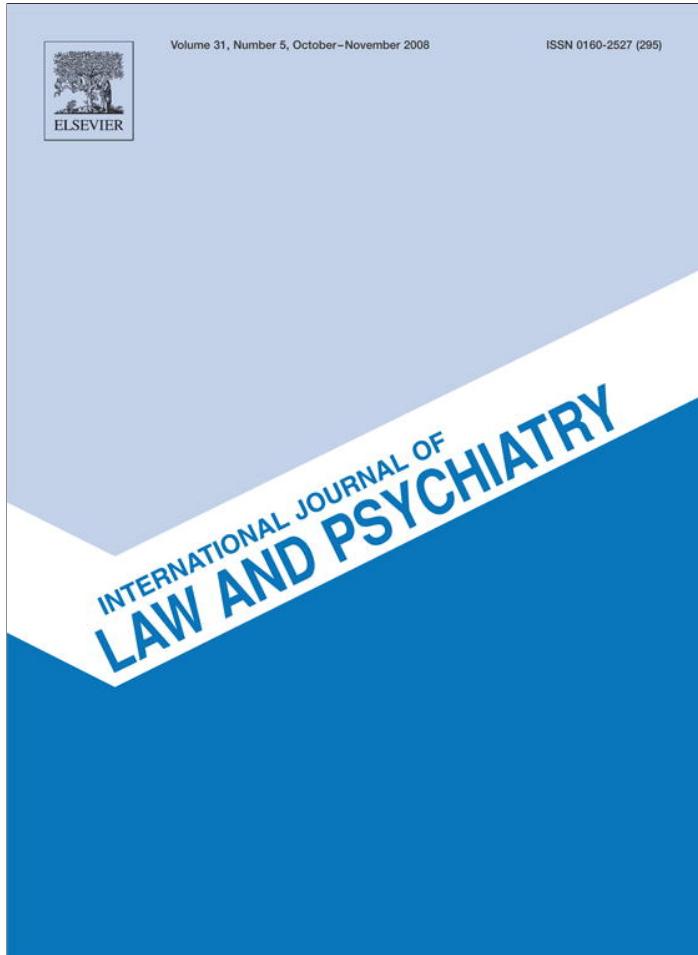


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Sex offender management using the polygraph: A critical review

Ewout H. Meijer ^{a,*¹}, Bruno Verschueren ^{b,1}, Harald L.G.J. Merckelbach ^a, Geert Crombez ^b

^a Department of Psychology, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

^b Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium

ARTICLE INFO

ABSTRACT

Reducing recidivism is a central goal in the treatment of sex offenders. In Europe, there is an increased interest in using the polygraph ("lie detector") as a tool in the treatment and risk assessment of convicted sex offenders. This interest originated from optimistic reports by American clinicians who argued that polygraph testing in the treatment of sex offenders is akin to urine analysis in the treatment of drug addiction. In this article, we critically examine the validity and utility of post-conviction sex offender polygraph testing. Our review shows that the available evidence for the claims about the clinical potential of polygraph tests is weak, if not absent. We conclude that portraying post-conviction polygraph testing as analogous to urine analysis is inaccurate, misleading, and ultimately, risky.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Recidivism among sex offenders poses a serious problem. Data from the United States [Bureau of Justice Statistics \(2003\)](#), show that 5% of about 10,000 sex offenders who were released from prison in 1994 were rearrested for a sex crime within 3 years. At 43%, rearrests for any type of crime were substantially higher. A review of 61 studies ($n=23,393$) demonstrated that sexual recidivism was 13% after a mean follow-up of 4–5 years ([Hanson & Bussiere, 1998](#)). Recidivism for any reoffense was 36%. Moreover, these statistics are an underestimation of true recidivism rates. Follow-up periods during which recidivisms are monitored are by definition restricted, and not all crimes lead to rearrest or reconvictions ([Kilpatrick, Edmunds, & Seymour, 1992](#)). Unfortunately, highly effective treatment programs do not yet exist. Meta-analyses on the efficacy of sex offender treatment programs have shown that cognitive-behavioral therapy can decrease recidivism, but that this decrease is only modest ([Hall, 1995](#); [Hanson et al., 2002](#)). More recent and more carefully designed studies failed to find any treatment effects at all ([Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005](#); [Schweitzer & Dwyer, 2003](#)).

Sex offenders often deny the seriousness and impact of their crimes. Embracing a more truthful and complete report of their crimes is considered an important aspect of treatment. Risk assessment would also benefit from offenders' truthful self-report about their criminal history and pathology. Contemporary risk assessment is based on actuarial risk predictors, which are risk factors derived from empirical research ([Hanson, 2003](#)). Hanson and Bussiere's meta-analysis showed that sexual offense recidivism was best predicted by measures of sexual deviancy (e.g., deviant sexual preferences, prior sexual offenses) and, to a lesser extent, by general criminological factors (e.g., age, total prior offenses). Although actuarial risk assessment has improved the prediction of recidivism compared to clinical judgement ([Dawes, Faust, & Meehl, 1989](#); [Grove & Meehl, 1996](#); [Grove, Zald, Lebow, Snitz, & Nelson, 2000](#); [Quinsey, Harris, Rice, & Cormier, 2006](#)), this prediction is not faultless ([Hanson, 2003](#); [Harris et al., 2003](#)). Langton et al. (2007) assessed the predictive ability of 6 contemporary actuarial instruments in 468 sex offenders who were followed for an average of 5.9 years. Predictive accuracy was expressed in terms of the area under the Receiver Operating Curve

* Corresponding author. Tel.: +31 433884505; fax: +31 433884196.

E-mail address: eh.meijer@psychology.unimaas.nl (E.H. Meijer).

¹ The first two authors contributed equally to the manuscript and are listed in alphabetical order. Bruno Verschueren is a post-doctoral fellow of the Scientific Research Foundation – Flanders, Belgium (FWO).

(ROC *a*). This statistic can be interpreted as the probability that a randomly chosen sex offender who reoffended obtained a higher score than a sex offender who did not. Values range from 0 to 1, with .5 reflecting chance level. The results showed that for all instruments, accuracy was significantly above chance. With ROC *a* values ranging from .57 to .72 for any crime, and from .61 to .71 for sex crimes, accuracy was, however, only modest.

Offender's records are used to score actuarial risk assessment instruments, and, although objective, this information is often incomplete. As an example, the number of prior sex offenses present in the records is probably an underestimation of the true number of sex offenses. Relying exclusively on file information, thus, poses a serious problem, and good practice requires corroborating offender's self-report. This, in turn may prove problematic as offenders have a vested interest in concealing their true behavior, as good behavior is often rewarded with such benefits as parole. In this context, some authors have referred to offenders' fake good tendencies as "supernormality" (Cima et al., 2003). This supernormality is a major obstacle for accurate risk assessment. All together, both treatment and risk assessment in sex offenders could benefit from more honest self-report.

In this review, we critically evaluate a method for verifying the veracity of offenders' self-reports; post-conviction sex offender polygraph testing (PCSOT). This method has steadily gained popularity among clinicians over the last decade, and was used in 70% of community sexual abuser programs for adult males in 2002 in the United States (McGrath, Cumming, & Burchard, 2003). Proponents argue that this method contributes to a more accurate and more complete sexual history taking, more accurate offense descriptions, superior assessment of high-risk behaviors, and better compliance with probation conditions. The popularity of the method among practitioners is illustrated in the opinion expressed by the former chair of the American Polygraph Association: "Those of us who conduct post-conviction sex offender testing know that it has been the missing link in preventing recidivism (Consigli, 2002, p. 239)". American practitioners have presented the role of the polygraph in the treatment of sex offenders as similar to that of urine analysis in the treatment of drug addicts (English, Jones, Patrick, & Pasini-Hill, 2003, p. 414). Both are viewed as objective and accurate methods to verify whether the offender's reports are truthful. These positive claims received much attention, and PCSOT was recently introduced in Europe. In the United Kingdom and the Netherlands, highly secure treatment hospitals have started PCSOT studies. European users appear to share Consigli's enthusiasm, promoting widespread use. At the time of writing, in the United Kingdom, the Management of Offenders and Sentencing Bill was in progress, extending the use of polygraph tests for sex offenders on license.

In this review, we discuss whether this enthusiasm is justified. We will first describe the general principles of polygraph testing and how it is used in PCSOT. We will then critically examine the results from PCSOT studies. Finally, we will discuss whether there is any evidence for the notion that PCSOT reduces recidivism, and whether PCSOT is, indeed, the urine test of sex offender treatment.

2. The polygraph and the detection of deception

The type of polygraph test used in PCSOT is a version of the Comparison Question Technique (Grubin, Madsen, Parsons, Sosnowski, & Warberg, 2004). This type of test is also known as the Control Question Test (CQT; National Research Council, 2003; Reid, 1947), and is best known from its use by law enforcement agencies in criminal investigations. During a CQT, relevant and comparison questions are asked. The relevant questions are directly related to the incident under investigation (e.g., "On the 14th of May, did you kill X?"). The comparison questions are emotionally provocative, and deal with undesirable behavior that is unrelated to the crime itself (e.g., "During the last 25 years, have you ever done anything illegal?"). Physiological responses, including skin conductance response, respiration and blood pressure are monitored throughout. The test procedure is based on the assumption that the questions which pose the biggest threat to the individual elicit the strongest physiological responses. It is further assumed that the relevant questions are most threatening for guilty suspects, whereas comparison questions form the biggest threat to innocent suspects. This latter assumption requires some explanation. A CQT is preceded by a lengthy interview, in which the examinee is led to believe that the polygraph is almost infallible. Thus, the innocent examinee can confidently and honestly answer "no" to the relevant questions. Furthermore, the examinee is led to believe that a deceptive answer to any of the questions will lead to a 'deceptive' test outcome. At the same time, the examinee is pressured into answering "no" to the comparison questions by the examiner suggesting that confessing illegal activities will negatively influence his/her opinion (see e.g., Offe & Offe, 2007). As a result, the innocent examinee is assumed to show the strongest physiological responses to the comparison questions, fearing that his/her deceptive answer to this question will get him/her convicted for the crime under investigation.

3. Polygraph validity

Ever since its introduction, the CQT has been the object of an intense debate, and remains controversial (Ben Shakhar, 2002; Ben Shakhar & Furedy, 1990; Fiedler, Schmid, & Stahl, 2002; Furedy, 1996a,b; Honts, Kircher, & Raskin, 1995; Lykken, 1998; Raskin & Honts, 2002). The debate focuses on the assumptions underlying the CQT, thus calling into question the test's accuracy. Many critics have argued that it is impossible to determine the origins of the physiological responses. Stronger responses to the relevant questions may indicate a guilty suspect's fear of being detected, but may also indicate an innocent suspect's fear of a wrongful conviction (Lykken, 1998). Indeed, whether an innocent examinee will be worried most by the comparison questions heavily depends on the skills of the polygrapher. Additionally, the detection accuracy of the CQT is hard to establish. In laboratory studies, participants are not threatened by severe consequences of failing a test. The studies therefore lack ecological validity. Field studies, on the other hand, suffer from a lack of "ground truth", i.e., conclusive exonerating or incriminating evidence that can corroborate test outcome (Iacono, 1991).

A large number of studies addressed the accuracy rates of the CQT in specific-incident testing (e.g., tests in criminal investigations). Mock crime study estimates range from 74–80% for guilty participants, and 63–66% for innocent participants (Ben Shakhar & Furedy, 1990; Kircher, Horowitz, & Raskin, 1988). The false positive ratio (i.e., innocent examinees classified as guilty) was estimated at 12–15%, and the false negative ratio (i.e., guilty examinees classified as innocent) was 7–8%. Field study estimates of the accuracy of the CQT are slightly higher, ranging from 84–89% for guilty, and 59–72% for innocent suspects, with estimated false positive rates of 12–23% and false negative rates of 1–13% (Ben Shakhar & Furedy, 1990; Raskin & Honts, 2002).

Recently, the National Research Council (NRC, 2003) evaluated the validity of the CQT for screening purposes. These tests are used in the United States to screen job applicants and monitor employees of government agencies whose work involves security risks (e.g., FBI applicants or nuclear scientists; Krapohl, 2002). When reviewing the evidence, the NRC concluded that there were no high quality studies on the use of the CQT for screening purposes. Therefore, the council based its review on studies that addressed the validity of the CQT in specific-incident testing. Reviewing this evidence, the NRC noted that most studies do not reach high levels of scientific quality. For example, nearly all studies are based on naïve participants, threatening ecological validity. Still, the Council selected 37 laboratory studies and 7 real-life field studies that met the minimum criteria for its analysis. The Council concluded that specific-incident polygraph tests can discriminate lying from truth telling at rates well above chance, though well below perfection.² Furthermore, the Council pointed out that accuracy for screening purposes is almost certainly lower than what can be achieved for specific-incident testing. The latter is especially relevant because the use of the CQT in PCSOT predominantly³ involves questioning sex offenders about unknown incidents, thus bearing a stronger resemblance to screening than to specific incidence testing (Abrams & Abrams, 1993). The main difference between specific incidence and screening tests is that during a specific-incident polygraph test, a suspect is questioned about a *known* incident, such as a murder, theft or arson. In a screening test, however, someone is questioned about incidents of which it is *unknown* whether they have taken place, such as disclosing classified information.

4. The polygraph and sex offenders

An important reason for lower accuracy in screening applications than in specific-incident testing is that the former involves considerably more ambiguity for the examinee (National Research Council, 2003). This becomes clear when looking at the function of the CQT in PCSOT more closely. PCSOT intends to verify self-reported information about past or current behavior. Firstly, it entails verifying the accuracy and completeness of the sexual history information provided by the offender. This "sexual history disclosure test" is generally administered at the beginning of the treatment. Secondly, it entails verifying whether a probationer or parolee is complying with the conditions of community supervision and is cooperating with treatment expectations (English et al., 2003). Such "maintenance tests" are administered periodically, for example every 6 months (Abrams & Abrams, 1993). Thus, in contrast to specific-incident testing, the offender is questioned about incidents of which it is unknown whether they have taken place at all. This means that the relevant questions need to be phrased in a very broad way. Consequently, relevant questions (e.g., "Have you had unsupervised contact with children over the last 3 months?") and comparison questions (e.g., "Have you done anything over the last 3 months that would concern your probation officer?"; Grubin et al., 2004 p. 213) become more similar. As a result, test outcomes will be more ambiguous and diagnostic decisions less accurate.

We found no studies that made a serious attempt to estimate the validity of the CQT in PCSOT. The only studies that tried to estimate it limited themselves to offenders' self-reports of test accuracy. Kokish, Levenson, and Blasingame (2005), for example, collected self-report data from 95 sex offenders who underwent a total of 333 tests. Eighteen offenders (19%) reported having been incorrectly labeled deceptive, and 6 (6%) claimed they had incorrectly been found truthful. Similar statistics were reported by Grubin and Madsen (2006). They collected self-report data from 126 offenders which completed a total of 263 tests. Of these 126 offenders, 27 (21%) reported having been incorrectly found deceptive, and 6 (5%) reported they had been wrongly labeled truthful.

Although such results might look encouraging, these studies suffer from a major methodological pitfall, namely sampling bias. Participation in a polygraph treatment program is voluntary. In a prospective study by Grubin et al. (2004), only 21 out of the 116 offenders (18%) that were initially approached, completed both polygraph tests that were part of the study. Accuracy rates based upon this small sample of offenders may thus very well be an overestimation. Offenders that were confronted with an incorrect test result may have simply dropped out.

Clearly, validity research on the CQT in PCSOT is needed. Apart from the ambiguity of the relevant questions, there are a number of other factors that specifically threaten PCSOT validity (see also Branaman & Gallagher, 2005; Cross & Saxe, 1992, 2001; Faller, 1997). These factors are related to cognitive distortions of sex offenders, and the effects of repeated testing.

Pertaining to cognitive distortions, sex offenders are known to justify their sexual involvement with children in diverse ways. They may minimize harm caused by the abuse, perceive children as desiring sexual contact with adults, or perceive their sexual contact as socially acceptable (Gannon & Polaschek, 2006; Ward, Hudson, Johnston, & Marshall, 1997). In specific-incident testing, the effects of cognitive distortions can be minimized by asking very specific and straightforward questions. For example, the

² The National Research Council did not report accuracy in terms of percentage correct decisions. This is because percentage correct decisions rely on an arbitrary cutoff point. The choice of where the cutoff is placed depends on the preference to reduce either the false positive ratio or the false negative ratio. Rather, the National Research Council expressed accuracy in terms of the area under the Receiver Operating characteristic Curve (ROC a). They report a median ROC a of .86 for the CQT in specific-incident testing.

³ There are some exceptions to this rule: Some authors (e.g., Abrams & Abrams, 1993; English et al., 2003) also report the use of specific-incident testing in the treatment of sex offenders, for example, when the offender continues to deny committing the crime of conviction.

suspect might be asked "On the night of July first 1994, did you put your penis into Sarah's vagina?" As argued earlier, in PCSOT one cannot ask questions as specific as this because the examiner does not know whether deviant behavior actually took place. The relevant questions are therefore formulated more vaguely; e.g., "Since July first 1994, have you had any sexual contact with a minor?" (Hagler, 1995 p. 107). This type of question is problematic when dealing with sex offenders, because sexual contact is not explicitly defined. It allows interpretation, and in the distorted mind of the sex offender, "sexual contact" per se may not have taken place. It may, for example, have been labeled as normal father to daughter interaction. A sex offender who has cognitive distortions of this type may therefore not respond to relevant questions. It is easy to see how this may contribute to false negative test outcomes.

The second factor that threatens CQT accuracy in PCSOT is repeated testing of the same offender. This can deteriorate the test outcome for a number of reasons. To begin with, it is well known that physiological responsivity decreases upon repeated presentation of the same stimulus (i.e., habituation; Andreassi, 2000). Thus, the "guilty" sex offender may not show a marked physiological response to the relevant question after having been repeatedly confronted with it, thereby raising the probability of a false negative test outcome. Repeated testing is also likely to reduce the emotionally provocative loading of the comparison question, thereby increasing the probability of a false positive test result in "innocent" sex offenders. Furthermore, repeated testing increases the probability of the effective use of countermeasures by the offender. Countermeasures refer to everything the examinee can do in order to alter the test outcome (Honts & Amato, 2002). The examinee might use physical countermeasures such as biting of the tongue in order to create a physical response to the comparison question. Alternatively, he or she might try to alter physiological responses using mental countermeasures, by thinking about something exciting during the comparison question, for example. Several studies have shown that mental countermeasures are especially problematic because they are not easily detected by the examiner (for a review, see Honts & Amato, 2002). Honts (2004) has suggested that countermeasures may only be effective if the examinee is able to practice them. Clearly, repeated testing not only familiarizes the examinee with the procedure, but also gives the examinee the opportunity to practice countermeasures.

In sum, the validity of the CQT has been, and still is, debated. The accuracy of this test for specific-incident testing is far from perfect, and when used in PCSOT, matters worsen even more. High-quality research on the use of the CQT for screening purposes is non-existent, and there is ample reason to assume that its accuracy will not reach the level obtained in specific-incident testing. Furthermore, several factors (i.e., cognitive distortions, repeated testing and countermeasures) seriously threaten its validity in PCSOT.

5. Disclosure through polygraph testing

Some have argued that the polygraph has its utility, regardless of its accuracy. "I am less concerned about the reliability of polygraph. The fact is that every time I refer a client for polygraph testing, I end up with significantly more information than I started with (Veeder, 1995, in Hagler, 1995)". Several reports have recently flourished, suggesting that the polygraph is highly successful in obtaining previously undisclosed information (e.g., Ahlmeyer, Heil, McKee, & English, 2000; Emerick & Dutton, 1993; English et al., 2003). This includes information on the number of previous offences, age of the first offence, number of victims and prevalence of high-risk behaviors.

In this particular respect, recent European PCSOT pilot studies also report encouraging results. Among them, Wilcox and Sosnowski (2005) describe the results of 14 sex offenders on probation who were administered a single CQT. Important new information was obtained from all offenders: 93% of the sample admitted they had committed both contact (e.g., indecent assault) and non-contact (e.g., voyeurism) sexual offenses, compared to the known 29% in the probation records. In addition, the offenders reported earlier onset of offending and a wider range of paraphilic interests than had previously been reported. The prospective study by Grubin et al. (2004) also demonstrated that 97% of the offenders disclosed high-risk behavior at their first test, and 71% did at the second test. Preliminary results from a Dutch pilot study also showed an increase in the disclosure of number of offenses, high-risk behavior and paraphilic interest in a sample of 23 men convicted of child pornography possession (Buschman, personal communication, October 28, 2006).

Overall, there is little doubt that PCSOT leads to an increase in the amount of information disclosed by offenders. There are, however, a number of methodological points that qualify the optimistic interpretation that one might give to this effect of polygraph testing. First, as discussed earlier, participation in a treatment program involving PCSOT is voluntary. Thus, the offenders included in the studies cited above probably represent a subgroup of highly motivated (and perhaps quickly disclosing) offenders. Second, the ground truth is unknown. Thus, it remains unknown how much information has gone undetected and how accurate the newly disclosed information is. As acknowledged by several authors (Grubin & Madsen, 2006; Kokish, Levenson, & Blasingame, 2005), offenders might have fabricated stories after deceptive test outcomes, in order to satisfy examiners or to obscure the actual reason for failing the test. Finally, all studies so far lack a proper control condition, making it impossible to draw firm conclusions on the utility of the polygraph per se. The additional value of PCSOT in obtaining new information can only be established by comparing its effect to that of a control condition such as simply interviewing offenders. Researchers have now compared information obtained through PCSOT solely with what was already known from the offender's files. In this context, any newly disclosed information is reported as evidence for the success of PCSOT. Yet, this approach remains silent about whether a thorough interview might be sufficient to obtain this information and whether the polygraph is additionally needed. Thus, the superiority of PCSOT can only be demonstrated by comparing it to another method that has a similar intensity (i.e., similar amount of time spent with the examinee). Only such a comparison could demonstrate the additional value of PCSOT in obtaining new information. If future research were to demonstrate that this is the case, then the question arises as to where this effect comes from. There is ample reason to assume that it would have to do with the bogus pipeline effect.

6. Bogus pipeline

There is a long tradition in social psychology to develop methods that aim to reveal a person's true opinion on sensitive topics. Standard paper-and-pencil tests are believed to be unreliable, because they are subject to response biases such as social desirability (Crowne & Marlowe, 1964). One approach social psychologists have used to circumvent social desirability is attaching the person to an impressive but useless machine. People respond more truthfully when they are led to believe that this "bogus pipeline" can reveal the actual truth. Hooked up to a fake lie detector, people have been found to be more honest about social prejudices, sexual fantasies, and addictive behaviors (Roe & Jamieson, 1993). Likewise, the capability of the polygraph to obtain new confessions in sex offenders is probably more related to this bogus pipeline effect than to its accuracy. Some proponents of sexual history disclosure tests have acknowledged this. According to Abrams and Abrams (1993), there are three points in time when sex offenders can disclose information: (1) when they are told that they will face a test in the near future, (2) during the pre-test interview, and (3) during the confrontation with the test outcome. Note that the first two points are before the polygraph test is actually conducted. Indeed, Grubin et al. (2004) found that most information was disclosed during the pre-test interview, well before the sensors of the polygraph were actually connected. Apparently, the expectation of an upcoming polygraph test is sufficient to make offenders disclose information. This means that disclosure of new information may have little to do with the polygraph as a method for the detection of deception per se. More likely, it is the questioning and the intimidation by the lie detector that makes him or her bring up new information. Recent empirical data supports the bogus pipeline effect in sex offenders. Gannon, Keown, and Polaschek (2007) asked 41 child molesters to what extent they agreed with cognitive distortions such as "Sometimes, touching a child sexually is a way to show love and affection" (MOLEST; Bumby, 1996). All items were read out loud and the participants gave a verbal response. Four to six weeks later, this procedure was repeated, but now 18 participants were attached to a bogus lie detector when answering the items. The remaining 23 participants formed a control condition in which the procedure was identical to the first administration. Results showed that child molesters in the bogus pipeline condition, but not in the control condition, reported more cognitive distortions than in the first administration, confirming the success of lie detectors in obtaining more honest self-reports, regardless of their accuracy. Whether the polygraph should be used to extract confessions regardless of demonstrated accuracy is an ethical dilemma.

Another ethical dilemma PCSOT faces is that sex offenders are required to be completely honest for successful completion of their treatment program. This requirement stands in sharp contrast to the position of the clinician who, when using the CQT is relying on a test that is based on deception. That is, the examinees are erroneously told that a deceptive answer to one of the comparison questions will lead to a deceptive test outcome (Offe & Offe, 2007). When discovered, this manipulative behavior of the clinician may seriously affect the therapeutic relationship.⁴

7. Reducing recidivism

The most critical question is not to what extent PCSOT leads to the disclosure of information, but whether it reduces recidivism. Such reduction could be either direct, due to the deterrent effect of an upcoming polygraph on behavior, or indirect through increased disclosure improving risk assessment and/or treatment outcome. We will argue that there is evidence against the former, and no evidence for the latter.

PCSOT has been used since the 1960s, and has been described as the missing link in preventing recidivism. Going through all relevant scientific databases (i.e., Psychinfo, Medline, Web of Science), however, we could not identify a single study that directly addressed the critical issue of whether PCSOT actually reduces recidivism. There is little reason to assume that PCSOT will have a direct effect on recidivism. Consider, for example, the study by Grubin and colleagues. These authors investigated whether the prospect of a polygraph test would influence offenders' behaviors. Clearly, the authors expected that the prospects of a polygraph test would result in a reduction in high-risk behavior (Grubin et al., 2004). One-hundred and sixteen convicted sex offenders were approached, of which less than half ($n=50$) agreed to participate. High-risk behaviors were identified for each individual and two groups were created. One group was told that they would undergo a polygraph test in 3 months, whereas the other group was told their behavior would be reviewed, with no specific reference to the use of the polygraph. In fact, both groups were subjected to a polygraph test 3 months later. Thirty-two of the original 50 offenders reported for the polygraph test. Thirty-one of them (97%) disclosed having engaged in high-risk behavior, with no differences between the polygraph aware and the polygraph unaware group. For a second polygraph test, performed 3 months later, another 11 offenders dropped out. Of the remaining 21 offenders 15 (71%) disclosed high-risk behavior. This study shows that although PCSOT is able to reveal new information in a highly selected sample of motivated sex offenders (18% of the original sample completed both tests), the knowledge of an upcoming polygraph test does not seem to prevent offenders from engaging in high-risk behavior.

⁴ Besides the ethical issues, PCSOT also faces legal issues. The question arises on how to handle criminally-relevant information that is disclosed by the offender. There are two options: either the offender is prosecuted or he/she gets immunity. When prosecuting the offender for the newly disclosed information, it is likely that defense attorneys will challenge the charge because it is based on self-incriminatory evidence that violates the right to remain silent and/or the right to have council present during interrogative questioning (see Malesky & Packer, 2004). Likely, prosecuting will imply the end of the treatment in most cases. Moreover, it is likely that offenders who risk being charged will not participate in polygraph-supported treatment programs. The other possibility is to promise the sex offender immunity, in that no legal consequences will be attached to the newly disclosed criminal information. This is the strategy that is most often chosen (Schwartz & Cellini, 1995). However, this strategy may prove very difficult to explain to the victims of the disclosed offences.

8. Concluding remarks

Treatment of sex offenders is a serious issue. Not only does recidivism occur frequently, the public in many countries is highly sensitive to the topic (Lösel, 2000). With this in mind, it is understandable that practitioners resort to unorthodox techniques such as polygraph testing. At present, however, there is no evidence supporting the accuracy of the CQT in PCSOT. Relevant theory and research suggests that this accuracy will be lower than that obtained in criminal investigations. Users have argued that accuracy is less important than the utility of the polygraph in obtaining disclosures. We have argued that this effect has little to do with the polygraph as a diagnostic test, and can largely be accounted for by its intimidating effect. Moreover, this effect does not deter sex offenders from engaging in high-risk behavior.

Proponents of PCSOT have argued that it is akin to urine analysis in the treatment of drug addiction. It should be clear from our review that this analogy is seriously flawed. Unlike urine analysis, PCSOT involves deception of the examinee. PCSOT is not as standardized as urine analysis, and test scoring of PCSOT protocols requires far more interpretation. Unlike urine analysis, accuracy of the CQT in PCSOT is unknown and it is highly likely that errors will frequently occur. If false positives occur, the false accusations that follow from these may be highly damaging to the therapeutic relationship. Even more problematic is that some sex offenders at risk for recidivism will be pled free by the polygraph.

PCSOT advocates seem to have unrealistic confidence in the polygraph. This confidence creates the possibility for offenders to engage in sex crimes under the safe conduct of a "non-deceptive" polygraph result. Therefore, we discourage the use of sanctions and privileges for deceptive and non-deceptive polygraph results, as proposed by Ahlmeyer et al. (2000). The occurrence of errors is especially problematic because, unlike urine drug analysis, accurate retesting is impossible with PCSOT. We seriously doubt that clinicians would use urine analysis in the treatment of drug addiction if it would have the same characteristics as PCSOT.

References

- Abrams, S., & Abrams, J. B. (1993). *Polygraph testing of the pedophile*. Portland: Ryan Gwinner Press.
- Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 12, 123–138.
- Andreassi, J. L. (2000). *Psychophysiology: Human behavior and physiological response*, 4th ed. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ben Shakhar, G. (2002). A critical review of the Control Questions Test (CQT). In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 103–126). San Diego: Academic Press.
- Ben Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Springer-Verlag Publishing.
- Branaman, T. F., & Gallagher, S. N. (2005). Polygraph testing in sex offender treatment: A review of limitations. *American Journal of Forensic Psychology*, 23, 45–64.
- Bumby, K. M. (1996). Assessing the cognitive distortions in child molesters and rapists: Developments and validation of the MOLEST and RAPE scales. *Sexual Abuse: A Journal of Research and Treatment*, 8, 37–54.
- Bureau of Justice Statistics (2003). *Recidivism of sex offenders released from prison in 1994*. Washington, D.C.: U.S. Department of Justice.
- Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., et al. (2003). The other side of malingering: Supernormality. *Clinical Neuropsychology*, 17, 235–243.
- Consigli, C. J. E. (2002). Post-conviction sex offender testing and the American Polygraph Association. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 237–250). San Diego, CA, US: Academic Press.
- Cross, T. P., & Saxe, L. (1992). A critique of the validity of polygraph testing in child abuse cases. *Journal of Child Sexual Abuse*, 1, 19–33.
- Cross, T. P., & Saxe, L. (2001). Polygraph testing and sexual abuse: The lure of the magic lasso. *Child Maltreatment*, 6, 195–206.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: John Wiley.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668–1674.
- Emerick, R. L., & Dutton, W. A. (1993). The effect of polygraphy on the self report of adolescent sex offender: Implications for risk assessment. *Annals of Sex Research*, 6, 83–103.
- English, K., Jones, L., Patrick, D., & Pasini-Hill, D. (2003). Sexual offender containment: Use of the postconviction polygraph. *Annuals New York Academy of Sciences*, 989, 411–427.
- Faller, K. C. (1997). The polygraph, its use in cases of alleged sexual abuse: An exploratory study. *Child Abuse and Neglect*, 21, 993–1008.
- Fiedler, K., Schmid, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection. *Basic and Applied Social Psychology*, 24, 313–324.
- Furedy, J. J. (1996). The North American polygraph and psychophysiology: Disinterested, uninterested, and interested perspectives. *International Journal of Psychophysiology*, 21, 97–105.
- Furedy, J. J. (1996). Some elementary distinctions among, and comments concerning, the 'control' question 'test' (CQT) polygrapher's many problems: A reply to Honts, Kircher and Raskin. *International Journal of Psychophysiology*, 22, 53–59.
- Gannon, T. A., Keown, K., & Polaschek, D. L. (2007). Increasing honest responding on cognitive distortions in child molesters: The bogus pipeline revisited. *Sex Abuse*, 19, 5–22.
- Gannon, T. A., & Polaschek, D. L. L. (2006). Cognitive distortions in child molesters: A re-examination of key theories and research. *Clinical Psychology Review*, 26, 1000–1019.
- Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.
- Grubin, D., & Madsen, L. (2006). Accuracy and utility of post-conviction polygraph testing of sex offenders. *British Journal of Psychiatry*, 188, 479–483.
- Grubin, D., Madsen, L., Parsons, S., Sosnowski, D., & Warberg, B. (2004). A prospective study of the impact of polygraphy on high-risk behaviors in adult sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 16, 209–222.
- Hagler, H. L. (1995). Polygraph as a measure of progress in the assessment, treatment, and surveillance of sex offenders. *Sexual Addiction and Compulsivity*, 2, 98–111.
- Hall, G. C. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, 63, 802–809.
- Hanson, R. K. (2003). Who is dangerous and when are they safe? Risk assessment with sexual offenders. In B. J. Winick & J. Q. La Fond (Eds.), *Protecting society from sexually dangerous offenders: Law, justice, and therapy* (pp. 63–74). Washington, DC: American Psychological Association.
- Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348–362.
- Hanson, R. K., Gordon, A., Harris, A. J., Marques, J. K., Murphy, W., Quinsey, V. L., et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sex Abuse*, 14, 169–194 discussion 195–167.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413–425.

- Honts, C. R. (2004). The psychophysiological detection of deception. In P. A. Granhag & L. A. Stroemwall (Eds.), *The detection of deception in forensic contexts* (pp. 103–126). Camebridge: Camebridge University Press.
- Honts, C. R., & Amato, S. L. (2002). Countermeasures. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 251–264). San Diego: Academic Press.
- Honts, C. R., Kircher, J. C., & Raskin, D. C. (1995). Polygrapher's dilemma or psychologist's chimaera: A reply to Furedy's logico-ethical considerations for psychophysiological practitioners and researchers. *International Journal of Psychophysiology*, 20, 199–207.
- Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests. In J. R. Jennings, P. K. Ackles, & M. G. H. Coles (Eds.), *Advances in psychophysiology*, Vol. 4 (pp. 201–207). London: Jessica Kingsley Publishers.
- Kilpatrick, D. G., Edmunds, C. N., & Seymour, A. K. (1992). Rape in America: A report to the nation. Arlington, VA: National Center for Victims of Crime. Charleston, SC: Medical University of South Carolina.
- Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the Control Question Polygraph Technique. *Law and Human Behavior*, 12, 79–90.
- Kokish, R., Levenson, J. S., & Blasingame, G. D. (2005). Post-conviction sex offender polygraph examination: Client-reported perceptions of utility and accuracy. *Sexual Abuse: A Journal of Research and Treatment*, 17, 211–221.
- Krapohl, D. J. (2002). The polygraph in personnel screening. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 217–236). San Diego: Academic Press.
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior*, 34, 37–59.
- Lösel, F. (2000). The efficacy of sexual offender treatment: A review of German and international evaluations. In P. J. van Koppen & N. Roos (Eds.), *Rationality, information and progress in psychology and law* (pp. 145–170). Maastricht: Metajuridica Publications.
- Lykken, D. T. (1998). *A Tremor in the Blood*. Reading: Perseus Publishing.
- Malesky, L. A., & Packer, I. K. (2004). Sex offender laws. *Journal of the American Academy of Psychiatry and the Law*, 32, 209–211.
- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommen, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's sex offender treatment and evaluation project (SOTEP). *Sex Abuse*, 17, 79–107.
- McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey*. Brandon, VT: Safer Society Press.
- National Research Council (2003). *The Polygraph and Lie Detection*. Committee to review the scientific evidence on the polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academic Press.
- Offe, H., & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior*, 31, 291–303.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 1–47). San Diego: Academic Press.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542–547.
- Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and analysis. *Psychological Bulletin*, 114, 363–375.
- Schwartz, & Cellini (1995). *The sex offender: Corrections, treatment and legal practice*. Kingston, NJ: Civic Research Institute.
- Schweitzer, R., & Dwyer, J. (2003). Sex crime recidivism: Evaluation of a sexual offender treatment program. *Journal of Interpersonal Violence*, 18, 1292–1310.
- Veeder, G. (1995). *Developing standards for polygraph testing*. Lecture at the Colorado Polygraph Association Meeting, Denver (January 21).
- Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review*, 17, 479–507.
- Wilcox, D. T., & Sosnowski, D. E. (2005). Polygraph examination of British sexual offenders: A pilot study on sexual history disclosure testing. *Journal of Sexual Aggression*, 11, 3–23.