
The (F)utility of Post-Conviction Polygraph Testing

Sexual Abuse: A Journal of
Research and Treatment
XX(X) 1-23
© The Author(s) 2012
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1079063212455668
<http://sax.sagepub.com>



Jeffrey W. Rosky¹

Abstract

The apparent utility of the polygraph to work both as a treatment and supervision aid and as a deterrent for future offending is cited as ample justification for its use. This article examines these claims to demonstrate that although post-conviction polygraph testing may have some utility by increasing disclosures of prior offending and, within specific cases, admissions of treatment and supervision violations, the limited evidence accumulated thus far does not adequately ascertain its accuracy nor support its efficacy or effectiveness as a deterrent. The article concludes with recommendations for creating a real evidentiary base beyond polygraph testing's apparent ability to elicit more information from offenders to evidence that can determine whether it is efficacious and effective in reducing criminality and deviance.

Keywords

polygraph, recidivism, sexual offender treatment, evaluation, assessment

Introduction

Within the past 35 years, post-conviction use of the polygraph has been expanding (Abrams & Ogard, 1986; Abrams & Simmons, 2000), most notably as part of supervision of sex offenders due to the repeated nature of their crimes and the secrecy they use to aid in their offending (Seto, 2004). In a 2000 survey of U.S. state probation and parole departments more than half of all respondents regularly used polygraph testing to monitor sex offenders (English, Jones, Pasini-Hill, Patrick, & Cooley-Towell, 2000). Faigman, Fienberg, and Stern (2003) reported that polygraph testing to monitor sex offenders is required in more than 30 states in the United States. More recently, McGrath, Cumming, Burchard, Zeoli, and Ellerby (2010) noted that the use of post-

¹University of Central Florida, Orlando, FL, USA

Corresponding Author:

Jeffrey W. Rosky, Department of Criminal Justice, College of Health & Public Affairs, University of Central Florida, HPA1-323, Orlando, FL 32816, USA

Email: jrosky@ucf.edu

conviction polygraph in sex offender programs in the United States stood at 79% for adults and 50% for adolescents.

Within the context of post-conviction sex offender supervision, advocates argue that there are three main benefits (i.e., utility) to be gained in the use of post-conviction polygraph testing by correctional agencies: (a) A significant increase in the reporting of past criminal histories by an offender including both sexual and nonsexual offenses and wider array of victims; (b) better assessment of an offender's therapeutic progress including identification of the offender's main risk factors; and (c) a deterrent effect provided by the polygraph itself in preventing both supervision technical violations and new offenses (California Coalition on Sexual Offending, 2004; Heil & English, 2009; Kokish, 2003; La Fond & Winnick, 2003; Levenson, 2009). Moreover, post-conviction polygraph testing is claimed to be an evidence-based intervention that aids in the supervision of sex offenders (Levenson, 2009). But what does the evidence really say about the validity, accuracy, efficacy, and effectiveness of post-conviction polygraph testing?

The Validity of Polygraph Testing

The basic assumption of any good test or measure is that it has construct validity; that is it is actually measuring the phenomena it was designed to capture (Maxfield & Babbie, 2011). There is no argument that the instrumentation used in polygraph testing is actually measuring blood pressure, breathing, heart rate, and perspiration; rather, it is the assumption within polygraph testing that if a subject shows some physiological response assumed to be related to deception during the polygraph examination, then the subject is deceptive. As many authors (Crosse & Saxe, 1992; Crosse & Saxe, 2001; Faigman, et al, 2003; Fienberg & Stern, 2005; Iacono, 2001; Lykken, 1998; Seto, 2004) have noted, it requires a logical leap to assume that the response is due solely to deception because this response can also be attributed to fear, anxiety, anger, and many medical or mental conditions. If we cannot establish definite construct validity that polygraph testing detects deception, this undermines any scientific or practical usage. Largely, this construct validity criticism is not discussed in the post-conviction polygraph literature with most proponents assuming that the test is measuring deceit. Hence, it is hard to ascertain whether there is construct validity in post-conviction polygraph testing.

Moreover, the problems with polygraph testing were highlighted by the U.S. National Research Council (NRC; 2003) in their exhaustive review of the polygraph, examining its validity and reliability, and its utility in screening employees engaged in governmental defense and classified work for whether these employees were engaged in espionage. The NRC (2003) found that certain polygraph testing techniques including similar tests used in post-conviction applications lacked sufficient scientific validity and most importantly, they found no support for the use of the polygraph in screening situations (i.e., individuals are tested on general questions about crimes and actions that may have happened) rather than specific incident testing (i.e., individuals

are tested about crimes and actions that have actually occurred). Moreover, the NRC (2003) concluded that overconfidence in polygraph testing created a significant risk to national security because it had no validity in screening employees for espionage and would not catch spies. As we will see, this same logic in overconfidence extends to post-conviction use.

The Accuracy of Polygraph Testing

Within the context of post-conviction use of polygraph testing with sex offenders, Crosse and Saxe (2001) noted that supporters often report accuracy rates exceeding 90% while Iacono and Lykken (1997) have shown similarly to the NRC conclusions that most of the studies cited in support of polygraph testing lacked sufficient peer review and were methodologically flawed. Faigman et al. (2003, p. 48) noted that with regard to estimates of the accuracy for post-conviction use of the polygraph that they could not find “a single controlled randomized trial or field trial in connection with polygraph testing with anything approaching credibility.”

Several polygraph proponents often cite accuracy numbers for polygraph testing from the 2003 NRC report as if the NRC endorsed what was found in their literature review (see Buschman et al., 2010; Grubin, 2005; Grubin & Madsen, 2006; Heil & English, 2009; Kokish, Levenson, & Blasingame, 2005) or at best, downplay the NRC’s (2003) findings (Levenson, 2009). In reality, the NRC (2003) concluded that no credible estimate of polygraph accuracy could be determined beyond the appearance that the polygraph seemed to detect deception at rates greater than chance for incident-specific tests only and that increases in the reliability and accuracy of the polygraph were unlikely. Hence, any number that is reported in the literature supporting post-conviction polygraph as being verified by the NRC is not true. Moreover, the NRC (2003) also could find nothing approaching scientific credibility for estimates of polygraph testing accuracy in screening applications, under which most post-conviction polygraph testing situations fall.

Since Faigman et al. (2003) published their findings, only two studies regarding the accuracy of post-conviction polygraph testing could be located searching across a variety of social science and public affairs citation databases. The first, by Kokish et al. (2005), asked a convenience sample of 95 offenders if the polygraph test incorrectly indicated deception when the offender was telling the truth or if it indicated no deception when they were lying. In the second study, Grubin and Madsen (2006), using another convenience sample, surveyed 126 sex offenders on supervision in Georgia who had undergone periodic polygraph testing. Using a confidential survey, they asked these offenders to rate whether the polygraph test accurately measured their truthfulness or deception for each test the offender took.

The results from Kokish et al. (2005) and Grubin and Madsen (2006) are interesting if taken at face value. First, the two contradict each other on how accurate the polygraph test was if it indicated deception with Kokish et al. (2005) showing that 94% of offenders indicated that the test correctly identified their deception and only 6%

claimed being falsely accused of deception whereas Grubin and Madsen (2006), with values of 48% and 53% for the same measure, essentially reported a coin flip whether the offender was really being deceptive. Given these disparate results, this is poor evidence on which to base accuracy as the estimates range from a coin flip to near perfection. Moreover, how is a treatment provider or supervisory personnel supposed to trust any deceptive test result if the accuracy is somewhere between 50% and 95%?

However, when the test indicated no deception, both studies were in near agreement with Kokish et al. (2005) reporting that 97% of offenders indicated that the test correctly identified their truthfulness and Grubin and Madsen's (2006) two analyses revealing that 97% and 91% of offenders also indicated the test correctly identified their truthfulness. Now, this not bad a result, with only between 3% and 9% of offenders indicating that they got away with deception, but without knowing what they deceitful on, for example, minor deviance, technical violations or new felony offense, it is hard to ascertain the impact of this deceit.

But beyond face value, the results and study designs in Kokish et al. (2005) and Grubin and Madsen (2006) can also be criticized on several levels from a methodological perspective. First, both were convenience samples precluding any generalizability of the results while Kokish et al. (2005) eliminated any offender with an immediate prior deceptive test, certainly introducing selection bias. Hence, both results need to be viewed with significant discretion; Kokish et al. (2005) especially. Second, self-report data by offenders (or anyone for that matter) have many limitations including recall bias, social desirability bias, and self-serving answers (Maxfield & Babbie, 2011). Also, Huizinga and Elliot (1986) in their exhaustive review of the reliability and validity of self-report data in criminology cautioned that having "a vested interest in producing a positive evaluation of the validity of either official data or self-reports" (p. 308) could cause researchers to overlook reliability and validity because negative ones undermine the study. Hence, the reliability, validity, or quality of any self-reported measure cannot be taken for granted. For instance, the high rate reported by offenders claiming the polygraph test correctly identified they were telling the truth by both Kokish et al. (2005) and Grubin and Madsen (2006) deserves further examination. Even with confidentiality, what is the incentive for an offender to report that the polygraph test failed to catch them in a lie? Revealing this would expose the fact that the offender has been getting away with some form of crime or deviance, or that he or she has hidden a potentially critical risk, but both studies downplay this possibility.¹

In addition, both Kokish et al. (2005) and Grubin and Madsen (2006) aggregated repeated tests on individual offenders, potentially violating any assumption of independence for each test. Research has shown that conditional dependence within diagnostic tests can severely underestimate error rates (Vacek, 1985), and both studies clearly have repeated measures on the same offender used in their respective calculations. Hence, in addition to other methodological concerns, both studies' estimates of accuracy may have severely underestimated their respective error rates.

In summary, although both Kokish et al. (2005) and Grubin and Madsen (2006) tried valiantly to estimate the accuracy of post-conviction polygraph testing, several

issues compromise their efforts including the inability to validate offender self-reports limiting any faith in reliability of the measures; the fact that they give contradictory information regarding the accuracy of the polygraph testing to correctly identify deception; that they fail to adjust for potential conditional dependence that could cause severe underestimation of error; and finally, lack of a representative sample means no generalized conclusions about the accuracy of post-conviction polygraph testing can be drawn. Hence, the accuracy claims reported by proponents of post-conviction polygraph do not have any meaningful support backed by methodologically rigorous research.

The Base Rate Problem

Moreover, even if we were to grant polygraph testing a high accuracy rate, the real determinant of how well the test performs is derived from its positive and negative predictive values, not its accuracy. Accuracy in tests indicating the presence or absence of some condition is actually defined by two different measures, sensitivity and specificity. Sensitivity in polygraph testing is defined as the test correctly indicating deception given that the subject is actually deceitful while specificity is defined as the test correctly indicating no deception given that a subject is actually truthful. Positive and negative predictive values, however, measure the reverse conditional relationships seen in sensitivity and specificity. Positive predictive value in polygraph testing is defined as the subject actually being deceitful given that the test indicates deception whereas negative predictive value is defined as the subject actually being truthful given that the test indicates no deception (Gastwirth,² 1987). What is interesting is that these two measures are dependent on the base rate of deception or how rare or common deception is in the population being tested. As the rate of deception becomes rarer, the positive predictive value decreases and the false positive rate increases whereas the rate of deception becomes more common, the negative predictive value decreases and the false negative rate increases; both occur regardless of test sensitivity and specificity. In addition to concerns raised about the lack of scientific validity and reliability of the polygraph, the NRC (2003) determined that security was compromised by the low base rate of espionage and government employees were at a high risk of being labeled deceptive when in fact they were telling the truth. Within the context of post-conviction polygraph testing, false positives do not pose a threat to public safety. However, they errantly increase supervision and incarceration costs and they are constitutionally troubling in that these offenders are being punished for offenses and violations they did not commit.

False negatives, however, do pose a significant threat to public safety. The NRC's (2003) assertion that a low base rate of espionage compromised effective use the polygraph and posed a threat to national security can be logically extended to the conclusion that a high base rate of deception among sex offenders compromises polygraph effectiveness and poses a significant threat to public safety. Indeed, Seto (2004) noted that secrecy is one of the hallmarks of a sex offending, and hence, it is a reasonable

assumption that most sex offenders are lying (at some point) and the base rate of deception (at some point) will be quite high. With an unknown base rate of deception within the context of these tests, we have no way, beyond scenarios, of measuring the false negative or positive rates of these tests or adequately assessing their real impact. In other words, how many offenders are we needlessly revoking to prison or higher custody levels due to false positives and therefore incurring higher but unnecessary costs in tight budget times?³ Indeed, recent research has shown the fiscal pressures correctional agencies are under with expanding prison populations and these agencies can ill afford to erroneously add more inmates (Austin, 2010). But more importantly, if a polygraph test has a false negative, and absent any other information gained from traditional supervision methods, how much more crime, deviance, and victimization are we allowing to occur? What is truly unknown is how much practitioners who value post-conviction polygraph overvalue the evidence it provides and ignore other pertinent factors; not knowing these true opportunity costs is a significant flaw that severely undermines confidence in the accuracy of post-conviction polygraph testing.

Habituation and Sensitization

Give the regularity with which post-conviction polygraph tests are administered, there is concern that repeated administration of polygraph tests may habituate or sensitize offenders (Branaman & Gallagher, 2005). If Branaman and Gallagher (2005) are correct, a prior test could influence the accuracy of the current test. Vacek (1985) and Hui and Walter (1980) gave methods for estimating the effect of conditional dependence on diagnostic test accuracy that are independent of the base rate of the condition. For instance, if we assume that a polygraph test is 90% accurate and a minimal correlation of 0.1 between sequential tests, accuracy is reduced from 90% in test 1 to 89% in test 2, 88% in test 3, and so on to the point that the accuracy of polygraph in detecting deception would not be acceptable to those who propose its use. Note that using the methods provided by Vacek (1985) and Hui and Walter (1980), it was only assumed that the prior test was correlated with the present test, not that there were any correlations with any prior tests. What these methods show is that any impact of prior test, even a minor impact, can negatively and significantly affect polygraph accuracy. Yet some proponents such as Heil and English (2009) claim, without any empirical support, that a different polygraph examiner should be used from test to test to ward off the potential for habituation. However, as it is an uncontested fact that the subject being polygraphed is unchanged from test to test and presumably remembers the outcome of his or her prior tests, it is an untenable assumption that current tests are independent of prior tests regardless of who is administering the test. Moreover, as Vacek (1985) and Hui and Walter (1980) show, the conditional dependence of a prior test (or tests) on a current test will cause any measure of error rates to severely underestimate true error and lead to a false sense of security even with a test with supposedly high accuracy.

Other Threats to Accuracy

In addition, no research in the post-conviction polygraph testing literature was found, which discusses the effect on polygraph accuracy by certain diseases or conditions that impact physiological measures used in polygraph testing such as metabolic syndrome, hypertension, thyroid disease, or early stage obstructive pulmonary disease. Nor is there discussion on the effect on accuracy of certain mental illnesses such as bipolar disorder or depression.⁴ Although subjects serve as their own baseline in polygraph testing, one might argue that these diseases and conditions are irrelevant, but this is also an untenable assumption. For instance, it has been well established that changes in blood pressure occur with changes in blood sugar levels, even in normal populations (Rebello, Hodges, & Smith, 1983), so if someone has a spike or drop in blood sugar, they will also have a concomitant spike or drop in blood pressure, which can easily happen after baseline has been established, possibly leading to a false positive or false negative result. Hence, diseases directly related to the measures used in polygraph testing introduce variability in these measures and can decrease accuracy. In addition, given the prevalence of any of these conditions in the general and offender populations, not knowing these potential impacts on accuracy is a severe limitation.

More significant though is the assertion by Heil and English (2009) that examiner skill may affect accuracy. If this is true, how does an offender ensure that his or her examiner is proficient enough to administer the test? Where and how do examiners gain this proficiency? How long do they have to administer polygraphs before they are allowed to render opinions? Moreover, what sort of appeals process do offenders have if they feel they have been wrongly judged deceitful by an inept or neophyte examiner? Who are treatment providers or field personnel going to believe? The incorrect polygraph test result or the innocent protesting offender? As Heil and English (2009) assert that neither treatment providers nor supervisory officers have the requisite background to assess the skill level or proficiency of an examiner and that this is best handled by a neutral examiner, it begs the question of how does the provider or officer know that this examiner is proficient enough to render judgment? Heil and English (2009) state that some treatment programs require some polygraph tests to be videotaped, and charts and reports to be randomly audited. But until the true impact of neophyte examiners on accuracy is known, it might be prudent to require that all tests, given their forensic setting, be videotaped and all charts and reports to be checked by a third party.

Lastly, given the stakes involved with polygraph testing in general, it is not surprising that techniques have evolved to “beat the polygraph” through the use of what are called countermeasures. These techniques are designed, ironically enough, to deceive the polygraph examiner into finding no deception. Honts, Raskin, and Kircher (1994) demonstrated that practiced countermeasures reduce the accuracy of the polygraph exam. In addition, a widely available document on the internet by Maschke and Scalabrini (2005) discussed countermeasures at length. Written with the intent to aid

law-abiding citizens in reducing the likelihood of a negative polygraph test that could damage employment opportunities, it provided several countermeasures involving both physical and mental techniques to help aid those taking a test to render a favorable decision. However, the likelihood of success of polygraph countermeasures, much like its accuracy, is unknown but given the availability of these techniques on the internet and elsewhere, many polygraph examiners have attempted to detect countermeasures with little documented success in peer-reviewed literature, and some polygraph examiners resort to accusing subjects of countermeasures without any proof (Maschke & Scalabrini, 2005). Indeed, reaction by the polygraph community to the prevalence of information on countermeasures is to claim that distributing such material is unethical and should be made illegal (Menges, 2002), highlighting the concern for how they could affect test accuracy.

For an example over these concerns, a 2004 Iowa court case, *Willis v. Smith et al.* (2004), illustrates the potential threat that countermeasures may have for post-conviction polygraph testing accuracy. The case involved a lawsuit over a civilly committed sex offender's attempt to access to countermeasure literature, specifically Maschke and Scalabrini's (2005) aforementioned book. He sued the institution that housed him because staff denied him access to the book due to fears that countermeasures would disrupt the potential effectiveness of the offender's maintenance polygraph and might be used by other civilly committed offenders. The judge allowed the book to be given to the offender but ordered all discussion of countermeasures to be redacted. Hence, this case is documented evidence that offenders are aware of countermeasures and their potential effectiveness.

Does Accuracy Really Matter?

Despite all of the evidence above regarding polygraph accuracy and factors that may affect it, some proponents have recently put forth arguments that accuracy is immaterial to the discussion of polygraph testing's utility. Indeed, Buschman et al. (2009) claim that accuracy assessments for polygraph testing are irrelevant because "[n]o behavioural base rate, ground truth, or proportion of people in a population (as they relate to a particular trait or propensity for offending) can predict what will happen tomorrow" (p. 12) and that polygraph testing's "[u]tility becomes apparent, as it has nothing to do with predicting" (p. 12).

However, what Buschman et al. (2009) apparently have done is confuse statistical prediction, which has to do with how closely an estimate approximates the reality it is trying to measure, with predicting future human behavior, which assessment of the accuracy of polygraph testing does not and cannot do. For instance, a bathroom scale estimates how much a person weighs at the moment he or she steps on the scale; it does not predict how much the person will weigh tomorrow (even though it is probably a good prediction for tomorrow), but we can still assess how accurate the scale is for its current measure. In fact, most of us would probably want a highly accurate

bathroom scale, one that is 99.9% accurate, that is, that the weight it tells us is within $\pm 0.01\%$ of reality. In fact, this is exactly what discussions of polygraph testing accuracy are trying to elicit, and as shown above, tried to be estimated by Kokish et al. (2005) and Grubin and Masden (2006).

But more importantly, Buschman et al. (2009) bely their claim that polygraph testing is not used to help with prediction because they note the tests are incorporated into “[p]ost-conviction decision making in probation and prison settings” (p. 12). As these decisions are based on risk level and risk level is about predicting and preventing future behavior, it seems that results from the polygraph test do help (or potentially hurt) risk prediction. Therefore, when these results are inaccurate and incorrect decisions are made, it needs to be emphasized that these poor decisions based on faulty information result in either increased costs to the system by needlessly increasing the supervision level or incarcerating offenders, or, most importantly, allowing further criminality and victimization to occur. In addition, the information gleaned for polygraph testing is used in determining types and levels of treatment or therapy; if the information is flawed, then they undermine any effectiveness because these interventions target the wrong people. Hence, accuracy is very important, polygraph test information informs decision making, and is therefore related to risk prediction. No substantial argument can be made otherwise.

Increased Reporting of Offending Behaviors, Earlier Onset of Offending, and Wider Victim Pools

Having assessed the literature on the validity and accuracy of post-conviction polygraph testing, we now move to the utility of post-conviction polygraph, the first of which is that advocates claim it reveals higher admissions of historical criminal behavior by sexual offenders including wider array of both sexual and nonsexual offenses, earlier onset of offending, and a wider array of victims and victim types. Indeed, the literature supporting the use of the polygraph in post-conviction surveillance of sex offenders does show that polygraphed offenders report more deviant behavior, more victims, and a larger pool of victim types than a comparison group (Ahlmeyer, Heil, McKee, & English, 2000; English et al., 2000; Heil, Ahlmeyer, & Simons, 2003; Heil & English, 2009).

But what do these results tell us about sex offenders that have not been learned with research on all types of offenders? With regard to a wider array of offenses, if this research had found otherwise, it would have gone against what has been found in research that shows most offenders, including violent offenders, do not discriminate in the types of deviant and criminal activities they participate in (Blumstein, Cohen, Das, & Moitra, 1988; Brame, Mulvey, & Piquero, 2001; Cohen, 1986; DeLisi, 2005; Piquero, 2000; Piquero, Farrington, & Blumstein, 2007). Moreover, what post-conviction polygraph testing is trying to do in this context is estimate an offender’s criminal career—that is, offending history—to determine how versatile or specialized an offender is in

his or her criminal behavior. But it is not noted in the literature on post-conviction polygraph whether these findings of wider criminal and deviant activity hold for different types of sex offenders. That is, are the same findings of offending diversity revealed by the polygraph consistent across rapists, pedophiles, epebophiles, voyeurs, and child porn consumers?

More importantly, what does acquiring this knowledge do to improve therapeutic and supervision placements and their outcomes? Ostensibly, knowledge of prior offending patterns should place offenders in better treatment protocols and appropriate supervision levels so that providers and community supervision officers can more effectively manage their caseloads. Indeed, Gannon, Beech, and Ward (2008) performed an exhaustive examination of the literature on post-conviction polygraph testing that supports the idea that this testing enhances risk prediction—both static and dynamic—for sexual offenders. Not surprisingly, they found within the context of criminal history disclosure that small sample sizes, lack of randomization, lack of adequate comparison groups, poor or missing experimental controls, or an inability to remove confounding variables such as treatment effects apart from polygraph testing itself in such studies as Ahlmeyer et al. (2000), Emerick and Dutton (1993), English et al. (2000), Heil et al. (2003), and Hindman and Peters (2001) rendered the ability to derive firm conclusions from any of these studies difficult. Hence, although the evidence base for these disclosure effects is large, it is also frustratingly bereft of substantial findings from anything approaching a randomized, controlled trial to account for selection bias, confounding factors, and other nonpolygraph treatment effects. Moreover, there is nothing to suggest in this evidence base that these disclosures actually help to reduce offending.

Finally, although polygraph testing does elicit confessions at a higher rate than other forms of interrogation, it is questioned whether this is due to the test or the ability of the test to serve as an interrogation prop that tricks some into confessing. Some authors (Crosse & Saxe, 2001; Ford, 1996; Gannon 2006; Gannon, Keown, & Polaschek, 2007) have drawn comparisons to the bogus pipeline effect where it has been shown that subjects attached to a nonfunctioning apparatus will make admissions if they believe the apparatus can detect what the machine is purported to measure. These authors argue that the utility of the polygraph lies only in its placebo effect. Indeed, in the 2004 Iowa court case discussed earlier, this placebo effect was noted by the testimony of a staff member who admitted that “it is more important for patients to believe the polygraph is valid than for the test actually to be valid” (*Willis v. Smith et al.*, 2004, pg. 8). From this testimony, the judge concluded that “the polygraphs act similarly to a placebo for some patients, in that if the patient is worried about being caught in a deception, the patient may admit things before the test is administered” (*Willis v. Smith et al.*, 2004, pg. 8). However, few polygraph proponents refer to this placebo effect in their reviews of the literature. Lastly, this case provides evidence that the utility of the polygraph to elicit higher disclosure of prior criminal history or disclosure of new crimes or field violations may be compromised if offenders are aware of the placebo effect.

The Polygraph Test as a Deterrent

Finally, Heil and English (2009) and Grubin (2008) both argue that the greatest utility of the polygraph is to augment treatment by providing clinical personnel information on more entrenched and recalcitrant behavior so that agencies can more ably place these offenders in appropriate therapies. They also argue that knowledge of these more entrenched and recalcitrant behaviors augment supervision levels by allowing field personnel to place these offenders on more appropriate supervision regimes. Again, implied by both these arguments is that this additional information should reduce risk of additional offending and violations of supervision standards. And essentially, the argument is that polygraph testing, and hence the threat of being caught, will deter offenders from future behavior, both criminal and violative.

Deterrence, within the criminological literature, is divided into two types: General deterrence where the threat of sanction or detection prevents people from committing criminal and deviant acts to begin with, and specific deterrence, where individuals are deterred from future offending by these same threats of sanction or detection (Pratt, Cullen, Blevins, Daigle, & Madensen, 2006). Hence, given that polygraph testing is performed on probationers and parolees to prevent future offending, it is a form of specific deterrence.

It is interesting that both Heil and English (2009) and Grubin (2008) place the greatest utility in polygraph testing's deterrent effect because the empirical evidence for a specific deterrent effect for any kind of sanction or supervision is minimal (Andrews & Bonta, 2006; Gendreau, Goggin, French, & Smith, 2006; Lipsey, 2009; Pratt et al., 2006). In discussing specific deterrence, criminologists have also begun to separate out deterrence, which involves the calculation of risk and rewards of committing an offense, and deterrability, which is the ability to perform the calculation (Jacobs, 2010). That is, if an offender has low deterrability, he or she will not be able to adequately calculate the risk and rewards no matter how much greater the risk of detection is over the rewards of the criminal or deviant act. What this means to post-conviction polygraph testing is that to really demonstrate its deterrent effect, it must demonstrate a practically important reduction in future offending and supervision violations. To demonstrate its impact on deterrability, it must show that offenders respect and inculcate the polygraph's ability to detect offending sufficiently so that it is the main reason they actually behaved properly. For instance, Kokish et al. (2005) found 90% of 95 sex offenders from a convenience sample thought that polygraph testing was a helpful part of their treatment plan because it reduced their criminogenic behaviors; but without showing lower recidivism rates, that is an actual change in behavior, it is a meaningless statistic and in no way shows a deterrent effect for polygraph testing.

So Is Post-Conviction Polygraph Testing Useful?

Most of the literature that supports post-conviction polygraph testing spends an inordinate amount of time assessing whether offenders, treatment providers, and community

supervision officers find it a useful component of a treatment program (Grubin, 2008; Grubin, 2010; Heil et al, 2003). However, the real question is why would it matter whether an offender or provider find polygraph testing useful without evidence of changed behavior? For instance, research has shown that more expensive placebos are perceived to work better than cheaper placebos (Waber, Shiv, Carmon, & Ariely, 2008). The fact remains that both are placebos, so neither the expensive or cheaper variety are useful in treating any condition. Hence, opinions of individuals involved with polygraph testing, being they offenders or providers, do not provide an adequate measure of usefulness nor utility beyond the fact that people perceive them to be useful or utile. *The real measure of usefulness or utility of any correctional treatment or program is whether it delivers the desired change in whatever behavior it is trying to affect.* One can argue against this position, but ultimately, this is how correctional treatment programs and interventions are and should be judged. And indeed, the literature on “what works” in correctional treatment stresses these behavioral change outcomes as indicative of program effectiveness (Gendreau, 1996; Gendreau, Goggin, Cullen, & Paparozzi, 2002; Gendreau, Smith, & French, 2006; Lowenkamp, Latessa, & Smith, 2006; MacKenzie, 2000, 2005, 2007; Polizzi, MacKenzie, & Hickman, 1999)

Which brings us to the main thrust of this article: Does post-conviction polygraph testing reduce further offending? Discussions of increased disclosures, accuracy, habituation, sensitization, and deterrence aside, when the rubber meets the road, does polygraph testing deliver on its promise to reduce criminality and deviance? To put it in the words of a polygraph proponent, “Here, the ‘weight of evidence’ is less heavy” (Grubin, 2008, p. 185). And when Grubin (2008) refers to the weight as less heavy, he is not joking; there are exactly three studies that assess the impact of post-conviction polygraph on subsequent offending⁵

The first is a study by Abrams and Ogard (1986) compared recidivism rates between a group of 35 offenders on post-conviction polygraph testing and another group of 243 offenders not on post-conviction polygraph testing. They found that the first group had a 2-year recidivism rate of 31% and the second group had a 2-year rate of 74%. In the second study, Edson (1991) found that of 173 sex offenders under community supervision who were required to take a periodic polygraph test, 95% of these offenders did not reoffend within 9 years.

The third study, the most robust out of the three in terms of research design, by McGrath, Cumming, Hoke, and Bonn-Miller (2007) examined 104 sex offenders on post-conviction polygraph testing matched by type of treatment and supervision with 104 sex offenders not on polygraph testing. No significant differences between the two groups were found on age, educational attainment, sex offense type, or risk levels. They then recorded 5-year rates for new sex convictions, new violent (but nonsex) convictions, new nonviolent convictions, field violations, and prison returns. They also matched previous findings by showing increased disclosures of wider prior criminal history for the polygraph group. However, the only statistically significant difference (i.e., $p < 0.05$) in offending they reported was on new violent convictions where the polygraph group had 2.9% or 3 new violent offenses and the nonpolygraph group

had 11.5% or 12 new violent offenses. No other statistically significant differences were found including new sexual offenses, which were about even between the two groups.⁶

Now, polygraph testing proponents use this significant finding for lower new violent convictions as favorable evidence that testing works to reduce reoffending (Grubin, 2008). But as the incidence of new violent convictions is low in both the polygraph and nonpolygraph groups, 3 and 12 cases respectively, can 9 more cases out of 104 offenders really be considered a practical difference? Not really, because increasing the number of new violent offenses to 5 in the polygraph group versus 12 in the nonpolygraph group changes the p -value to 0.09. If two cases can change the conclusions of a test, it is hardly solid evidence that polygraph testing reduces new violent offending. In addition, although McGrath et al. (2007) do not speculate on why there is a lower rate of new violent convictions but a higher rate in new nonviolent offenses, Walker's (2006) concept of criminal justice thermodynamics gives a good basis for explaining these differences. Criminal justice thermodynamics occurs when shifts in discretion for a variety of factors are done to help the system to function effectively. For example, the difference seen in new violent convictions between the two groups could be due to plea bargaining down to a nonviolent offense to more easily obtain a conviction (the polygraph group did have a higher rate of new nonviolent crime, 35.6% vs. 29.8% or 6 more cases), or the use of prison revocation to ameliorate the cost and effort of trying a new case. Without case information, it is hard to be certain, but these are more plausible explanations than any effect polygraph testing might claim.

What is more interesting is that although no statistical differences were found between the polygraph and nonpolygraph groups for any new offense, field violations or prison returns, rates were higher in the polygraph group for any new offense (39.4% vs. 34.6%), field violations (54% vs. 47%) and prison returns (47% vs. 39%). Moreover, these differences are practically important⁷ as the promise of post-conviction polygraph testing is that increases in disclosure of prior offending will allow for better treatment that will reduce risk of future offending and that the polygraph will also serve as a deterrent for field violations and new offending. Even if this sample was too small to find a difference, shouldn't the nonpolygraph group have higher rates of new offenses, field violations and prison returns? Instead, the evidence goes in the opposite direction. Indeed, the conclusion from McGrath et al. (2007) is

The results of this study support research findings cited earlier indicating that individuals who have committed sexual offenses and who undergo polygraph compliance testing admit to engaging in previously withheld high risk behaviors and that providers find this information relevant for improving treatment and supervision services. *Although it seems logical that these outcomes would lead to lower recidivism rates, the present results do not provide much support for this hypothesis.* (p. 389; emphasis added)

However, proponents still may claim this study as a success by claiming that it detected violations that otherwise might have been missed. But this serves only as having one's cake and eating it too. Proponents cannot simultaneously claim a deterrent effect and a surveillance effect without explicitly stating what the expected effects will be. Moreover, how do we separate out the deterrent effect from the surveillance effect?

The overall conclusions that can be drawn by these three studies is that evidence for utility of the polygraph in reducing offending is weakly supported by the first two studies, but this conclusion is undermined by the lack of adequate controls for selection bias, lack of a random sample, and probable confounding in addition to the facts that the Abrams and Ogard (1986) study appeared in *Polygraph*, which is essentially a trade journal for polygraph examiners and the Edson (1991) study was a technical report never submitted for peer review. Which leaves the only academically peer-reviewed study of note, McGrath et al. (2007), and it directly contradicts the utility hypothesis by having nonsignificant but higher rates of new offending, field violations, and prison revocations for the post-conviction polygraph test group versus a control group. Rather than Grubin's (2008) assertion that the evidence for the post-conviction polygraph testing reducing criminal and deviant behavior is less heavy, it is nonexistent.

Discussion

Given the evidence presented within this article, it is clear that the evidence supporting the utility of post-conviction polygraph testing is far from the compelling picture that advocates paint. However, a distinction must be made between use and utility. Clearly, the polygraph test has found its niche within post-conviction supervision of sex offenders. But how should the evidence presented here be interpreted overall? That is, how are decision makers using the information gleaned from polygraph testing in supervision? Aside from the studies showing that treatment and supervisory personnel value information obtained in these tests, the McGrath et al. (2007) study is really the only evidence in the form of field violations and prison revocations that indicates decisions made by these staff and these results show that the polygraph group are both given technical violations and are revoked back to prison at higher rates. However, without information about why a decision was made, we can only speculate that a deceptive test is being used in the absence of other information (e.g., confession) by staff to make revocation and violation decisions. But mere speculation does not answer the question. Thus, this paucity of evidence makes it difficult to say that post-conviction polygraph testing has any meaningful impact on criminal justice decision making and offender outcomes—good or bad—beyond the fact that personnel and offenders seem to like it.

In fact, the widespread adoption of polygraph testing in sex offender treatment despite any evidence that it works to reduce offending (McGrath et al., 2010) is an all too common event seen within correctional systems (Latessa, Cullen, & Gendreau, 2002). Indeed, Latessa and colleagues (Cullen & Gendreau, 2001; Gendreau, 2000;

Gendreau, Smith, & Thériault, 2009; Latessa et al., 2002) use the term “correctional quackery” to label programs and treatments (e.g., boot camps) that continue to be used despite having either a lack of empirical and theoretical support, or substantial evidence that such programs do not work at all to reduce offending.

In summary, beyond two studies using convenient samples and self-reports, no adequate assessment or measure of accuracy across a variety of conditions, offenders, and examiner skill level has been performed and verified by independent researchers within the post-conviction polygraph test literature. Nor has any study measured or verified of the impact of habituation and sensitization or common diseases on accuracy. Without better understanding of the impact of these conditions on polygraph testing’s ability to render a deceptive or nondeceptive finding, the true accuracy of post-conviction polygraph testing remains a mystery. Lastly, we do not know how information provided by the polygraph is synthesized within decision making by practitioners. What happens when contradictory information is obtained the polygraph? Do practitioners ignore the result? Does it overwhelm all other evidence? There are opportunity costs involved in all decisions but we need better information to understand both the benefit and harm with using information obtained in post-conviction polygraph. We can ill afford to use programs such as post-conviction polygraph testing without establishing that they actually reduce offending because we do not have the time or money to waste on ineffective correctional programs nor can we accept the very real potential they pose for increased victimization.

Recommendations

Despite the negative conclusions drawn in this review, it would be remiss to not chart a path that would allow us to amass the needed evidence to effectively judge post-conviction polygraph testing’s real utility. With that in mind, first, we need researchers, clinicians, and polygraph examiners to stop selling post-conviction polygraph testing as an effective, evidence-based tool for supervising offenders and start adequately assessing its efficacy and effectiveness with viable outcome measures such as new offenses, revocations, and technical violations. We have correctional agencies with enormous databases containing this information. It would be easy enough to identify, throughout the United States and other countries, a proper sample with differing types of sex offenders with post-conviction polygraph test exposure, match them with a comparable cohort with no exposure, and measure these recidivism outcomes. In addition, we can use propensity score matching with these data to approximate randomized controlled trials which would make this evidence even more compelling (Stuart, 2010). And although it points toward a potential conclusion, simply relying on a single negative study of 208 sex offenders from Vermont is not enough evidence to dismiss post-conviction polygraph testing as an ineffective or countereffective tool. Also claiming that supervisory and treatment personnel and offenders “like it” or find it useful is not evidence because science is not nor should be a popularity contest. Instead, it must winnow out competing hypotheses with

compelling evidence of efficacy and effectiveness and in this case, recidivism, violations, and revocations are the only outcomes of interest to demonstrate polygraph testing's efficacy and effectiveness. And when we find substantial compelling evidence against a hypothesis, no matter how much we want it to be true, the scientific method requires us to dismiss it and adopt the hypothesis with real empirical and theoretical support (Sagan, 1996). That really is the beauty of the scientific method; it does not care whether we like the results, it merely requires that we change our views when we are wrong. We need to be willing to do that with all correctional treatments, including post-conviction polygraph testing.

Second, and echoing NRC (2003), more research—performed by researchers and scientists indifferent and, more importantly, invested in the success or failure of any polygraph theory or test type—is needed to (a) establish using randomized, controlled trials with appropriate placebo groups what type, if any, of polygraph testing has the most theoretical and empirical support; (b) determine how this type of testing would effectively aid criminal justice agencies in the supervision of offenders in the form of reduced recidivism; (c) assess the impact on accuracy by diseases and mental illnesses related to the physiological processes used in polygraph testing; and (d) if efficacy and effectiveness is found, determine the best way to incorporate these methods into agencies that minimizes adverse events from both false positives and false negatives.

Third, if any scientifically useful polygraph technique is identified above, we need to eliminate humans from administering the test as this would obviate the problem of neophyte or inept examiners. These objective polygraph techniques should be able to be automated using software to not only administer both examination and pre-examination questions similar to what is done in computer-aided surveys (Maxfield & Babbie, 2011) but also to score and record charts and videos of these tests.⁸ This software could then be used without the need for a trained examiner, allowing it to be used in probation offices, prisons, and clinical settings, giving real time feedback to treatment and field personnel, and reducing the cost burden to agencies and offenders. Indeed, one wonders why an enterprising polygraph examiner hasn't done this yet, even with techniques that have little empirical support.

Fourth, it may be futile to think that agencies will abandon use of post-conviction polygraph testing given its deep entrenchment in sex offender supervision. Indeed, programs for combatting substance abuse in children and teenagers like Drug Abuse Resistance Education, continue to survive and thrive despite the plethora of evidence showing that these programs are completely ineffective in preventing future drug use by its graduates and may, in fact, cause harm (Lilienfeld, 2007). But despite a real lack of evidence for efficacy and effectiveness of polygraph testing in sex offender supervision, it is now being extended for post-conviction use with other types of offenders such as those convicted of domestic violence (Wilson, Batye, & Riveros, 2008). This cannot be stated more forcefully, without any substantial evidence base to do so, agencies should not adopt post-conviction polygraph testing for any other types of offenders until, (a) clear outcomes can be established to measure its efficacy and effectiveness, and (b) the evidence base is sufficient to determine polygraph testing poses minimal

additional harm. Moreover, no revocation decisions based on the polygraph should be made without some substantiation of admitted criminal and deviant activity.

Lastly, this review of the research on post-conviction polygraph testing was intended to give an unvarnished view of the evidence, both good and bad. If something really works, it will provide the necessary evidence regardless of who is the author. If positive findings can only be found by those who support a technique and not by those without a vested interest, then it is probably pseudoscience. And to reiterate above, if something does not live up to its promise despite our best intentions, we need, above all, to be intellectually honest and abandon these programs and treatments without evidentiary support in favor of those that do because we have neither the time nor money to waste on such programs given our current economic state. In conclusion, and so far, post-conviction polygraph testing has not provided an ample evidentiary base to demonstrate its utility for decreasing future criminality and deviance for sex offenders and what it has shown is futile in assessing its real worth.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Both Kokish et al. (2005) and Grubin and Madsen (2006) devote discussion to false confessions. Although they are troubling, they are not only a problem confronting polygraph testing but also are part and parcel of a larger problem within criminal justice and mental health (Loftus, 2004). Hence, this review does not examine the problem of false confessions but the reader should note that it is a serious problem.
2. Coincidentally, Gastwirth (1987) used the polygraph to demonstrate the positive and negative predictive values of diagnostic tests and the effect of the condition prevalence or base rate. If either of the predictive values is low, the diagnostic test is deemed insufficient for the matter at hand.
3. An anonymous reviewer questioned whether it was standard practice for offenders to be sent to prison for deceptive results. Indeed, although post-conviction polygraph standards say that no decision should be made solely on polygraph results, we really do not know what field personnel are doing in practice as it is not measured, but it is an interesting question. In addition, the reviewer questioned whether post-conviction polygraph results could be legally admissible in court. In fact, a violation of supervision standards is enough to get a supervised offender revoked regardless of the admissibility of the evidence. Moreover, the exclusionary rule does not apply in revocation hearings (*Pennsylvania v. Scott*, 1998), so any evidence gathered may be admissible. For instance, an Idaho case, *State v. Travis* (1994), held that revocation decisions based solely on the polygraph could be used to return a sex offender probationer back to prison. A sampling of other cases upholding the

right of state and federal government to use post-conviction polygraph on sex offenders includes *Jones v. Virginia* (2003), *Kansas v. Lumley* (1999), *U.S. v. Dotson* (2003), *U.S. v. Lee* (2003), *U.S. v. Saxena* (2000), *U.S. v. Scheffer* (1998), *U.S. v. Taylor* (2003), *U.S. v. York* (2004), and *U.S. v. Zinn* (2003).

4. Kokish et al. (2007) and Grubin and Madsen (2006) did discuss certain personality types being more likely to put forth a false confession, but they did not discuss the impact that mental and physical illness might have on polygraph accuracy.
5. Grubin (2008) also counted in his assessment of post-conviction polygraph testing's ability to reduce reoffending his and colleagues 2004 study on whether offenders thought that testing reduced their subsequent offending and his 2006 study where probation officers found polygraphy to be useful in 90% of cases. In addition, Grubin's (2010) study measured only whether disclosures helped case managers, not whether it reduced offending. These are not counted here because they do not measure actual changes in offending or field violations.
6. As noted by an anonymous reviewer, offense categories, violations, and revocations within the McGrath et al (2007) study were probably not independent of one another within subjects or across time. This introduces statistical bias, most notably from repeated measures and should have been accounted for in their analyses.
7. An anonymous reviewer pointed out that that I was using statistically insignificant findings to argue these points. Hence, I use the term "practical difference" rather than "significant difference." This is the key point because within the McGrath et al (2007) study, the polygraph group's results were consistently higher than the control group's results. Indeed, the practical difference is what is scientifically or theoretically important while statistical significance is largely a function of sample size. See Ziliak and McCloskey (2008) for an excellent discussion of statistical versus. practical difference.
8. An anonymous reviewer expressed skepticism that automation could be achieved, claiming that the lengthy pretest interview precluded this from happening because it is where "important information is gathered, questions are formulated, and it is confirmed that the offender understands what is being asked, amongst other things." However, if this is true, then post-conviction polygraph testing is an unstandardized technique that changes from subject to subject; hence, no meaningful assessment of its accuracy, validity, or effectiveness can be made. In other words, the machine is merely a prop used to elicit confession.

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*, 63-68.
- Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). Impact of polygraph on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *12*, 123-138.
- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: Anderson.
- Austin, J. (2010). Reducing America's correctional populations: A strategic plan. *Justice Research and Policy*, *12*, 9-40.

- Blumstein, A., Cohen, J., Das, S., & Moitra, S. D. (1988). Specialization and seriousness during adult criminal careers. *Journal of Quantitative Criminology*, 4, 303-345.
- Brame, R., Mulvey, E. P., & Piquero, A. R. (2001). On the development of different kinds of criminal activity. *Sociological Methods and Research*, 29, 319-341.
- 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.
- Buschman, J., Bogaerts, S., Foulger, S., Wilcox, D., Sosnowski, D., & Cushman, B. (2010). Sexual history disclosure polygraph examinations with cybercrime offences: A first Dutch explorative study. *International Journal of Offender Therapy and Comparative Criminology*, 54, 395-411.
- California Coalition on Sexual Offending. (2004). *Position paper for clinical polygraph examinations in sex offender treatment*. Stockton, CA: Author.
- Cohen, J. (1986). Research on criminal careers: Individual frequency rates and offense seriousness. In A. Blumstein, J. Cohen, J. A. Roth, & C. Visher (Eds.), *Criminal careers and "career criminals"* (pp. 292-418). Washington, DC: National Academy Press.
- Crosse, T. P., & Saxe, L. (1992). A critique of the validity of polygraph testing in child sexual abuse cases. *Journal of Child Sexual Abuse*, 1, 19-33.
- Crosse, T. P., & L. Saxe. (2001). Polygraph testing and sexual abuse: The lure of the Magic Lasso. *Child Maltreatment*, 6, 195-206.
- Cullen, F. T., & Gendreau, P. (2001). From nothing works to what works: Changing professional ideology in the 21st century. *The Prison Journal*, 81, 313-338.
- DeLisi, M. (2005). *Career criminals in society*. Thousand Oaks, CA: SAGE.
- Edson, C. F. (1991). *Sex offender treatment*. Jackson County, OR: Department of Corrections.
- 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). *The value of polygraph testing in sex offender management* (Research report submitted to the National Institute of Justice, No. D97LBVX0034). Denver: Colorado Department of Public Safety, Division of Criminal Justice, Office of Research and Statistics.
- Faigman, D. L., Fienberg, S. E., & Stern, P. C. (2003). The limits of the polygraph. *Issues in Science and Technology*, 20, 40-48.
- Fienberg, S. E., & Stern, P. C. (2005). In search of the Magic Lasso: The truth about the polygraph. *Statistical Science*, 20, 249-260.
- Ford, C. V. (1996). *Lies! Lies! Lies! The psychology of deceit*. Washington, DC: American Psychiatric Press.
- Gannon, T. A. (2006). Increasing honest responding on cognitive distortions in child molesters: The bogus pipeline procedure. *Journal of Interpersonal Violence*, 21, 1-18.
- Gannon, T. A., Beech, A. R., & Ward, T. (2008). Does the polygraph lead to better risk prediction for sexual offenders? *Aggression & Violent Behavior*, 13, 29-44.
- Gannon, T. A., Keown, K., & Polaschek, D. L. L. (2007). Increasing honest responding on cognitive distortions in child molesters: The bogus pipeline revisited. *Sexual Abuse: A Journal of Research and Treatment*, 19, 5-22.

- Gastwirth, J. L. (1987). The statistical precision of medical screening devices: Application to polygraph and AIDS antibodies test data. *Statistical Science*, 2, 213-222.
- Gendreau, P. (1996). The principles of effective intervention with offenders. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 117-130). Thousand Oaks, CA: SAGE.
- Gendreau, P. (2000). 1998 Margaret Mead Award address: Rational policies for reforming offenders. In M. McMahon (Ed.), *Assessment to assistance: programs for women in community corrections* (pp. 329-338). Lanham, MD: American Correctional Association.
- Gendreau, P., Goggin, C., Cullen, F. T., & Pappozzi, M. (2002). The common sense revolution and correctional policy. In J. McGuire (Ed.), *Offender rehabilitation & treatment: Effective programs & policies to reduce re-offending* (pp. 360-386). Chichester, UK: Wiley.
- Gendreau, P., Goggin, C., French, S., & Smith, P. (2006). Practicing psychology in correctional settings: "What works" in reducing criminal behavior. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (3rd ed., pp. 722-750). New York, NY: Wiley.
- Gendreau, P., Smith, P., & French, S. (2006). The theory of effective correctional intervention: Empirical status & future direction. In F. T. Cullen, J. P. Wright, & R. Blevins (Eds.), *Taking stock: The status of criminological theory—advances in criminological theory* (pp. 419-446). New Brunswick, NJ: Transaction.
- Gendreau, P., Smith, P., & Thériault, Y. L. (2009). Chaos theory and correctional treatment: Common sense, correctional quackery, and the Law of Fartcatchers. *Journal of Contemporary Criminal Justice*, 25, 384-396.
- Grubin, D. (2005). Lie detection and the polygraph: A historical review. *Journal of Forensic Psychiatry and Psychology*, 16, 357-369.
- Grubin, D. (2008). The case for polygraph testing of sex offenders. *Legal and Criminological Psychology*, 13, 177-189.
- Grubin, D. (2010). A trial of voluntary polygraphy testing in 10 English probation areas. *Sexual Abuse: A Journal of Research and Treatment*, 22, 266-278.
- Grubin, D., & Madsen, L. (2006). Accuracy and utility of post-conviction polygraph testing of sex offenders. *British Journal of Psychiatry*, 188, 479-483.
- Heil, P., Ahlmeyer, S., & Simons, D. (2003). Crossover sexual offences. *Sexual Abuse: A Journal of Research and Treatment*, 15, 221-236.
- Heil, P., & English, K. (2009). Sex offender polygraph testing in the United States: Trends and controversies. In D. T. Wilcox (Ed.), *The use of the polygraph in assessing, treating, and supervising sex offenders: A practitioner's guide*. NY: Wiley (pp. 181-216).
- Hindman, J., & Peters, J. M. (2001). Polygraph testing leads to better understanding adult and juvenile sex offenders. *Federal Probation*, 65, 8-15.
- 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.
- Hui, S. L., & Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, 36, 167-171.
- Huizinga, D., & Elliott, D. S. (1986). Reassessing the reliability and validity of self-report delinquency measures. *Journal of Quantitative Criminology*, 2, 293-327.

- Iacono, W. G. (2001). Forensic "lie detection": Procedures without scientific basis. *Journal of Forensic Psychology Practice, 1*, 75-86.
- Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82*, 426-433.
- Jacobs, B. A. (2010). Deterrence and deterrability. *Criminology, 48*, 417-441.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349-364.
- Jones v. Virginia, No. 2637-02-1 (2003).
- Kansas v. Lumley, WL 218704 (1999).
- Kokish, R. (2003). The current role of post-conviction sex offender polygraph testing in sex offender treatment. *Journal of Child Sexual Abuse, 12*, 175-194.
- 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.
- La Fond, J. Q., & Winick, B. J. (2003). Sex offender reentry courts: A cost effective proposal for managing sex offender risk in the community. *Annals of New York Academy of Sciences, 98*, 300-223.
- Latessa, E., Cullen, F., & Gendreau, P. (2002). Beyond correctional quackery: Professionalism and the possibility of effective treatment. *Federal Probation, 66*, 43-49.
- Levenson, J. S. (2009). Sex offender polygraph examination: An evidence-based case management tool for social workers. *Journal of Evidence-Based Social Work, 6*, 361-375.
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53-70.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders, 4*, 124-147.
- Loftus, E. F. (2004). The devil in confessions. *Psychological Science in the Public Interest, 5*, 1-2.
- Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter: The impact of adhering to the principles of effective intervention. *Criminology and Public Policy, 5*, 201-220.
- Lykken, D. (1998). *A tremor in the blood: Uses and abuses of the lie detector* (2nd ed.). Reading, MA: Perseus.
- MacKenzie, D. L. (2000). Evidence-based corrections: Identifying what works. *Crime and Delinquency, 46*, 457-471.
- MacKenzie, D. L. (2005). The importance of using scientific evidence to make decisions about correctional programming. *Criminology & Public Policy, 4*, 249-258.
- MacKenzie, D. L. (2007) *What works in corrections? Reducing the criminal activities of offenders and delinquents*. Cambridge, UK: Cambridge Press.
- Maschke, G. W., & Scalabrini, G. J. (2005). *The lie behind the lie detector* (4th digital ed.). Available from <http://antipolygraph.org>
- Maxfield, M. G., & Babbie, E. R. (2011). *Research methods for criminal justice and criminology* (6th ed.). Stamford, CT: Cengage Learning.

- McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.
- McGrath, R. J., Cumming, G. F., Hoke, S. E., & Bonn-Miller, M. O. (2007). Outcomes in a community sex offender treatment program: A comparison between polygraphed and matched non-polygraphed offenders. *Sexual Abuse: A Journal of Research and Treatment, 19*, 381-393.
- Menges, P. M. (2002). Ethical considerations of providing polygraph countermeasures to the public. *Polygraph, 31*, 254-262.
- Pennsylvania v. Scott, 524 U.S. 357 (1998).
- Piquero, A. (2000). Frequency, specialization, and violence in offending careers. *Journal of Research in Crime and Delinquency, 37*, 392-418.
- Piquero, A. R., Farrington, D. P., & Blumstein, A. (2007). *Key issues in criminal career research: New analyses of the Cambridge Study in Delinquent Development*. New York, NY: Cambridge University Press.
- Polizzi, D. M., MacKenzie, D. L., & Hickman, L. (1999). What works in adult sex offender treatment?: A review of prison- and non-prison-based treatment programs. *International Journal of Offender Therapy and Comparative Criminology, 43*, 357-374.
- Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L., & Madensen, T. (2006). The empirical status of deterrence theory: A meta-analysis. In F. T. Cullen, J. P. Wright & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory—advances in criminological theory*, (pp. 367-396) New Brunswick, NJ: Transaction.
- Rebello, T., Hodges, R. E., & Smith, J. L. (1983). Short-term effects of various sugars on antiatriuresis and blood pressure changes in normotensive young men. *American Journal of Clinical Nutrition, 38*, 84-94.
- Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark*. New York, NY: Ballantine.
- Seto, M. C. (2004). Pedophilia and sex offenses against children. *Annual Review of Sex Research, 14*, 321-361.
- State v. Travis, 125 Idaho 1, 867 P. 2d 234 (1994).
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1-21.
- United States v. Dotson, 324 F.3d 256, 261 12 (2003).
- United States v. Lee, 315 F.3d 206, 213 (2003).
- United States v. Saxena, 229 F.3d 1, 5 n.1 (2000).
- United States v. Scheffer, 523 U.S. 303, 309 (1998).
- United States v. Taylor, 338 F.3d 1280, 1284 n.2 (2003).
- United States v. York, 357 F.3d 14, 22 11 (2004).
- United States v. Zinn, 321 F.3d 1084, 1089-90 14 (2003).
- U.S. National Research Council Committee (2003). *The polygraph and lie detection*. Washington, DC: National Academic Press.
- Vacek, P. M. (1985). The effect of conditional dependence on diagnostic tests. *Biometrics, 41*, 959-968.

- Waber, R. L., Shiv, B., Carmon, Z., & Ariely, D. (2008). Commercial features of placebo and therapeutic efficacy. *Journal of the American Medical Association*, 299, 1016-1017.
- Walker, S. (2006). *Sense and nonsense about crime and drugs: A policy guide* (6th ed.). Belmont, CA: Thomson Higher Education.
- Willis v. Smith et al., C04-4012-MWB (Iowa). (2004).
- Wilson, D. B., Batye, K., & Riveros, R. (2008). *Testing and evaluation of the use of polygraphs to combat violence against women* [Document No.: 222115]. Washington, DC: National Institute of Justice.
- Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Bio

Jeffrey W. Rosky is an assistant professor in the Department of Criminal Justice at the University of Central Florida. His research interests include criminological theory, correctional treatment programs and research methods. He has a Ph.D. in Criminal Justice from Washington State University, an M.S. in Biometrics from the University of Colorado, and a B.A. in Statistics from Rutgers University.