**

LARS MADSEN; DON GRUBIN & SHAUN PARSONS

Please initial box

1. I confirm that I have read and understand the information sheet dated 13 / 01 / 2003 for the above study and have had the opportunity to ask questions of a researcher.

☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

☐
3. I understand that I will be required to complete the National Adult Reading Test (NART-2) and the NEO – Personality Inventory (NEO-PI-R).

☐
4. I understand that I may be required to supply a hair sample and that the polygraph exam may include a question about drug use. I give permission for individuals from the University of Newcastle to have access to my probation file or appropriate records and polygraph results.

☐
5. I understand that by completing the psychometric testing and providing a hair sample, if requested, I am entitled to a \$ 25.00 rebate on the polygraph examination that will be used in the research.

☐
6. I agree to take part in the above study.

\_\_\_\_\_  
Name of Participant

\_\_\_\_\_  
Date  
Signature

\_\_\_\_\_  
Name of Person taking consent

\_\_\_\_\_  
Date  
Signature

## APPENDIX F

### Study 2 Questionnaire



THE INFORMATION THAT YOU PROVIDE IS COMPLETELY CONFIDENTIAL. YOUR RESPONSES TO THESE QUESTIONS WILL NOT BE SHARED WITH YOUR PROBATION OFFICER, TREATMENT PROVIDER OR POLYGRAPH EXAMINER.

THESE QUESTIONS REFER TO POLYGRAPH EXAMINATIONS THAT YOU HAVE HAD WHILE ON SUPERVISION.

1. Have you ever failed a polygraph exam (been accused of lying) when you were telling the truth? *(that is, the polygraph got it wrong?)*

NO (go to question 2)                      YES

If YES, how many times have you been wrongly accused of lying? \_\_\_\_

a. At these times, when you were telling the truth and accused of lying have you ever reported engaging in behaviour that you had not engaged in?

NO                      YES (Go to question 2.a)

2. Have you ever reported false information during a polygraph examination? *(that is, information about your behaviour that you knew was not true?)*

NO                      YES

a. If YES, what false information did you report?

---

---

---

a. What were your reasons for reporting false information?  
(Circle relevant reasons)

- 1) I would get in trouble with my supervisors or therapist if I failed and did not report anything
- 2) I was confused
- 3) I wanted to give a good impression to the polygraphers
- 4) Wanted to ensure that I passed the polygraph test
- 5) I felt pressured to report something
- 6) I wanted to demonstrate that I was committed to the treatment
- 7) Other reason/s:

---

---

3. Have you ever passed a polygraph exam when you knew that you were lying? (that is, not answering truthfully?)

NO

YES

If YES, how many times have you done this? \_\_\_\_\_

4. Have you ever deliberately used drugs, physical or mental strategies to pass a polygraph exam?

NO

YES

If YES, how many times have you done this? \_\_\_\_\_

Which strategies have you used (tick)?

- Drugs (e.g. tranquillisers) \_\_\_\_\_

- Physical:
  - Controlling breathing \_\_\_\_\_
  - Moving during the test \_\_\_\_\_
  - Tensing muscles \_\_\_\_\_
  - Relaxing muscles \_\_\_\_\_
  - Self- hypnosis \_\_\_\_\_
- Mental:
  - Controlling thoughts \_\_\_\_\_
  - Disassociating \_\_\_\_\_

*INFORMATION ABOUT THE POLYGRAPH’S USEFULNESS*

5. How helpful is the polygraph for you with avoiding high-risk behaviours and situations?

1	2	3	4	5
None	Minimal	Moderately	Quite	Extremely

6. How helpful is the polygraph for you with avoiding re-offending?

1	2	3	4	5
None	Minimal	Moderately	Quite	Extremely

7. Overall, how helpful is the polygraph in your treatment?

1	2	3	4	5
None	Minimal	Moderately	Quite	Extremely

8. Identify which, if any, of the following behaviours decreased because due to the use of the polygraph:

\_\_\_\_\_ Masturbation using deviant or inappropriate fantasies

\_\_\_\_\_ Drug usage

\_\_\_\_\_ Alcohol use

\_\_\_\_\_ Contact with children and/or vulnerable adults

\_\_\_\_\_ Visiting places to view children (e.g. arcades, gyms, and swimming pools)

- \_\_\_\_\_ Collecting pictures of children from clothing catalogues for masturbation purposes
- \_\_\_\_\_ Missing group meetings / not completing home work tasks
- \_\_\_\_\_ Probation / supervision violations
- \_\_\_\_\_ Use of pornography
- \_\_\_\_\_ Visiting adult book stores
- \_\_\_\_\_ Use of internet (e.g. visiting chat rooms; pornography sites)
- \_\_\_\_\_ Engaging in extra-marital affairs or sexual relationships outside main relationship
- \_\_\_\_\_ Use of prostitutes
- \_\_\_\_\_ Use of telephone sex lines
- \_\_\_\_\_ Other identified risk behaviours

9. Because of having to complete regular polygraph tests are you more or less honest with your treatment provider or probation officer about your behaviour?

1 Less truthful	2 No Change	3 More truthful
--------------------	----------------	--------------------

10. Because of having to complete regular polygraph tests are you more or less likely to report information about your offence and offence-related behaviour to your supervisor or treatment provider?

1 Less likely to disclose	2 No Change	3 More likely to disclose
------------------------------	----------------	------------------------------

11. Because of having to complete regular polygraph tests are you more or less truthful with your family and friends?

1 Less truthful	2 No Change	3 More truthful
--------------------	----------------	--------------------



*INFORMATION ABOUT THE ACCURACY OF THE POLYGRAPH*

12.How accurate do you believe the polygraph is?

1 Not accurate	2 Slightly	3 Moderately	4 Quite	5 Extremely accurate
-------------------	---------------	-----------------	------------	----------------------------

## APPENDIX G

### The relationship between the five-factor model and personality disorder

Examining the correlations offers some insight into the shared FFM features of the different DSM PDs. Excluding Antisocial PD, all were positively correlated with Neuroticism, suggesting that these share the core quality of emotional maladjustment. Again this appears theoretically consistent, as a tendency to experience negative affectivity is not a feature of Antisocial PD. These findings are consistent with other research (e.g. Widiger & Costa, 2002). In the other domains, Extraversion was negatively associated with Avoidant, Depressive and Schizotypal PDs, indicating that these share features of social withdrawal. Agreeableness was negatively correlated with Passive-aggressive, Depressive, Paranoid, Schizotypal, Histrionic, Narcissistic, Borderline and Antisocial PDs. This suggests that these PDs share antagonistic features, such as uncooperativeness, vindictiveness and lack of trust, which again is consistent with theoretical conceptualizations (American Psychiatric Association, 1994). Finally, the Conscientiousness domain was negatively correlated with the Paranoid PD, indicating these individuals are likely to be less scrupulous and reliable. Taken together these findings indicate that most of the PDs shared common FFM characteristics. This suggests that the various PDs may cluster together to the degree that they represent the same traits, which may explain the problems with comorbidity.

The inability of the FFM to distinguish between different PD categories is a limitation noted in the literature. There are perhaps a number of reasons why the FFM may struggle to do this. Firstly, this study did not examine the relationship between all the facets scores and PD. Widiger, Trull, Clarkin, Sanderson and Costa (2002) have argued that the diagnostic utility of the FFM will come when

more is known about how the facets that underlie each of the domains relate to specific PDs. They contend, for instance, that Histrionic PD is best captured by high scores on the following facets: depression, self-consciousness, warmth, gregariousness, excitement seeking, positive emotions, openness to fantasy and feelings and trust. The little research that have explored the facet characteristics, suggest that the relations between the FFM and PDs are both stronger and clearer when these are considered (e.g. Trull, et al., 2003; Miller, Reynolds & Pilkonis, 2004).

Of course another possibility is that extreme and maladaptive variants of the traits believed to compose PD are not adequately represented in measures of the FFM. Haigler and Widiger (2001) showed that the NEO-PI-R is heavily weighted, in terms of social desirability, towards certain poles of each of the five domains. They found that only 17% of the high Agreeableness and 10% of the high conscientiousness items described behaviour that would be characterized as maladaptive or dysfunctional. This finding suggests that the current measure of the FFM, the NEO-PI-R, is less well suited to assessing PDs thought to be related to high levels of agreeableness or conscientiousness, such as Obsessive-Compulsive PD.

Finally, it is perhaps debatable that the FFM would predict DSM PD. The limitations of these categories have been well documented and include an inadequate scientific base and comorbidity. Indeed, some researchers have argued against the idea of considering PD as discreet categories of pathology (Widiger, Trull, Clarkin, Sanderson & Costa, 2002). It would seem, however, that research



has endeavored to establish the value of the FFM by comparing it to same categorical system, which has been so heavily criticized and disputed (Livesley, 2001).

There are several limitations that need to be noted in regards to these conclusions that should be acknowledged. Firstly, the sample size was reasonably small, a larger sample may have yielded different findings. Similarly, it needs to be borne in mind that only 43% of the sex offenders approached agreed to take part. It could be argued that the individuals that volunteered might have different personality traits from those that did not. Secondly, it is possible that low or high prevalence rates for certain PDs may have affected the results. Several disorders were infrequent in this sample, and it is conceivable that additional or different associations between the FFM domains and PD might have emerged in another sample of sex offenders. Replication of the present results is, therefore, necessary. Finally, the study is also limited by its reliance on the self-report measures for assessing FFM traits and PD symptomatology. Future research in this area might include informant versions of both NEO and SCID to assess whether the present results are replicated.