



The
British
Psychological
Society

A review of the current scientific status and fields of application of Polygraphic Deception Detection

Final report (6 October 2004) from the BPS Working Party

A review of the current scientific status and fields of application of Polygraphic Deception Detection

Final report (6 October 2004) from the BPS Working Party

Members of the Working Party

Professor Ray Bull (Chair)
Dr Helen Baron
Professor Gisli Gudjonsson
Professor Sarah Hampson
Professor Gina Rippon
Professor Aldert Vrij

Contents

- 7 Executive summary
- 8 Introduction
 - What is the polygraph?
 - What constitutes a good psychometric procedure?
- 10 Polygraph testing in criminal investigations
 - The Relevant/Irrelevant Technique
 - The Control Question Test
 - The Directed Lie Test
 - The Guilty Knowledge Test
 - Accuracy (validity) of polygraph tests
 - Countermeasures
 - Base rates
- 18 The use of the polygraph in employment and security screening
- 19 The use of the polygraph in the clinical setting
- 22 Other possible methods to detect deception
 - Event-related brain potentials
 - Functional brain imaging
 - Behaviour/demeanour
 - Linguistic analysis
 - Voice stress analysis
 - Evaluation of different approaches to measuring honesty and integrity
 - Faking
 - Situational influences on honesty
 - Combining methods
- 27 Human rights and codes of conduct
- 29 Main conclusions
- 31 References

Executive summary

- This document presents a review of the current scientific status and fields of application of polygraphic lie detection.
- It presents an overview of:
 - What is the polygraph?
 - What constitutes a good psychometric procedure?
 - Polygraph testing in criminal investigations.
 - Polygraph testing in employment and security screening.
 - The use of the polygraph in the clinical setting.
 - Other possible methods to detect deception.
 - Human rights and codes of conduct.
- The polygraph is a set of equipment that accurately measures various sorts of bodily activity such as heart rate, blood pressure, respiration, and palmar sweating. In recent years brain activity has also begun to be measured in this setting. This bodily (and brain) activity can be displayed via ink writing pens on to charts or via a computer's visual display unit.
- A good psychometric procedure should demonstrate reliability and validity.
- Most published research on polygraphic deception detection has been concerned with its possible use in criminal investigations. The results of better quality research studies demonstrate that while the correct classification of deceivers can sometimes be fairly high, incorrect decisions about who is or is not being deceptive occur at rates that are far from negligible.
- Use of countermeasures may well result in deceivers not being detected.
- Use of the polygraph in employment and security screening is not justified by the available research evidence.
- Use of the polygraph in the clinical setting, with specific reference to its use with sex offenders, has received too little research attention.
- More research is needed on other possible methods to detect deception, honesty and integrity.
- Use of the polygraph in attempts to detect deception raises issues concerning human rights and professional codes of conduct.
- Over confidence in the ability of any procedure designed to detect deception can have serious consequences, especially if the deceivers are few among many non-deceivers.

Introduction

In 1986 the British Psychological Society (BPS) published the report of its first working group on the use of the polygraph in criminal investigation and personnel screening. Since then a considerable further amount of research has been published on this general topic. In 2003 the BPS decided to convene another working group to produce this report. Members of the BPS who may be in a position to consider the use of the polygraph in attempts to detect deception should make themselves fully aware of the contents of this report.

Similar developments have taken place in the USA. In its 2003 review of the scientific evidence on the polygraph and lie detection, the National Research Council of the National Academies (NRC) stated that: 'The polygraph continues to be the subject of a great deal of scientific and public controversy in the United States' (p.12).

The 2003 NRC review noted that in the USA the 1983 report by the Office of Technology Assessment on the validity of polygraphic lie detection had raised many criticisms that were still being raised 20 years later and that the 1988 Employee Polygraph Protection Act had reduced considerably the use of polygraphy in employment settings.

This report overviews:

- What is the polygraph?
- What constitutes a good psychometric procedure?
- Polygraph testing in criminal investigations.
- The use of the polygraph in employment and security screening.
- The use of the polygraph in the clinical setting.
- Other possible methods to detect deception.
- Human rights and codes of conduct.

The report ends with a concluding section.

What is the polygraph?

Throughout history (Bull, 1988) it has often been assumed that lying is accompanied by a change in the body's physiological activity. The polygraph is a set of equipment that accurately measures various sorts of bodily activity such as heart rate, blood pressure, respiration, and palmar sweating. In recent years brain activity has also begun to be measured in this setting. This bodily (and brain) activity can be displayed via ink writing pens on to charts or via a computer's visual display unit.

The polygraph (from the Greek 'poly' = 'many', and 'graph' = 'to write') is widely used by analytical staff in a variety of medical and scientific settings for purposes other than lie detection.

In lie detection situations its use is based on the premise that lying is accompanied by changes in the activity measured by the polygraph.

What constitutes a good psychometric procedure?

One of the major topics that psychologists and others have focused on across the decades is how best to determine if a testing procedure can be relied upon. Obviously, many issues are involved in this, but the most important ones include validity and reliability.

Validity

Validity concerns the extent to which a procedure measures what it is claimed to measure, and validity can have several components including:

- Face validity;
- Content validity;
- Predictive (or criterion) validity;
- Construct validity;
- Incremental validity.

Face validity refers to the extent to which a procedure 'on the face of it' appears to measure what it is claimed to measure.

Content validity concerns the relationship of the contents within a test to the construct being measured.

Predictive (or criterion) validity is the extent to which scores on the measure relate to or can predict outcomes. Thus, regarding polygraph testing, how accurate the procedure is at classifying people (e.g. as lying or truth telling). This typically has been the focus of polygraph validation work.

Construct validity is concerned with the relationship of the measure to underlying theory and constructs. This is a broad notion that includes the quality and extent of the (research) evidence that underpins such theory and concepts. Freedom from bias would generally be assessed as part of construct validity. (Criminal courts, for example, are now focusing more on this form of validity.)

Incremental validity is concerned with how well a procedure compares with other procedures designed to achieve the same ends (e.g. determine whether a person is being deceptive).

Reliability

Reliability, within psychology, 'refers not to the accuracy of evaluations but to their consistency across time or among examiners' (Carroll, 1988, p.20). Within psychology it is often the word 'validity' that is concerned with accuracy. A number of different forms of reliability are relevant to polygraphic lie detection.

One main form is '*inter-examiner*' *reliability* which concerns, for example, whether different examiners when looking at a person's physiological activity during responses to incriminating and non-incriminating questions come to the same conclusions.

'Intra-examiner' *reliability* is concerned with the extent to which the same examiner, on different occasions, comes to the same conclusion when looking at a record of a person's physiological activity.

'Test-re-test' *reliability* involves the extent to which when re-tested a person's activity leads examiners to the same decision on both occasions.

'Inter-item' *reliability* refers to the extent, for example, that different incriminating questions lead to the same conclusion and/or that observation of the different types of bodily activity during a test lead to the same conclusion.

It should be noted that a reliable procedure (in the senses described above) is not necessarily a valid procedure, for it might produce consistent but inaccurate results.

The European Federation of Psychologists' Associations has produced guidelines for the evaluation of psychological instruments that set standards for assessing their reliability, validity, and other properties (Bartram, 2002) and these have informed the contents of this report.

Polygraph testing in criminal investigations

The use of the polygraph in criminal investigations is described first because this is the use that has been most researched. A number of the points made in this section also apply to other types of use (which are described later in this report) but for brevity are not repeated in those other sections.

Polygraph tests are currently used in criminal investigations in many countries including Belgium, Canada, Israel, Japan, Turkey, Singapore, South Korea, Mexico, Pakistan, the Philippines, Taiwan, Thailand, and the USA (Lykken, 1998; Raskin, 1990; Vrij, 2000).

A polygraph is sometimes called a lie detector, but this term is misleading. A polygraph does not detect lies, but only arousal which is assumed to accompany telling a lie. Polygraph examiners have no other option than to measure deception in such an indirect way, as a pattern of physiological activity directly related to lying does not exist (Saxe, 1991). Three of the four most popular lie detection procedures using the polygraph (Relevant/Irrelevant Test, Control Question Test and Directed Lie Test, all discussed below) are built upon the premise that, while answering so-called 'relevant' questions, liars will be more aroused than while answering so-called 'control' questions, due to a fear of detection (fear of getting caught lying). This premise is somewhat naive as truth tellers may also be more aroused when answering the relevant questions, particularly: (i) when these relevant questions are emotion evoking questions (e.g. when an innocent man, suspected of murdering his beloved wife, is asked questions about his wife in a polygraph test, the memory of his late wife might re-awaken his strong feelings about her); and (ii) when the innocent examinee experiences fear, which may occur, for example, when the person is afraid that his or her honest answers will not be believed by the polygraph examiner. The other popular test (Guilty Knowledge Test, discussed below) is built upon the premise that guilty examinees will be more aroused concerning certain information due to different orienting reactions, that is, they will show enhanced orienting responses when recognising crucial details of a crime. This premise has strong support in psychophysiological research (Fiedler, Schmidt & Stahl, 2002).

Four different tests

Many different polygraph tests exist, but we will restrict ourselves to briefly describing the four most popular polygraph tests. (See Kleiner, 2002, and Vrij, 2000, for detailed descriptions of these tests.)

The Relevant/Irrelevant Technique

One of the oldest polygraph procedures is the Relevant/Irrelevant Technique (RIT), developed by Larson in 1932. In the RIT, two types of questions are asked, crime-relevant questions and crime-irrelevant questions. Crime-relevant questions are related to the crime under investigation, such as 'Did you steal the money from the company office last night?' All examinees (i.e. employees of the company), both innocent and guilty, will answer 'no' to this question, otherwise they admit having committed the crime. Irrelevant questions have nothing to do with the crime, and the examiner knows for sure that the examinee will tell the truth while answering these questions. An example of a crime-irrelevant question is 'Is today Tuesday?' The examiner will then compare the physiological responses to both type of questions. The rationale behind the RIT is that larger responses to relevant crime-related questions than to irrelevant questions indicate that the examinee was lying while responding to the crime-relevant questions. However, the premise used in RIT polygraph tests is incorrect. A strong physiological response could also occur, for example, when truthful examinees are afraid of not being believed. Also, in the context of a criminal investigation, the question 'Did you steal that money?' is likely to be more arousal provoking than the question 'Is today Tuesday?' It is, therefore, clear that RIT is an inappropriate technique in polygraph testing, and there is agreement amongst polygraph researchers that such a test should not be used (Honts, 1991; Lykken, 1998; Raskin, 1986; Saxe, 1994). Larson himself acknowledged the limitations of his technique and declared that he was not particularly happy with the importance others gave to it. He said in 1961: 'I originally hoped that instrumental lie detection would become a legitimate part of professional police science. It is little more than a racket. The lie detector, as used in many places, is nothing more than a psychological third-degree aimed at extorting confessions as the old physical beatings were. At times I'm sorry I ever had any part in its development' (cited in Lykken, 1998, pp.28-29).

The Control Question Test

The Control Question Test (CQT, also labelled the Comparison Question Test) compares responses to relevant questions with responses to control questions (Elaad, 2003). *Relevant questions* are specific questions about the crime. A relevant question in a murder investigation could be: 'On March 12, did you shoot Scott Fisbee?' (Iacono & Patrick, 1997). *Control questions* deal with acts that are indirectly related to the crime under investigation, and do not refer to the crime in question. They are general in nature, deliberately vague, and cover long periods of time. They are meant to embarrass the suspects (both guilty and innocent) and to evoke arousal. This is facilitated by giving the suspect no choice but to lie when answering the control questions. Examiners formulate control questions for which, in their view, denials are deceptive. The exact formulation of these questions will depend on the examinee's circumstances, but a control question in an examination regarding a murder might be: 'Have you ever tried to hurt someone to get revenge?' (Iacono & Patrick, 1997), where the examiner believes that the examinee did indeed hurt someone earlier in his life. Under normal circumstances, some examinees might admit this (control) wrong-doing. However, during a polygraph examination they will not do this because the examiner will tell the examinee that admitting this would cause the examiner to conclude that the examinee is the type of person who would commit the crime in question and would, therefore, be considered guilty. Thus, the examinee has no other choice than to deny this (earlier) wrong-doing and thus to be untruthful in answering the control questions. Obviously, an examinee cannot be found guilty for having committed the crime under investigation by answering control questions untruthfully, as the control questions are not directly related to the crime. The examiner knows this and in that respect their statements are purposefully misleading.

The CQT is based on the assumption that in the *innocent suspect*, control questions will generate more arousal than the relevant questions. This pattern will emerge because the innocent examinee will become more concerned with regard to his or her answers to the control questions, because: (i) the examiner puts so much emphasis on the control questions; and (ii) the examinee knows he or she is lying to the control questions but is answering the relevant questions truthfully. However, the same control questions are expected to elicit less arousal in *guilty suspects* than the relevant questions. A guilty suspect gives deceptive responses to both types of question, which in principle should lead to similar physiological responses to both types of question. However, relevant questions represent the most immediate and serious threat to the examinee, which are expected to lead to a stronger physiological response than the control questions.

Several problems have been identified with this method (see Vrij, 2000; Ben-Shakhar, 2002), including the following three problems. First, although the method may be an improvement compared to the RIT, the alternative explanations for heightened arousal when answering relevant questions (i.e. (i) arousal evoking questions, and (ii) the fear of not being believed) still cannot be ruled out. Second, the examiner purposefully misleads the examinee. This misleading aspect is essential in the test, as creating arousal while answering control questions is necessary for innocent examinees to pass the test. However, misleading suspects may well be considered unethical in some countries. Guidelines on testing typically stress the importance of providing testees with clear information on assessment processes. Third, the test is not standardised, as the control questions that could be asked depend on the type of crime under investigation. When investigating a theft different control questions need to be asked than when investigating a murder. Also, control questions such as 'Have you ever tried to hurt someone to get revenge?' can only be asked to examinees who are known to have hurt someone in the past. The lack of standardisation means that much depends on the skills of the individual polygraph examiner who formulates the questions.

The Directed Lie Test

This standardisation issue is addressed in the Directed Lie Test. In a DLT test, the control questions are standardised and can be asked in all situations. Typical examples of such control questions are 'During the first 27 years of your life, did you ever tell even one lie?' and 'Before age 27, did you ever break even one rule or regulation?' (Raskin & Honts, 2002). Examinees will be instructed to answer 'No' to these questions. They will also be instructed to think about particular situations in which they did tell a lie or did break a rule during these (control) denials. The rationale behind the DLT is similar to the

rationale behind the CQT. Guilty suspects are thought to be mostly concerned with the relevant questions and are expected to show the strongest responses to these questions; innocent suspects are thought to be more concerned with the (control) directed lie questions since they will be concerned that their responses while lying (i.e. to directed lie questions) differ from their responses when telling the truth (i.e. to relevant questions). Although the standardisation problem of the CQT might be reduced with the DLT test, the issues of relevant questions which might be more arousal-evoking than the control questions and fear of not being believed while replying to relevant questions remain.

The Guilty Knowledge Test

The aim of the Guilty Knowledge Test (GKT) (sometimes known as the concealed information test) is to examine whether examinees possess knowledge about a particular crime that they do not want to reveal (Ben-Shakhar, Bar-Hillel & Krenmitzer, 2002; Ben-Shakhar & Elaad, 2003). For example, suppose that the examinee killed somebody with a knife, left the knife at the murder scene, but now denies any involvement in the crime. In a Guilty Knowledge Test the examiner will show this suspect several types of knife, including the one used in the murder. For each knife the examinee will be asked whether he or she recognises the knife as the one used in the murder. Both innocent and guilty examinees will deny each time that they have used such a knife. A guilty examinee, however, will probably recognise the knife he or she has used. It is assumed that this so-called *guilty knowledge* will produce a heightened physiological response which will be detected by the polygraph. Lykken (1998) described how the GKT could have been used in the O.J. Simpson murder case. Questions which could have been used in a GKT immediately after the body of Simpson's wife was found, included: (1) 'You know that Nicole has been found murdered, Mr. Simpson. How was she killed? – Was she drowned? Was she hit on the head with something? Was she shot? Was she beaten to death? Was she stabbed? Was she strangled? and (2) Where did we find her body? Was it – In the living room? In the driveway? By the side gate? In the kitchen? In the bedroom? By the pool? (Lykken, 1998, p.298).

The GKT test, when properly carried out, does not suffer as much from the problems (mentioned above) of arousal-evoking relevant questions and fear of not being believed.¹ The GKT test is theoretically more sound and is less disputed amongst scientists than the CQT test. Iacono and Patrick (1997) published a survey of scientific opinion about both these types of polygraph testing. They asked members of the American Society of Psychophysiological Research (who can be considered as experts) and fellows of the American Psychological Association (Division 1, general psychology) for their opinions regarding CQT and GKT polygraph tests. The findings revealed that the opinions of both groups of psychologists were very similar, and that they favoured the Guilty Knowledge Test. The majority of interviewees (about 75 per cent) considered the Guilty Knowledge Test to be based on scientifically sound psychological principles or theory whereas only a minority (about 33 per cent) thought that this was the case for the Control Question Test.

The Committee to Review the Scientific Evidence on the Polygraph (National Research Council, 2003) provided a thorough and outstanding discussion of the theoretical basis of polygraph testing. They discussed several theoretical accounts that have been used as justification for CQT testing (conflict theory, conditioned response theory, psychological set theory, threat-of-punishment theory, related arousal theory, and dichotomisation theory). All theories predict that deceptive examinees will show stronger physiological reactions to relevant than to control questions. However, as the Committee pointed out, all theories predict that truthful examinees, under certain conditions, will show physiological patterns that are similar to those expected from deceptive examinees. The Committee also criticised how cardiovascular (blood pressure), electrodermal (skin response) and respiratory measures are all lumped together into one general response, as the different measures respond in different ways to various psychological states. Moreover, they argued that current knowledge about physiological responses to social interactions (for example, issues such as: (i) expectancy effects; and

¹ However, these two problems can also occur in a GKT, for example, when the correct multiple choice alternative is a more arousal-evoking option than the other alternatives, or when the innocent suspect can guess what the correct alternative is. Whether this is the case could easily be checked by conducting mock tests with known innocent persons. The test is unfair when these mock suspects show stronger responses to the 'correct' alternatives.

(ii) stigmas which might cause examinees to feel a sense of discomfort and anxiety during social interactions) have been widely ignored in polygraph theory. More generally, the Committee pointed out that, although psychophysiological detection of deception is one of the oldest branches of applied psychology (it started in the late 1800s) it has been affected relatively little by advances made over more than a century in basic psychology and physiology.

The Committee concluded that, although the theory underlying GKT testing (orienting theory) is more plausible than the theory underlying CQT testing, GKT is not without theoretical problems either. The Committee's main concern is that reactions to familiar and unfamiliar stimuli should be thought of as a continuum rather than a dichotomy (as suggested by GKT polygraphers). That is, suppose that the murderer used a revolver and suppose that the innocent examinee owns an unregistered pistol. That examinee might show enhanced responses to questions that mention handguns among the alternatives, even when he or she has no concealed knowledge about the murder weapon.

Other criticisms concerning the Guilty Knowledge Test are largely related to its assumed limited applicability. The problem with GKT is that it is restricted to questions which *only the person who designed the questions/test and the guilty examinee* know the answers. The person who designs the test should know the correct answer, otherwise there is a risk that the correct answer is not in the set of alternatives. Moreover, the GKT only works when questions are asked about details that are actually known to the culprit, otherwise there is no guilty knowledge to detect. This is not always the case. The guilty suspect may not have perceived the details the examiner is asking about, or may have forgotten them by the time the test takes place. The longer the period between the crime and the polygraph test, the more likely it is that the suspect has forgotten certain details. The problem is that the person who designs the test can never be sure that the culprit knows the answer to the crucial questions.

Moreover, a suspect may admit having guilty knowledge but nevertheless deny guilt. This happens when the suspect admits being present but denies the specific alleged acts. The most common example is an alleged sexual assault in which the witness claims that force was used and the suspect admits the sexual acts but claims that they were consensual. Similar problems arise in cases where there are several suspects who were actually involved in the crime to an extent but all deny having been the principal actor (Raskin, 1988).

Finally, questions can be asked only about items to which innocent suspects do not know the answer (otherwise they will have guilty knowledge as well). In many cases the salient details of the crime are made available by the media, investigators or lawyers.² In order to minimise this problem, a decision could be made to ask questions about minor details which are not widely known. This, however, increases the likelihood that the guilty suspect does not know the answers either.

The result is that the number of cases in which the GKT can be used is limited. Podlesny (1995) analysed criminal case files of the American Federal Bureau of Investigation (FBI) and found that in only nine per cent of the cases in which CQTs were used, a GKT could have been used as well. Bashore and Rapp (1993) believe that the limited applicability of the test is the principal obstacle to broader use of the technique.

Lykken (1998) rejected the idea that the GKT cannot be used in many cases. He pointed out that at present FBI investigators are not trained to search fresh crime scenes for usable GKT items. If they were trained then the test could be used more often. Lykken made a comparison with the search for

² Making details about the crime available to suspects, so-called information leakage, is according to Ben-Shakhar, Bar-Hillel and Kremnitzer (2002) the main problem with GKT testing. It is common practice to disclose details of crimes to suspects in police interviews as it is seen as a possible way to make suspects confess (Inbau, Reid, Buckley & Jayne, 2001). Many suspects, including those who are innocent, therefore, might have guilty knowledge after being interviewed. Ben-Shakhar *et al.* (2002) point out that this might not be problematic as long as innocent suspects are aware of having acquired the guilty knowledge in this way and can account for it. However, if guilty knowledge is leaked without the innocent suspects' awareness, they might incriminate themselves. Also, leaking crime relevant details to suspects in police interviews might rescue guilty suspects as they then can point out that they obtained this guilty knowledge in the interview rather than through being involved in the crime.

fingerprints and stated that 'had Podlesny (*the one who analysed the FBI case files, just mentioned*) been working at Scotland Yard in 1900 at the time of the introduction of the Galton-Henry system of fingerprint identification, it is likely that he would also have found very few cases in the records of the Yard that included fingerprints of suspects.' (Lykken, 1998, p.305 – the words in italics have been added.)

Accuracy (validity) of polygraph testing in criminal investigations

Whether the polygraph is a useful procedure for criminal investigations should be tested utilising appropriate procedures. Scientific laboratory studies of polygraph testing often use a 'mock crime' paradigm. 'Guilty participants' are instructed to commit a mock crime and 'innocent participants' are told that they are suspected of such a crime. Both innocent and guilty participants are then submitted to a polygraph test. These laboratory studies, which generally show somewhat favourable results for polygraph testing, have been fiercely attacked by polygraph opponents. Amongst other things, they argue that the guilty participants have little incentive to try to beat the test and that innocent participants are unlikely to be concerned about the relevant questions (Iacono & Patrick, 1997). (See Raskin & Honts (2002), Vrij (2000), MacLaren (2001) and Ben-Skakhar & Elaad (2003) for recent reviews of laboratory-based polygraph studies.)

Numerous field studies have been published to date, but they are subject to debate. The problem is that the quality of polygraph field studies that have been carried out or published to date is low (National Research Council report, 2003, of the Committee to Review the Scientific Evidence on the Polygraph).³ One of the main problems is establishing the ground truth, that is, establishing with certainty whether the suspect is actually innocent or guilty. Ideally, this would be done with corroborative and conclusive evidence that is gathered independently of the polygraph test, such as DNA evidence. However, typically polygraph tests are conducted when no corroborative evidence is available, or as Honts, Kircher and Raskin (2002, p.452) pointed out: 'If there had been strong proof of guilt or innocence in the actual cases, polygraph tests would probably not have been conducted'. Therefore, in field studies confessions are widely accepted as ways to establish the ground truth, however, they are problematic. A suspect is considered guilty when he or she confesses to the crime and is considered innocent when another person confesses to the crime under investigation. The problem with confessions is that they are not independent from polygraph outcomes. For example, a guilty suspect who passes the test is unlikely to confess as there is no further evidence against him or her. Since that suspect is the culprit, it is unlikely that someone else will confess to that crime. In other words, in this case a confession will typically not occur. This, however, implies that the case will not be included in the field study as only cases in which someone made a confession will be included. The incorrect polygraph decision will, therefore, not be noted, and the result is that accuracy percentages reported in field studies which are based on confessions (almost all field studies) are likely to be inflated.⁴

In this section, we will limit ourselves to discussing the accuracy of CQT and GKT polygraph field tests. The RIT technique is disputed by most polygraph users, and therefore its accuracy will not be discussed. Regarding the Directed Lie Test, only one field study has been published (Honts & Raskin, 1988) which is not enough to draw conclusions from.

³ This Committee looked at 194 separate studies (mixture of laboratory and field studies) of which only 57 (50 laboratory and seven field studies) met their minimal standards of scientific adequacy. They looked at the following six minimal criteria: (1) documentation of examination procedures sufficient to allow a basic replication; (2) independently determined truth; (3) inclusion of both guilty and innocent individuals as determined by truth criteria; (4) sufficient information for quantitative estimation of accuracy; (5) polygraph scoring conducted blind to information about truth; and (6) in experimental studies, appropriate assignment to experimental groups germane to estimating accuracy (mainly, guilt and innocence). Regarding establishing the ground truth in field studies (criterion 2) a study passed if (i) truth was defined by a confession, (ii) adjudication by a legal process, or (iii) review of the case facts by a panel who were uninformed about the results of the polygraph examination. It should be noted that none of these three criteria determines the ground truth with absolute certainty.

⁴ See also Fiedler, Schmid & Stahl (2002) for a discussion about problems in establishing the ground truth. They summarise their discussion as follows: 'those (cases) who can confirm the test result are more likely to be included in an empirical validity study than those who can disconfirm the test result' (p.318).

Several reviews have been published to date regarding the accuracy of CQT. The results of field studies reviews are in Table 1 and involve accuracy scores regarding blind-to-guilt/innocence scoring of the polygraph charts. The different reviews produced somewhat different outcomes because the authors included different studies in their reviews (e.g. because the quality of the published field studies is generally low, authors often left out some which they felt didn't meet minimum quality criteria. However, the authors left out different field studies). Saxe, Dougherty and Cross (1985) attempted to provide 'an objective description, to the extent that is possible, of current psychological knowledge of polygraph testing' (p.356). They presented a review which was initiated by the USA congressional Office of Technology Assessment (OTA) to advise President Reagan about polygraph testing. They found 10 studies that met the OTA standards. It should be noted that the ground truth in most studies which were included in all reviews was confession-based. There is reasonable agreement between the reviews regarding guilty suspects. Correct classifications were made in 83 per cent to 89 per cent of the cases, whereas incorrect decisions (classifying a guilty suspect as innocent) were made in 10 per cent to 17 per cent of the cases. However, there is less agreement amongst the reviews regarding innocent suspects. Nevertheless, the findings for innocent suspects are less encouraging than for guilty suspects. Depending on the review, between 53 per cent and 78 per cent of innocent suspects were correctly classified and between 11 per cent and 47 per cent were incorrectly classified.⁵ (The National Research Council report, 2003, of the Committee to Review the Scientific Evidence on the Polygraph found seven studies which met their standards. Their review is not included in Table 1 because the Committee did not report accuracy rates.)

Table 1: Results of field study reviews testing the accuracy of the Control Question Test.

Reviews	Guilty			Innocent		
	Correct	Wrong	Inconclusive	Correct	Wrong	Inconclusive
Ben-Shakhar & Furedy (1990, N=9)	84%	13%	3%	72%	23%	5%
Carroll (1991, N=3)*	83%	17%	–	53%	47%	–
Honts & Perry (1992, N=3)	86%	11%	3%	59%	30%	11%
Iacono & Patrick (1997, N=3)	84%	–	–	56%	–	–
Lykken (1998, N=4)**	86%	–	–	61%	–	–
OTA report (Saxe <i>et al.</i> , 1985, N=10)	88%	10%	2%	78%	20%	2%
Raskin & Honts (2002, N=4)	89%	1%	10%	59%	12%	29%

N refers to the number of studies included in the review.

* Inconclusive cases were not included.

** Incorrect classifications and inclusive cases were not reported.

Only two field studies regarding the accuracy of the Guilty Knowledge Test seem to have been published (see Table 2) and their findings differ. Both revealed very good results regarding the classification of innocent suspects (94 per cent and 98 per cent of innocent suspects were correctly classified) but rather poor results regarding the classification of guilty suspects (76 per cent and 42 per cent of guilty suspects were correctly classified).⁶

Given the stronger theoretical underpinnings of GKT testing (see above), higher accuracy rates for GKT could be expected, particularly if better methods for selecting usable GKT items could be developed (see above). Given the weakness of the theoretical foundations of CQT testing (see above), achieving higher accuracy rates for CQT testing is less likely to happen.

⁵ This pattern, higher accuracy in classifying guilty suspects, also emerged in laboratory-based CQT tests. Vrij (2000) overviewed four reviews regarding laboratory based CQT testing, and found 73 per cent accuracy in correctly classifying guilty examinees (nine per cent incorrect and 18 per cent inconclusive decisions) and 66 per cent accuracy in correctly classifying innocent examinees (13 per cent incorrect and 21 per cent inconclusive decisions).

⁶ This pattern is also reflected in laboratory-based GKT testing. Vrij (2000) overviewed three reviews concerning laboratory based GKT testing and found a 96 per cent accurate rate in correctly classifying innocent examinees (four per cent incorrect decisions) and a 82 per cent accuracy rate in correctly classifying guilty examinees (18 per cent incorrect decisions).

Table 2: Results of field studies testing the accuracy of the Guilty Knowledge Test.

	Guilty		Innocent	
	Correct	Wrong	Correct	Wrong
Elaad (1990)	42%	58%	98%	2%
Elaad <i>et al.</i> (1992)	76%	24%	94%	6%

* The GKT does not have an inconclusive category.

Two additional issues

There are two additional issues which may affect the accuracy of polygraph tests, that is: (i) the ability of people to beat the polygraph test; and (ii) the base rate issue.

Countermeasures

Polygraph test outcomes will often have serious negative consequences for guilty examinees, and they might, therefore, try to influence polygraph outcomes and try to produce physiological responses that may lead the examiner to conclude that they are telling the truth. Methods to achieve this are called 'countermeasures'. Countermeasures are deliberate techniques that some guilty people use in order to beat the polygraph test (Gudjonsson, 1983, 1988). It is possible that innocent subjects may sometimes also use deliberate countermeasures to influence the outcome of the test (Clifton, 1991), but there are no data available on this subject population (National Research Council, 2003).

Gudjonsson (1988) noted that there are three main ways of beating the polygraph test:

- Attempting to suppress the physiological reactions to the relevant questions or items presented in order to make the difference in responsivity between the relevant and control questions minimal, or in the right direction. This typically involves mental countermeasures which include: general relaxation or meditation, examinees trying to distract or dissociate themselves from the anxiety associated with the relevant question, training in hypnosis prior to the polygraph examination to produce 'amnesia' for the offence, and biofeedback training. The advantage of using mental countermeasures is that they are usually more difficult to detect than physical countermeasures.
- Attempting to reduce overall anxiety or reactivity (i.e. to dampen physiological responses) during the assessment (e.g. by the use of drugs prior to the examination).
- Attempting to augment physiological reactions to control or neutral questions or items presented in order to reduce the differentiation between the relevant and control questions or items. This can be achieved by inflicting physical or mental pain or producing muscle tension when the control items are presented.

It is probably easier for examinees to increase their arousal while answering control questions than to lower their arousal while answering relevant questions. Therefore, countermeasures are generally designed to increase arousal to control questions. See Gudjonsson, (1988), Vrij (2000), and Honts and Amato (2002) for reviews of countermeasures studies.

Reid and Inbau (1977) did not seem to worry about the effectiveness of countermeasures. They argued that it is highly improbable that countermeasures can succeed, because properly trained examiners would notice that the examinee is trying to fool them. However, several studies, some conducted by polygraph supporters, have shown that the use of countermeasures can be very effective in defeating polygraph tests, and that they sometimes remain unnoticed by polygraph examiners (Ben-Shakhar & Dolev, 1996; Elaad, 1987, cited in Ben-Shakhar & Furedy, 1990; Honts, Devitt, Winbush & Kircher, 1996; Honts, Hodes & Raskin, 1985; Honts, Raskin & Kircher, 1994).

The most famous countermeasures test was probably conducted by Floyd 'Buzz' Fay, a man who was falsely convicted of murder in the USA on the basis of a failed polygraph examination. He took it on himself to become a polygraph expert during his two-and-half years of wrongful imprisonment. He coached 27 inmates, who all freely confessed to him that they were guilty, in how to beat the control question polygraph test. After only 20 minutes of instruction, 23 of the 27 inmates were successful in defeating the polygraph examination (Ford, 1995; Kleinmuntz & Szucko, 1984).

Base rates

The base rate issue refers to the number of people that can be expected to be guilty in criminal investigations. For example, suppose an organisation is interested in identifying potential spies within it. The number of people spying in an organisation is typically small, say one in 1000 employees. To catch the spy with a polygraph examination is virtually impossible. Polygraph examinations are likely to falsely accuse at least some innocent employees, and suppose that this happens in 30 per cent of the cases (this is not an unrealistic estimate, see Table 1). That implies that approximately 300 out of 1000 employees will fail the test, with the vast majority of them being innocent. Other techniques will be needed to further scrutinise these 300 employees. Obviously, scrutinising a selection of 300 employees is less time consuming than scrutinising all 1000 employees. However, clearing employees just on the basis of passing a polygraph test is risky as some guilty employees will be able to slip through the net and will pass the polygraph test. Such a chance is not theoretical, as people can beat polygraph tests (see above) and professional spies are trained how to beat polygraph tests.

Aldrich Ames, the CIA-agent who sold secrets to the Soviets for many years passed several polygraph tests during this time. Lykken (1998, p.3) pointed out that Ames 'succeeded in his spy career for as long as he did because his ability to beat the lie detector deflected official suspicions.' Ames' KGB controller Viktor Cherkashin, explained in an interview with *The Sunday Times* (8 February 1998, p.21) how he helped Ames to pass polygraph tests. Cherkashin had arranged a lunch between Ames and a Russian diplomat. To Ames' surprise, Cherkashin attended the lunch himself as well. Ames became nervous, as the FBI knew Cherkashin and had him watched. Cherkashin, however, attended the lunch on purpose. He knew that the CIA often made its officers undergo routine polygraph tests and knew that Ames would be asked 'If he had recently had unofficial contacts with KGB officers?' because this was a standard question they always asked. Since contacts between Ames and KGB officers were secret, Ames would typically have to lie while answering this question. After the lunch, however, there was no need for Ames to lie about this anymore, and he could safely say that he had been approached.

The use of the polygraph in employment and security screening

At one time the employment applications of the polygraph were widespread in the USA but the United States Employee Polygraph Protection Act of 1988 prohibited the use of the polygraph for pre-employment screening except by some government agencies and certain strategic industries such as nuclear power. However, some organisations might be considering using the polygraph in pre-employment screening to try to identify those likely to engage in counter productive behaviour or those who may be hiding criminal activity or other sensitive information that might disqualify them from employment. Organisations might also be considering using the polygraph in periodic screening of current employees. However, given the lack of a specific event to be investigated, the Relevant/Irrelevant questioning procedure is often used for such screening but this procedure has been heavily criticised (see above section on criminal investigations). A major problem with the use of the polygraph in employee screening is the large proportion of false positive outcomes.

Published studies of the validity of the polygraph in employee selection are flawed in a number of ways. They typically fail to use standard research paradigms to investigate validity. Results are generally quoted in terms of percentage accuracy rather than correlation coefficients. This approach is problematic when there is a low base rate of the behaviour to be identified. If only five per cent of the population are dishonest, then random selection classifying all those chosen as honest will achieve 95 per cent accuracy, yet be entirely ineffective at identifying dishonest individuals. Other studies compare polygraph results from prison populations and employees and show a better than chance ability to differentiate members of these two groups. However, there is no evidence that job applicants likely to engage in counterproductive behaviour will respond in the same manner to polygraph testing as convicted felons whose crime was serious enough to warrant a prison sentence (Sackett & Decker, 1979).

The 2003 report of the United States National Research Council (NRC), which had employee screening by government agencies as its main focus, stated that: 'There are no studies that provide even indirect evidence of the validity of the polygraph for making judgements of future undesirable behaviour from pre-employment screening tests. The theory and logic of the polygraph ... is not consistent with ... forecasts of future ... performance ...' (p.150).

The NRC report did state that: 'Although we believe it likely that polygraph testing has utility in screening contexts because it might have a deterrent effect, we were struck by the lack of scientific evidence concerning the factors that might produce or inhibit deterrence' (p.187). However, the deterrent effect of polygraph-based procedures depends upon people believing that these are valid and reliable.

The NRC report concluded that: 'The polygraph as currently used has extremely serious limitations for use in security screening to identify security risks and to clear valued employees' (p.218) and that: 'Its accuracy in distinguishing actual or potential security violators from innocent test takers is insufficient to justify reliance on its use in employee screening in federal agencies' (p.219).

The 1986 report of the BPS working group on the use of the polygraph concluded that: '... polygraph-based techniques ... are unlikely to be acceptable in the British context of employment and staff screening' (p.81).

The current BPS working group agrees with these two conclusions. The limited usefulness of the polygraph in employee screening justifies that efforts be made to consider other methods of employee screening. (Such methods are mentioned in a later section of the present report.)

The use of the polygraph in the clinical setting

The polygraph has been used for clinical purposes in the past in a number of ways. For example:

- to assist in establishing the identity of an allegedly amnesic patient (Gudjonsson, 1979);
- to diagnose genuine blood injury phobia relating to providing a blood sample (e.g. in cases of alleged 'drunk driving') (Gudjonsson & Sartory, 1983);
- to identify sexual preferences and arousal by the use of the penile plethysmograph (Salter, 1988).

More recently Grubin (2002) has suggested that the polygraph can also be used clinically:

- in the management and treatment of sex offenders (English, Jones, Pasini-Hill, Patrick & Cooley-Towell, 2000; Emerick & Dutton, 1993; Wilcox, 2000);
- in cases of suspected substance use;
- to improve compliance with prescribed medication;
- to explore mental states (e.g. '...in patients who are guarded about the nature and existence of delusions or hallucinations').

Of these last four, the most common so far has been the application of polygraph techniques with sex offenders (Kokish, 2004; Wilcox, Sosnowski & Middleton, 2000).

Sex offenders

It is well known that sex offenders typically minimise the nature and extent of their sexual offending (e.g. Ahlmeyer, Heil, McKee & English, 2000; Salter, 1988). The most likely reason is their feeling of shame. Indeed, the greater the degree of shame generated by the offence the more sex offenders are inhibited from admitting the offence and disclosing the details (Gudjonsson, 1992, 2003).

Gudjonsson and Sigurdsson (2000) compared the confession rate of violent offenders, rapists and child molesters. The highest rate was found for child molesters (83 per cent) and lowest for rapists (61 per cent). The confession rate of violent offenders fell in between the other two groups (77 per cent). These findings confirm the findings of Nugent and Kroner (1996) that rapists confess less readily to their crime than child molesters. In the Gudjonsson and Sigurdsson study the rapists were also significantly more likely to dispute their confession at trial than the child molesters. The most likely explanation for these findings, according to Gudjonsson and Sigurdsson, is that it is more difficult to find convincing evidence against rapists than child molesters (e.g. rapists typically claim that the sexual act was consensual when there is forensic evidence against them).

Considering the difficulties presented by sex offenders' denials, the seriousness of their offending, and the likelihood of reoffending, it is not surprising that the polygraph has in recent years begun to be applied with this population of offenders to aid their management and treatment. The 'surveillance of probationers' dates back to the 1980s (Abrams & Ogard, 1986), although there has been a growing interest in the application of the polygraph with sex offenders in the last few years in the USA (Abrams & Simmons, 2000; Hindman & Peters, 2001). It is still only employed in a minority of programmes (Abrams & Simmons, 2000). Unfortunately, there appears to be an absence of information concerning the scientific validity (i.e. evidence base) of the procedure.

According to Grubin (2002) the value in using the polygraph with sex offender monitoring, management and treatment is the substantial increase in the disclosure of important information, such as:

- the number and types of victims;
- the types of offences committed;
- deviant fantasies and behaviours.

Grubin (2002) and Grubin *et al.* (in press) discuss a British pilot project with 50 sex offenders, mainly child molesters. The innovative nature of this project requires that it be extensively described here. The aim of the study was to examine whether the introduction of polygraph examinations assists offenders in avoiding behaviours that had been identified as increasing their risk of re-offending; for example, using pornography, or going to places where potential victims might be located.

It did not seek to identify re-offences, but those behaviours that, if left unchallenged, could increase the likelihood of a re-offence.

Of 116 sex offenders approached for the study, 50 (43 per cent) agreed to participate. All of these were told to come back in three months' time, with just over half ($N=27$, 54 per cent) ('The Polygraph Aware Group') being told that they would then have to take a polygraph test. Twenty-two (44 per cent) were not specifically told about a polygraph test and were expecting to be attending for a 'review' ('The Polygraph Unaware Group'), and one (two per cent) was excluded due to 'acute mental illness'.

During the polygraph test the 'Comparison Question Technique' (CQT) was used. The three or four 'high risk' behaviours identified by the treatment providers were used as the relevant questions and comparison questions included such questions as 'Have you done anything over the last three months that would concern your probation officer?', which were expected to arouse anxiety but to a lesser extent than the relevant questions. The polygraph examination consisted of three phases: (a) a pre-test interview, where the procedure was explained and the 'high risk' behaviours were confirmed; (b) the polygraph examination itself, when the participant was asked if he had engaged in any of the high risk behaviours; and (c) post-test interview, during which the participant was told if he had failed the test and provided with the opportunity to explain the alleged deceptive responses.

At the time of the first polygraph examination, two of the 49 participants had been recalled to prison and 14 did not turn up for the evaluation. One participant out of 'The Polygraph Unaware Group' refused to take part after being asked to take a polygraph test. This left a total of 32 men who were polygraphed at 'Time 1' and 31 of these (97 per cent) admitted to having engaged in at least one high risk behaviour (e.g. masturbating to deviant fantasies, obtaining pictures of children for the purpose of sexual arousal, having unsupervised contact with children). One man admitted to about 50 separate incidents of frottage involving young girls. The mean number of high-risk behaviours reported was 2.45 ($SD = 1.7$, range 0–7). The one participant who disclosed no high-risk behaviour failed the polygraph test, but gave no explanation for the failure. The participants were told that because of their reporting of high-risk behaviour they would be polygraphed again in six months' time.

Out of the 32 participants who were polygraphed at Time 1, 21 (66 per cent) turned up for the second polygraph test. This represents 18 per cent of the 116 sex offenders originally asked to participate in the study and 42 per cent of those who at first interview agreed to participate in the study.

Grubin's (2002) main conclusions with regard to the utility of the polygraph with sex offenders were:

'Thus, polygraphy can contribute substantially to treatment programmes, as well as assisting offenders to avoid the sorts of behaviours that increase their risk of re-offending: it encourages offenders to disclose information that is relevant to their treatment and supervision. Determination of the reliability and validity of the technique itself is of less pertinence than when it is used in investigative settings, but the empirical base for its use in treatment and supervision nevertheless needs improving.' (p.48)

'We concluded that those offenders who were motivated not to re-offend found the procedure beneficial while those who were less motivated avoided it.' (p.51)

'Although a small study, our results suggest that the most effective use of the polygraph may be as a 'truth facilitator' rather than as a lie detector. It can bring worrying behaviours to the attention of supervisors and treatment providers, allowing effective intervention and additional treatment before offending occurs. For the offender himself, it can provide an additional external support that assists him to adhere to treatment and relapse prevention plan.' (p.51)

Grubin *et al.* (in press) recognise the possibility that some of the participants may have made false admissions to please the examiner (labelled by the researchers as 'defensive lying'), but this could not be independently verified.

Ethical issues in the clinical setting.

Although this report has a later section on human rights and codes of conduct, there are a number of possible ethical issues specifically related to the application of the polygraph with offenders as a part of their management, supervision, and treatment in the community, including:

1. The use of the polygraph with sex offenders would seem to be primarily designed to encourage them to discuss their involvement with high risk factors (e.g. deviant sexual fantasies and masturbatory practices, breaches of the conditions of their licence), this being a use of the polygraph outside its original remit.
2. The validity of the application of the polygraph with sex offenders as part of their supervision, management and treatment is not scientifically established, although there is growing evidence that it encourages offenders to disclose their deviant thoughts and actions, and may help them exercise self-control. Therefore, it may help with relapse prevention, but more research is needed to ascertain its effectiveness in practice. In particular, the false positive and negative error rates are unknown.
3. In the pilot study by Grubin *et al.* (2002), the results of which need to be treated with caution in view of the small number of participants, it is evident that there are high refusal and drop out rates. What would happen to prison inmates and other sex offenders who refuse to co-operate with the polygraph testing? Would this refusal hinder their being released on licence back into the community? It is possible that a failure to co-operate with the use of the polygraph and a negative result (i.e. apparent deception) may have negative consequences for the sex offender concerned.
4. The use of polygraph testing could be perceived by the sex offenders as being coercive rather than voluntary.

Other possible methods to detect deception

The polygraph traditionally has involved measures of peripheral physiological manifestations thought to be involved in deception or to relate to guilty knowledge. More recently attempts have been made to investigate alternative or additional measures, for example, brain function.

Event-related brain potentials

Electrical activity within the brain can be detected using electrodes on the scalp. One aspect of this is usually referred to as the 'P300' component, which relates to a response occurring in the brain approximately 300 milliseconds after a 'significant' event has taken place. There is a considerable body of published literature on the general psychological significance of this response which is thought to be related to information processing (see Donchin, 2003).

Of relevance to this report is the possible use of the P300 response in settings that are similar to the guilty knowledge test. Also, since the P300 may be small when a person is carrying out a concurrent task, reductions in it could occur when a person is having to concentrate on lying. The 2003 NRC report concluded (p.175) that as yet there had been published 'very limited data on accuracy' but that the use of P300 in lie detection settings could have an accuracy 'level similar to that of the polygraph'.

Of interest is the under-researched notion that since P300 and traditional polygraphy measure different types of activity, combining the two might result in higher accuracy. Also under-researched is the effect of the use of counter measures on P300 accuracy.

Functional brain imaging

Functional brain imaging involves measuring blood activity in the brain using positron emission topography (PET) or magnetic resonance imaging (MRI). These procedures are proving useful in studying neural correlates of psychological functioning. However, using such procedures to study brain activity that may be associated with lying has only recently commenced. One problem facing such work is that the parts of the brain that may be more active during deception also are probably more active during non-deceptive activities. Thus the false positive errors found in traditional polygraphy may occur using measures of brain activity.

Functional brain imaging is currently very expensive and time-consuming. It is also probably alarming to those asked to undertake it in attempts to detect deception, including innocent people.

More research is needed on the possible use of functional brain imaging to detect deception.

Behaviour/demeanour

In contrast to the very limited amount of research published concerning the possible links between brain activity and deception, there is a wealth of published research on the possible relationships between demeanour/bodily behaviour and deception. This research has been comprehensively reviewed by De Paulo *et al.* (2003) and by Vrij (2000).

A very large proportion of this type of research concerns lying in low stakes situations (e.g. merely for the purpose of the experiment – as is the case in many laboratory-based polygraph experiments). Much of such research has found that people are often poor at detecting lying/truth telling from others' bodily (including facial) behaviour, including relevant professionals. However, a few studies have employed more ecologically valid, high stakes situations (Ekman, O'Sullivan & Frank, 1999; Frank & Ekman, 1997; Mann, Vrij & Bull, 2004) and these have found performance levels of noticeably above chance level but yet far from perfect (i.e. levels of performance not that different from those sometimes found with the polygraph).

In most of these demeanour studies the observers have received no relevant training. However, the limited amount of published research on such training does not reliably find improvements in performance (Bull, 2004).

One possible advantage regarding the observation of demeanour is that, unlike traditional polygraphs, it can be done non-invasively (though this, of course, raises other ethical issues).

Linguistic analyses

Memon, Vrij and Bull (2003) pointed out that the literature overviews by De Paulo *et al.* (2003) and Vrij (2000) were in agreement that quite often increases in voice pitch and in speech errors appear to accompany lying. However, not all liars demonstrate this and not all those attempting to detect deception pick up on these cues. Furthermore, the differences in pitch are often hard for humans to notice. (See the following section on voice stress analysis.)

Memon *et al.* (2003) also reviewed the published research on criteria-based content analysis (CBCA), a procedure that examines the contents of accounts for criteria expected to be more often present in truthful accounts. This procedure was developed for analysis of statements by (alleged) victims and it has rarely been applied to the statements of suspects. While some of its criteria may be relevant to this report (i.e. those believed to be associated with cognitive difficulties relating to lying) others may not be. Nevertheless, there is a growing number of research studies that have found that CBCA criteria are more present in truthful accounts than in false accounts, though which criteria actually found to discriminate varies across studies.

Some other methods of language content analyses have begun to be researched but, as yet, insufficient findings are available on the effectiveness of these procedures. The National Research Council was of the view that 'efforts to design measures for the detection of deception based on language use may have untapped potential' (p.165).

Voice stress analysis

As stated in the previous section, overviews of the relevant research have concluded that some aspects of speech (e.g. changes in voice pitch) may relate, though imperfectly, to lying. However, in its review of various instruments and procedures designed to detect deception by voice stress analyses the National Research Council concluded that 'The practical performance of voice stress analysis for detecting deception has not been impressive' (p.168) and that the relevant research offers 'little or no scientific basis for the use of ... voice measurement instruments as an alternative to the polygraph ...' (p.168). Indeed, voice stress analysis shares many weaknesses with the polygraph (e.g. measuring stress but this may not be caused by deception). However, unlike the polygraph, voice analysis need not be invasive (Gunn & Gudjonsson, 1988).

'Personality' testing

Alternatives to the polygraph test for pre-employment screening purposes have proliferated in the wake of the USA Employee Polygraph Protection Act of 1988. Many USA companies previously engaged in polygraph-based employment screening developed alternative measures of integrity to meet their clients' needs. This has led to the creation of many tests specifically designed to assess honesty or integrity in the workplace. The more general term 'integrity' testing has replaced the term 'honesty' testing. The aim of such tests is to screen out job applicants who are likely to engage in various forms of counterproductive behaviour at work, such as theft, absenteeism, tardiness, drug abuse, sabotage or spying.

At the same time there has been a renewed interest among psychologists in personality assessment generally and its use for the purpose of employee selection. Personality testing is now widely used in pre-employment screening to help match applicants to jobs and has empirical support in terms of moderate predictive validity (e.g. Barrick & Mount, 1991; Robertson & Smith, 2001). Related to the growth in personality testing in pre-employment screening is the improvement in measures of personality that has resulted from the emergence of a widespread (but not unanimous) consensus on the structure of personality. There is now considerable agreement on a hierarchical model of personality traits with five dimensions or factors: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Openness (e.g. Digman, 1990). Valid and reliable measures of these five dimensions have been developed and used to predict various criteria, including job performance and workplace deviance (Cullen & Sackett, 2003). Measures of integrity for the assessment of aspects of

trustworthiness at work have been derived from broader personality inventories such as by using scores on the Conscientiousness dimension of five-factor personality inventories. Both types of measures can typically be administered using paper and pencil forms, or these days via the internet, and are widely used as alternatives to the polygraph in employee screening for integrity.

The range of behaviours encompassed by the integrity construct is extremely wide, and any one test is, therefore, unlikely to be a valid and reliable measure of the full range of deviant workplace behaviours. Overt integrity tests typically are designed to assess common workplace deviance, which can be seen as consisting of two broad categories of behaviour (Bennett & Robinson, 2000): organisational deviance (behaviours directed at harming the organisation such as theft and absenteeism) and interpersonal deviance (behaviours directed at harming other people in the workplace such as rudeness). Overt integrity tests are composed of items that directly ask about past workplace behaviour such as theft and tardiness. Their items often resemble the types of questions used in polygraph screening. In contrast, personality-based integrity testing uses measures of a personality trait such as Conscientiousness to provide a less direct way of assessing characteristics that are likely to predict workplace deviance.

Evaluation of the different approaches to measuring honesty and integrity

In response to the burgeoning of integrity testing, two reviews of these questionnaires were commissioned in the USA. The report by the USA Congress Office of Technological Assessment (1990) reviewed studies that examined the predictive validity of these tests to identify workplace theft. The report's conclusions were equivocal. It did not support banning such testing but nor did it report conclusive evidence in favour of integrity testing. The report commissioned by the American Psychological Association reviewed 46 commercially available integrity tests for their reliability and validity. The conclusions were somewhat more favourable. Where the evidence was available, these tests did appear to have modest predictive validity (Goldberg, Grenier, Guion, Sechrest & Wing, 1991). Recommendations for the proper use of integrity questionnaires and for further research on their validity were published in the *American Psychologist* (Camara & Schneider, 1994).

Effectiveness

As these reports conclude, there is a large body of evidence in the published research literature that both overt integrity tests and indirect personality-based measures predict workplace deviance. Meta-analyses suggest that integrity measures are valid predictors of workplace deviance of various kinds (e.g. Ones, Viswesvaran & Schmidt, 1993, 2003). Evidence has accumulated to indicate that personality traits, particularly the traits of Agreeableness, Conscientiousness, and Emotional Stability, predict various job-performance criteria to a modest but useful degree (e.g. Barrick & Mount, 1991; Cullen & Sackett, 2003).

Objectivity and reliability

The polygraph generates a sample of physiological data, which initially would seem to be more objective than self-report integrity tests. The reliability of the polygraph as a recording device for cardiovascular activity, respiration, and skin conductance is not in question. However, there often is a subjective element to the interpretation of the polygraph record. The same record can be viewed by one expert as indicating a 'pass' and by another as a 'fail' (Ben-Shakhar & Furedy, 1990). In contrast, self-report integrity tests are typically scored following a coding key, often by a computer, and so there is no element of subjectivity in arriving at the score. Moreover, an integrity or personality questionnaire is standardised in the sense that the items are the same for everyone every time it is administered. This leads to typically good levels of reliability for these questionnaires – both internal consistency and test retest. This is not true of the polygraph test. The questioning and the procedure vary with examiners and the purpose of the examination (Ben-Shakhar & Furedy, 1990).

False positives and false negatives

Given the subjective element in interpreting the polygraph record, mistakes are inevitable. A 'guilty' person could be deemed 'innocent' (false negative) and an 'innocent' person could be deemed 'guilty' (false positive). This is one consequence of a pass-fail test. There is a tendency to perpetuate this problem with self-report integrity tests by reducing a continuous score to a dichotomous pass or fail by

the use of a cut point. Because such cutting scores are often not based on adequate or appropriate empirical data, Goldberg *et al.* (1991) cautioned against their use. The OTA report (1990) strongly criticised dichotomised measures of integrity in this area. They argued that it was unethical to label people as 'dishonest' or 'lacking integrity' because honesty is such a fundamental societal value. This was particularly so when the test result might have only moderate validity. Continuous measures, such as those based on personality measurement, do not suffer in the same way because they use a more differentiated measurement scale and typically measure traits which are less contentiously labelled. It is not such a slur on a person's character to be assessed as somewhat lacking in conscientiousness as it is to be labelled as dishonest.

Faking

Probably the major concern of employers with self-report integrity measures is the danger that such tests are easily faked. It seems improbable that job applicants would knowingly describe themselves as dishonest. However, the evidence suggests that faking is not as big a problem as might be supposed (e.g., Cunningham, Wong & Barabee, 1994; Ones, Viswesvaran & Schmidt, 1993). There are several reasons why this may be so. People may believe that their responses will be verified from other sources. Individuals seriously lacking in integrity may not even realise that they are responding in the dishonestly coded direction. Those who do attempt to 'fake good' have to decide what pattern of responding would most closely resemble a truly honest person, and they may get it wrong. The use of less direct items such as those used in personality questionnaires can help reduce faking as can some test formats (e.g. Martin, Bowen & Hunt, 2002).

Faking is a problem with the polygraph also. As stated elsewhere in this report, individuals can coach themselves to alter their physiological responses to try to 'beat' the polygraph.

Situational influences on honesty and integrity

Psychologists have long recognised that behaviour is determined by both internal and external factors. Workplace deviance can be controlled, in part, by situational factors. An extreme example of situational controls on workplace deviance is the thorough searching of employees in diamond mines at the end of their shift. Workplaces can and should be designed to support good citizenship and to deter behaviours that damage the organisation or other employees. Interventions such as effective stock control procedures or activities which develop trust between employees and supervisors have been shown to reduce 'shrinkage' (Murphy, 1993).

However, there are situational influences exerted on employees which are external to the workplace and beyond the employers' control. Internal and external situational influences combine with personality traits to determine behaviour. Some investigation of external influences alongside measures of personality is, therefore, also pertinent to employee screening. References and other forms of checks are often undertaken in pre-employment vetting. However, the reliability and validity of these aspects of pre-employment screening for predicting honesty or integrity in the workplace remains unclear. Robertson and Smith (2001) found that references had some effectiveness as predictors of overall job performance – but less than integrity tests.

Even if valid and reliable measures of integrity, either overt or indirect, have been or can be developed, the question remains whether such measures would be able to detect the most extreme and damaging forms of workplace deviance that are becoming more prominent today. For example, leaks of classified information and acts of violence with the aid of insiders appear to be on the increase. In response, security professionals may be tempted to resort to the polygraph test in the hope that it can screen out people who have the potential to engage in these destructive acts. However, the evidence reviewed does not support such a use of the polygraph. And where acts are motivated by strong moral, political or religious beliefs (as with both whistleblowers and terrorists) it is unlikely that any measure of integrity would be useful in identifying potential perpetrators.

Combining methods

Many of the above alternative methods should only be conducted by those who are qualified to do so. One possible way to improve performance could be to combine the polygraph with one or more of such alternatives. Unfortunately, the NRC '...found no serious investigations of such multi-component screening approaches' (p.8) and had 'not located any studies that attempt directly to measure the incremental validity of the polygraph when added to any of these information sources' (p.198).

However, some polygraphers say that their conclusions are also informed by observing demeanour when preparing a person for and conducting with them a polygraph test (Lykken, 1998), but polygraphers may suffer from false beliefs about demeanour cues (Vrij, 2000).

Human Rights and Codes of Conduct

Given that there are concerns about the reliability and the validity of polygraphic lie detection, issues relating to human rights and professional/scientific codes of conduct need to be addressed.

Attempts to deceive

While in certain settings polygraphic lie detection may produce results better than chance, it is a far from perfect procedure. Anybody who attempts to persuade others (e.g. a polygraph examiner testing a police suspect) that polygraphic lie detection is an error-free procedure (by, for example, using a 'stimulation procedure' – Bull, 1988; National Research Council, 2003) will be attempting to deceive. To claim otherwise would be inconsistent, for example, with the Code of Conduct for psychologists promulgated by the British Psychological Society, as would lying about the outcome of a polygraph test. Indeed, Chartered Psychologists should consider informing participants of the known error rates concerning the procedures they plan to administer.

Furthermore, many of the studies designed to examine the effectiveness of polygraphic lie detection are themselves imperfect. In its overview of the many relevant studies the National Research Council commented that:

'The general quality of the evidence for judging polygraph validity is relatively low: the substantial number of studies most relevant for this purpose were below the quality level typically needed for funding by the National Science Foundation or the National Institutes of Health' (p.2).

and that:

'because the great bulk of polygraphic research has been funded by agencies that rely on the polygraph for law enforcement or counterintelligence purposes, there is significant potential for bias and conflict of interest in polygraph research' (p.119)

though some quality, independent research studies have also been undertaken.

Given that polygraphic lie detection is a far from perfect procedure, people's human rights may well be infringed by requiring them to be subject to polygraphic lie detection or by making the assumption that those who refuse to take a polygraph test have something to hide. Attempts at coercion should be avoided. It would seem inappropriate to suggest that people be required to take a polygraph test to prove their innocence (Clifton, 1991). The possibility of false admissions/confessions (Gudjonsson, 2003) should always be borne in mind.

Even so, many of those who conduct polygraphic lie detection tests seem to have (National Research Council, 2003) 'strong, apparently unshakeable beliefs ... in its efficacy on the basis of their experiences' (p.20). Yet, 'if polygraph testing is not in fact highly accurate in distinguishing truthful from deceptive responses ... agencies that use polygraphs are infringing civil liberties for insufficient benefits to the national security' (National Research Council, 2003, p.58).

A related aspect of civil liberties is the right that people taking part in procedures freely give their informed consent to this. The BPS Code of Conduct emphasises the importance not only of consent and the absence of duress, but also that consent be based on full and valid explanations.

In its 1986 report the BPS raised the issue of whether personality differences (e.g. in anxiety) among people or 'ethnic differences in physiological responding' (p.90) would relate to error rates in polygraphic lie detection. Similar issues were raised in the 2003 review by the National Research Council. To date, insufficient research has been published on such issues, especially the possibility that members of certain groups in society would be associated with more 'false positive' errors (e.g. falsely deciding that a person is lying). Also, people who in the past have been wrongly accused based on a polygraph test may be more likely to be wrongly accused in the future (e.g. because they may be more reactive to control questions).

It is these false positive errors plus, of course, 'false negative' errors (e.g. deciding that a person is telling the truth when he or she is lying), that are a serious human rights issue. While civil societies, professional groups and governments have the responsibility of deciding what are acceptable error rates, those people who actually are the 'false positives' and those who are damaged by the 'false negatives' have their own individual rights.

Polygraphic error rates are still a crucial issue, particularly when the 'base rate' is low (e.g. in security screening when trying to determine which few people among many are a threat). The NRC review commented that screening populations with low rates for the target transgressions (e.g. national security threats such as spying or terrorism) requires diagnostic procedures of very high accuracy, certainly higher than what polygraph testing is ever likely to produce. The review pointed out that with a target transgression rate of 10 in 10,000, and a test sensitivity to detect 80 per cent or more of the targets, then about 1606 persons would 'fail' the test (of whom eight would be targets), resulting in 1598 innocent employees failing. Alternatively, to keep the number of such false alarms to about 40, eight of the ten targets would 'pass' the test. Thus the review stated that: the 'Available evidence indicates that polygraph testing as currently used has extremely serious limitations in such screening applications ...' (p.6).

However, polygraph screening in such security settings could have utility if it deterred people. Unfortunately, it is likely that the very people one would wish to deter will be aware of the limitations of polygraph testing. Inappropriate confidence in the accuracy of polygraph testing by organisations could well be a grave danger to security objectives.

In terms of security screening the NRC review commented that polygraph testing could 'lead to credible claims that agencies that use polygraphs are infringing civil liberties for insufficient benefits to national security' (p.7).

In a number of countries the courts have been apprehensive about admitting testimony concerning the 'outcomes' of polygraphic lie detection. As courts become more sophisticated concerning the admissibility of expert testimony supposedly based on 'scientific' evidence (Faigman, 2003), they are seeking to evaluate the scientific validity on which such testimony is based. This BPS report and the NRC review should be of assistance in this regard.

Main conclusions

Although polygraph equipment does accurately measure a number of physiological activities, these activities do not reflect a single underlying process. Furthermore, these activities are not necessarily in concord either within or across individuals.

The NRC review stated that the 'inherent ambiguity of the physiological measures used in the polygraph suggests that further investments in improving polygraph technique and interpretation will bring only modest improvements in accuracy' (p.2).

The vast majority of published studies on polygraphic lie detection have been laboratory-based and they have not been that successful in recreating many of the factors that would be present in real life (in the 'field') such as high costs of being judged as lying. Thus serious questions persist about generalising beyond laboratory situations. However, the scarcity of good field studies is a substantial impediment to appraising the scientific validity of polygraphic lie detection.

The NRC review noted that 'Polygraph accuracy may be reaching a point of diminishing returns. There is only limited room to improve the detection of deception from the physiological responses the polygraph measures' (p.102). and that 'Almost a century of research in scientific psychology and physiology provides little basis for the expectation that a polygraph test could have extremely high accuracy' (p.2). We agree with these points.

While few good quality studies of the effects of countermeasures have been published, the use of countermeasures does pose a serious threat because the physiological activities measured by the polygraph could be affected either by physical or cognitive means. People who are major threats may well have strong incentives and sufficient resources to use countermeasures effectively (including countermeasures unknown to polygraphers).

Given that even in the most favourable circumstances polygraphic lie detection accuracy is not high, an over-reliance on an imperfect procedure may lead to undue relaxation concerning the developing of: (i) other methods of identifying or screening wrongdoers; and (ii) other ways of ensuring security and preventing crime. The belief that people who 'pass' a polygraph test are, therefore, cleared of suspicion is a false belief.

People incorrectly judged by polygraphers as lying or having guilty knowledge may be falsely convicted or lost to an organisation. Quality personnel may choose not to join organisations that use procedures known to have inaccuracy rates that are not negligible.

Quality of research

The NRC review offered the view: (i) that researchers of international standing have rather rarely chosen to conduct research on this topic ('... promising young scientists from a number of relevant fields have not flocked to forensic science to make their careers', p.98); and (ii) that policy makers, investigators, and members of the legal profession have rarely chosen to have relevant training in science. In its review of all the relevant literatures the BPS working party also noted how few quality researchers had devoted most of their careers to work on this aspect of polygraphy.

The NRC review concluded that:

'Research on the polygraph has not progressed over time in the manner of a typical scientific field. It has not accumulated knowledge or strengthened its scientific underpinnings in any significant manner. Polygraph research has proceeded in relative isolation from related fields of basic science and has benefited little from conceptual, theoretical, and technological advances in those fields that are relevant to the psychophysiological detection of deception' (p.213).

and that:

'Forensic science has not kept up with the state of science more generally for two basic reasons: the legal community's basic ignorance of science and statistics, and the lack of interest among research scientists in the practical (and especially forensic) applications of science. In lie detection, for instance, policy makers have not demanded better work, and few scientists have been interested in pursuing the subject. This powerful combination of ignorance and apathy has ... led to convictions of the innocent, acquittals for the guilty, and numerous costs to individuals, ranging from job loss to social ostracism'! (p.207).

It is probably the case that law enforcement and security agencies need to improve their capability to evaluate effectively the claims made by advocates of techniques for detecting deception. Similarly, while a number of investigators around the world claim that various techniques elicit confessions, there exists little published research on this topic. Indeed, the utility of a technique to elicit confessions probably depends on the possible confessors' beliefs about the accuracy/validity of that technique.

The NRC review noted that the USA government had not made a serious effort to help develop the scientific basis of physiological (or other methods of) lie detection. Such a criticism also applies in Britain. What is needed is an expanded research programme on various possible methods to be administered by an organisation independent of those with vested interests in one or more possible methods for detecting deception. The Association of Chief Police Officers (of England and Wales) has a working group currently examining the effectiveness of a number of possible methods of detecting deception. The work of that group should be built upon.

The polygraph is one among a number of procedures that could be used in attempts to detect deception and integrity but, like all procedures, it has inherent weaknesses. Error rates in polygraphic deception detection can be high. The most appropriate procedure or combination of procedures will depend on the circumstances. Polygraphic deception detection procedures should not be ascribed a special status. We must not deceive ourselves into thinking that there will ever be an error-free way of detecting deception.

References

- 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*, 63–67.
- Ahlmeier, S., Heil, P., McKee, B. & English, K. (2000). The impact of polygraphy on admissions of victims and offences in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *12*, 123–139.
- Barrick, M.J. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.
- Bartram, D. (2002). EFPA review model for the description and evaluation of psychological instruments: version 3.26. www.efpa.be/documentsonprofissues/efpareview_model.htm.
- Bashore, T.R. & Rapp, P.E. (1993). Are there alternatives to traditional polygraph procedures? *Psychological Bulletin*, *113*, 3–22.
- Bennet, R.J. & Robinson, S.L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, *85*, 349–360.
- Ben-Shakhar, G. & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*, *88*, 131–151.
- Ben-Shakhar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp.103–126). London: Academic Press.
- Ben-Shakhar, G. & Dolev, K. (1996). Psychophysiological detection through the Guilty Knowledge Technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*, 273–281.
- Ben-Shakhar, G. & Furedy, J.J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Springer-Verlag.
- Ben-Shakhar, G., Bar-Hillel, M. & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the Guilty Knowledge Technique in court. *Law and Human Behaviour*, *26*, 527–541.
- British Psychological Society (1986). Report of the working group on the use of the polygraph in criminal investigations and personnel screening. *Bulletin of the British Psychological Society*, *39*, 81–94.
- Bull, R. (1988). What is the lie-detection test? In A. Gale (Ed.), *The polygraph test: Lies, truth and science*. (pp.10–18). London: Sage.
- Bull, R. (2004). Training to detect deception from behavioural cues: Attempts and problems. In P.A. Granhag & C.A. Stromwall (Eds.), *Deception detection in forensic contexts*. Cambridge: Cambridge University Press.
- Camara, W.J. & Schneider, D.L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist*, *49*, 112–119.
- Carroll, D. (1988). How accurate is polygraph lie detection? In A. Gale (Ed.), *The polygraph test: Lies, truth and science* (pp.19–28). London: Sage.
- Carroll, D. (1991). Lie detection: Lies and truths. In R. Cochrane & D. Carroll (Eds.), *Psychology and social issues: A tutorial test* (pp.160–170). London: The Falmer Press.
- Clifton, C. (1991). *Deception detection: Winning the polygraph game*. Boulder, Co: Paladin.
- Cullen, M.J. & Sackett, P.R. (2003). Personality and counterproductive workplace behavior. In M.R. Barrick & A.M. Ryan, (Eds.), *Personality and work: Reconsidering the role of personality in organisations* (pp.150–180). San Francisco: Jossey-Bass.
- Cunningham, M.R., Wong, D.T. & Barabee, A.P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology*, *79*, 643–658.
- De Paulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.
- Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. In L.W. Porter & M.R. Rosenzweig (Eds.), *Annual Review of Psychology: Vol. 41* (pp.417–440). Palo Alto: Annual Reviews.
- Ekman, P., O'Sullivan, M. & Frank, M.G. (1999). A few can catch a liar. *Psychological Science*, *10*, 263–266.
- Elaad, E. (2003). Is the inference rule of the control question polygraph technique plausible? *Psychology, Crime and Law*, *9*, 37–47.
- Emerick, R.L. & Dutton, W.A. (1993). The effect of the polygraph on the self-report of adolescent sex offenders: Implementation for risk assessment. *Annals of Sex Research*, *6*, 83–103.
- English, K., Jones, L., Pasini-Hill, D., Patrick, D. & Cooley-Towell, S. (2000). Department of Public Safety. *The value of testing in sex offender management. Research report submitted to the National Institute of Justice*. Denver: Colorado, Division of Criminal Justice, Office of Research and Statistics.
- Faigman, D. (2003). Expert evidence: The rules and the rationality the law applies (or should apply) to psychological expertise. In D. Carson & R. Bull (Eds.), *Handbook of psychology in legal contexts* (2nd ed.) (pp.367–400). Chichester: Wiley.
- Fiedler, K., Schmid, J. & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology*, *24*, 313–324.
- Ford, Ch. V. (1995). *Lies! Lies!! Lies!!! The psychology of deceit*. Washington, DC: American Psychiatric Press.
- Frank, M.G. & Ekman, P. (1997). The ability to detect deceit generalises across different types of high-stake lies. *Journal of Personality and Social Psychology*, *72*, 1429–1439.
- Goldberg, L.R., Grenier, J.R., Guion, R.M., Sechrest, L.B. & Wing, H. (1991). *Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions: An APA task force report*. Washington, DC: American Psychological Association.
- Grubin, D. (2002). The potential use of polygraph in forensic psychiatry. *Criminal Behaviour and Mental Health*, *12*, 45–55.

- Grubin, D., Madsen, L., Parsons, S., Sosnowski, D. & Warberg, B. (in press) A prospective study of the impact of polygraphy on high risk behaviours in adult sex offenders. *Sexual Abuse: A Journal of Research and Treatment*.
- Gudjonsson, G.H. (1979). The use of electrodermal responses in a case of amnesia (a case report). *Medicine, Science and the Law*, 19, 138–140.
- Gudjonsson, G.H. (1983). Lie detection: Techniques and countermeasures. In S.M.A. Lloyd-Bostock & B.R. Clifford (Eds.), *Evaluating witness evidence* (pp.137–153). Chichester: Wiley.
- Gudjonsson, G.H. (1988). How to defeat the polygraph tests. In A. Gale (Ed.), *The Polygraph Test. Truth, Lies and Science* (pp.126–136). London: Sage.
- Gudjonsson, G.H. (1992). *The psychology of interrogations, confessions, and testimony*. Chichester: Wiley.
- Gudjonsson, G.H. (2003). *The psychology of interrogations and confessions. A handbook*. Chichester: Wiley.
- Gudjonsson, G.H. & Lebeque, B. (1989). Psychological and psychiatric aspects of a coerced-internalised false confession. *Journal of the Forensic Science Society*, 29, 261–269.
- Gudjonsson, G.H. & Sartory, G. (1983). Blood-injury phobia: A 'reasonable excuse' for failing to give a specimen in a case of suspected drunken driving. *Journal of the Forensic Science Society*, 23, 197–201.
- Gudjonsson, G.H. & Sigurdsson, J.F. (2000). Differences and similarities between violent offenders and sex offenders. *Child Abuse and Neglect*, 24, 363–372.
- Gunn, J. & Gudjonsson, G.H. (1988). Using the psychological stress evaluator in conditions of extreme stress. *Psychological Medicine*, 18, 235–238.
- Hindman, J. & Peters, J.M. (2001). Polygraph testing leads to better understanding of adult and juvenile sex offenders. *Federal Probation*, 65, 8–15.
- Honts, C.R. (1991). The emperor's new clothes: The application of the polygraph tests in the American workplace. *Forensic Reports*, 4, 91–116.
- Honts, C.R. & Amato, S.L. (2002). Countermeasures. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp.251–264). London: Academic Press.
- Honts, C.R. & Perry, M.V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behaviour*, 16, 357–379.
- 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, 56–61.
- 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.R. & Hodes, R.L. (1982). The effects of multiple physical countermeasures on the physiological detection of deception. *Psychophysiology*, 19, 564 (SPR Abstract).
- Honts, C.R., Hodes, R.L. & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, 70, 177–187.
- Honts, C.R., Kircher, J.C. & Raskin, D.C. (1996). Polygrapher's dilemma or psychologist's: A reply to Furedy's logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 20, 199–207.
- Honts, C.R., Kircher, J.C. & Raskin, D.C. (2002). The scientific status of research on polygraph techniques: The case for polygraph tests. In D.L. Faigman, D. Kaye, M.J. Saks & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (Volume 2) (pp.446–483). West: St. Paul Minnesota.
- Honts, C.R., Raskin, D.C. & Kircher, J.C. (1984). Effects of spontaneous countermeasures on the detection of deception. *Psychophysiology*, 20, 446 (SPR Abstract).
- Honts, C.R., Raskin, D.C. & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252–259.
- Iacono, W.G. & Patrick, C.J. (1997). Polygraphy and integrity testing. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp.252–281). New York: The Guildford Press.
- Inbau, F.E., Reid, J.E., Buckley, J.P. & Jayne, B.C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, Maryland: Aspen Publishers.
- Kleiner, M. (2002). *Handbook of polygraph testing* (Edited book). London: Academic Press.
- Kleinmuntz, B. & Szucko, J.J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist*, 39, 766–776.
- Kokish, R. (2004). The current role of post-conviction sex offender polygraph testing in sex offender treatment. *Journal of Child Sexual Abuse*, 12, 175–194.
- Larson, J.A. (1932). *Lying and its detection: A study of deception and deception tests*. Chicago, Ill: University of Chicago Press.
- Lykken, D.T. (1998). *A tremor in the blood: Uses and abuses of lie detection* (2nd ed.). New York: Plenum.
- MacLaren, V.V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology*, 86, 674-683.
- Mann, S., Vrij, A. & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89, 137–149.
- Martin, B.A., Bowen, C.C. & Hunt, S.T. (2002). How effective are people at faking on personality questionnaires?. *Personality and Individual Differences*, 32, 247–256.
- Memon, A., Vrij, A. & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility* (2nd ed.). Chichester: Wiley.
- Murphy, K.R. (1993). *Honesty in the workplace*. Pacific Grove, CA.: Brooks/Cole Publishing.
- National Research Council (2003). *The polygraph and lie detection*. Committee to Review the Scientific Evidence on the Polygraph. Washington, DC: The National Academic Press.
- Nugent, P.M. & Kroner, D.G. (1996). Denial, response styles, and admittance of offences among child molesters and rapists. *Journal of Interpersonal Violence*, 11, 75–486.

- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.
- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17*, S19–S38.
- Podlesney, J.A. (1995). *A lack of operable case facts restricts applicability of the Guilty Knowledge Deception Detection Method in FBI criminal investigations*. FBI Technical Report, Quantico, VA.
- Raskin, D. (1990). Polygraph techniques for the detection of deception. In D.C. Raskin (Ed.), *Psychological methods in criminal investigations and evidence* (pp.247–296). Springer: New York.
- Raskin, D.C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding acceptance of polygraph evidence. *Utah Law Review, 29*, 29–74.
- Raskin, D.C. (1988). Does science support polygraph testing? In A. Gale (Ed.), *The polygraph test: Lies, truth and science* (pp.96–110). London: Sage.
- Raskin, D.C. & Honts, C.R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp.1–48). London: Academic Press.
- Reid, J.E. & Inbau, F.E. (1977). *Truth and deception: The polygraph (lie detector) technique*. Baltimore: Williams & Wilkins.
- Roberston, I.T. & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology, 74*, 441–472.
- Rosenfeld, J.P. (2002). Event-related potentials in the detection of deception, malingering, and false memories. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp.1–48). London: Academic Press.
- Sackett, P.R. & Decker, P.J. (1979). Detection of deception in the employment context: A review and critical analysis. *Personnel Psychology, 32*, 487–506.
- Salter, A. (1988). *Treating child sex offenders and victims. A practical guide*. London: Sage.
- Saxe, L. (1994). Detection of deception: Polygraph and integrity tests. *Current Directions in Psychological Science, 3*, 69–73.
- Saxe, L., Dougherty, D. & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist, 40*, 355–366.
- The value of testing in sex offender management. Research report submitted to the National Institute of Justice*. Denver: Colorado Department of Public Safety, Division of Criminal Justice, Office of Research and Statistics.
- USA Congress, Office of Technology Assessment. (1990). *The use of integrity tests for pre-employment screening* (OTA-SET-442). Washington, DC: USA Government Printing Office.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.
- Vrij, A. (in press). Committee to Review the Scientific Evidence on the Polygraph. The polygraph and lie detection. *The Howard Journal of Criminal Justice*.
- Wilcox, D. (2000). Application of the clinical polygraph examination to the assessment, treatment and monitoring of sex offenders. *Journal of Sexual Aggression, 5*, 134–152.
- Wilcox, D.T., Sosnowski, D. & Middleton, D. (2000). Polygraphy and sex offenders. *Forensic Update, 61*, 20–25.

The British Psychological Society was founded in 1901 and incorporated by Royal Charter in 1965.

Its principal objects are to:

- promote the advancement and diffusion of a knowledge of psychology pure and applied;
- promote the efficiency and usefulness of Members of the Society by setting up a high standard of professional education and knowledge;
- maintain a Code of Conduct for the guidance of Members;
- compel the observation of strict rules of professional conduct;
- maintain a Register of Chartered Psychologists.

The Society has 40,000 members and:

- has branches in England, Northern Ireland, Scotland and Wales;
- accredits nearly 700 undergraduate degrees;
- accredits nearly 100 postgraduate professional training courses;
- accredits higher degrees, in partnership with the Open University Validation Service;
- confers Fellowships for distinguished achievements;
- confers Chartered Status for professionally qualified psychologists;
- awards grants to support research and scholarship;
- publishes 10 scientific journals;
- publishes books, CD-ROMs, videos and other educational resources;
- publishes *The Psychologist* each month;

- publishes newsletters for its constituent groups;
- maintains a Web site;
- has international links with psychological societies and associations throughout the world;
- provides an information service for the news media and the public;
- has an Ethics Committee;
- provides service to the Disciplinary Board;
- maintains a Register of more than 9,000 Chartered Psychologists;
- prepares policy statements on matters of social policy;
- holds conferences, workshops, continuing professional development and training events;
- recognises distinguished contributions to psychological science and practice through individual awards and honours.

The Vision Statement of the Society plans that by 2006:

- there will be 50,000 members;
- the Society will have offices in the major constituent parts of the United Kingdom;
- the public at large will have a clear understanding of psychology as both a research and applied discipline;
- the Society will seek to influence public policy on matters relating to education, the family and the community;
- all its members will contribute so that the strengths and diversity of its membership are fully utilised.

The British Psychological Society

St. Andrews House, 48 Princess Road East, Leicester LE1 7DR, UK

Telephone 0116 254 9568 Facsimile 0116 247 0787 E-mail mail@bps.org.uk Website www.bps.org.uk