# Information Gain of the Relevant-Irrelevant Test

**Charles R. Honts[1] and Mark Handler[2]**

## Abstract

We adapted the Information Gain (IG) analysis procedures described by Honts and Schweinle (2009) to the Relevant – Irrelevant technique (RIT) using data from Krapohl & Rosales (2014). The RIT is used in a number of screening settings, including public safety applicant screening and security screening for U.S. Federal employees (Honts & Amato, 2007). IG provides an index of the increase in information gained by using a test result over predicting the base rate of the target event. We found that deceptive outcomes from the RIT examinations in Krapohl & Rosales (2013) failed to provide a significant increase in IG over unassisted professional lie catchers across the complete range of base rates of deception. For truthful outcomes IG peaked at 0.268 at a base rate of deception of 0.63. For truthful results, there was a significant (1-tailed, $p < .05$) increase in IG over unassisted professionals for base rates ranging from 0.12 through 0.98 inclusive. In comparison to available alternatives, such as the Directed Lie Screening Test (Handler, Honts, & Nelson, 2013), the RIT provides markedly inferior performance across the range of base rates with both truthful and deceptive outcomes.

## Introduction

Psychophysiological Detection of Deception (PDD) screening tests are routinely included in the evaluation process of applicants for employment in the public safety setting (American Polygraph Association [APA], 2009a; Handler, Honts, Krapohl, Nelson & Griffin, 2009; Honts, Raskin & Kircher, 2008; National Research Council, 2003) as well as post-conviction sex offender monitoring (APA, 2009b; Consigli, 2002). Screening tests are any tests conducted in the absence of a known or reported problem and can address single or multiple target issue(s), depending on situational needs. In January 2012, the APA defined the requirements for screening techniques conducted by members. APA bylaws and standard of practice require the following of screening tests (APA, 2012):

> 3.9.1.4 Polygraph techniques used for screening purposes shall be those for which there exists research demonstrating an unweighted accuracy rate significantly greater than chance, and should be used in a "successive hurdles" approach which entails additional testing with validated methods when the screening test is not favorably resolved.

[1] Charles Honts received his Ph.D. in Psychology from the University of Utah in 1986, and joined the Boise State University Psychology faculty in 1995 where he currently holds the position of Professor of Psychology. Dr. Honts research interests involve scientific credibility assessment with adults and children, specifically the psychophysiological detection of deception. Other research and practice interests include interviewing, including child witnesses, interrogation and false confession phenomena. He is the author of more than 300 scientific publications, papers and presentations.

[2] Mark Handler is the research chairperson for the American Association of Police Polygraphists, has published articles and studies on various aspects of the polygraph, and can be reached at polygraphmark@gmail.com. The views expressed are those of the author and not necessarily the AAPP or the APA.

Correspondence conceding this article should be directed to, Charles R. Honts, Ph.D. at Honts, Handler, and Hartwig, LLC, 3105 Sweetwater Drive, Boise ID, 83716 or honts@mac.com.

The authors would like to acknowledge and thank Raymond Nelson for his helpful comments and suggestions after reviewing drafts of this manuscript.

> 3.9.1.4.1 (Effective January 1, 2015, section 3.9.1.4 shall be replaced with the following). Polygraph techniques used for screening purposes shall be those for which there exist at least two published empirical studies, original and replicated, demonstrating an unweighted accuracy rate that is significantly greater than chance, and should be used in a "successive hurdles" approach which entails additional testing with validated methods when the screening test is not favorably resolved. (p.6)

Krapohl, Senter and Stern (2005) reported a by-case accuracy of 73% for a blind evaluator performing global analysis with a matched sample of field cases conducted with the Relevant Irrelevant technique (RIT). That study was the only available study that satisfied the minimal requirements for inclusion in the APA (2011) meta-analytic review and committee report. The RIT was thus only included in an appendix due to the lack of availability of any other qualifying published study or replication on the RIT.[3] For APA members who use the RIT to remain in compliance with published standards after January 2015, a second study of the RIT meeting minimum APA research standards and requirements was required. Krapohl and Rosales (2014) published the results a second study on the RIT in an effort to meet that requirement.

Using the information provided by Krapohl and Rosales (2014) we created a traditional contingency table, shown here as Table 1, of outcomes for their 100 subjects. One half of those subjects were guilty[4] and half were innocent, establishing the common experiment base rate for deception of 0.50. To assess the diagnostic power of the RIT data in Krapohl and Rosales we calculated a detection efficiency coefficient (DEC; Kircher, Horowitz, & Raskin, 1988). The DEC is simply a direct application of the Pearson Product Moment Correlation to two vectors of data. In the first vector Deceptive, Inconclusive, and Truthful outcomes are coded -1, 0, and 1, respectively. In the second vector, Guilty and Innocent status is coded -1 and 1 respectively. Kircher et al., describe the DEC as follows:

**Table 1. Decision outcomes from Krapohl and Rosales (2014)**

| | Decision | | |
|---|---|---|---|
| Status | Deceptive | Inconclusive | Truthful |
| Guilty | 41 | 3 | 6 |
| Innocent | 27 | 5 | 18 |

[3] One other survey (Crewson, 2001) reported an average accuracy rate of .74 for RIT screening test results, using studies for which the study report and study data are unpublished, unavailable, and inscrutable, or which did not meet the minimum standards for inclusion in the meta-analytic survey.

[4] Throughout this paper we use the terms "deceptive," "truthful," and inconclusive to describe the outcomes of a PDD test. We use the terms "guilty" and "innocent" to describe the real-world status of the test subject. This usage follows the standard usage in scientific publications and serves to clearly maintain the distinction between real-world status and test results. We realize guilt and innocence in a legal setting is determined by the trier of fact. That usage is not represented here.

Our use of the correlation coefficient is based on the assumption that there is an underlying order to the polygraph interpreters' judgments, with inconclusive outcomes being treated as intermediate values along a truthful/deceptive continuum. Although inconclusive outcomes may be viewed as failures of the technique and their occurrence would reduce the value of the correlation coefficient, they would not be weighted as heavily as false positive or false negative decision errors. Thus, the correlation coefficient provides a measure of detection efficiency that is consistent with the real-world consequences of various types of polygraph outcomes. (p. 84)

The DEC for the RIT outcomes in Krapohl and Rosales (2014) was 0.305. Although this value does indicate better than chance discrimination of truthful and deceptive subjects ($p$ = 0.002) the discriminative power of the test is very low. To provide a contrast to other polygraph test data we provide detection efficiency statistics for some studies of the RIT, the Directed Lie Screening Test scored with ESS (American Polygraph Association, 2011), the Law Enforcement Pre Employment Test (Department of Defense Polygraph Institute, 2002; Raskin & Kircher, 2014), and some well known studies of the comparison question test in Table 2. Table 2 reveals the RIT to be

**Table 2. Detection Efficiency Coefficients for Selected Studies**

| Test Type | Study | DEC[1] |
|---|---|---|
| RIT | Horvath (1988)[2] | -0.14 |
| | Horowitz, Kircher, Honts, & Raskin (1997) | 0.38 |
| | Krapohl, Senter, & Stern (2005) | 0.48 |
| | Krapohl & Rosales (2014) | 0.30 |
| | **Average for the RIT** | **0.25** |
| DLST | APA (2011)[3] | 0.68 |
| LEPET | Raskin & Kircher (2014)[4] | 0.91 |
| CQT - Utah | Kircher, Horowitz, & Raskin (1988)[5] | 0.79 |
| Mixed | Raskin & Kircher (2014)[6] | 0.88 |
| Arther | Horvath (1977)[7] | 0.25 |

[1] Detection efficiency coefficient calculated with the methods described in Kircher et al., 1988.

[2] Data from the two visual analysis evaluators was combined. Insufficient information was provided to allow for the assignment of small number of inconclusive outcomes to guilt condition. They were distributed randomly.

[3] Data from DLST scored with ESS studies were combined from the data provided in the table in Appendix E-6.

[4] Scored with OSS-2, Table 3.4.

[5] Average DEC for the five studies with the highest generalizability scores, see Tables 1 and 2.

[6] Scored with OSS-2, Table 3.3.

[7] The small number of inconclusive outcomes were counted as errors. This DEC is thus a slight underestimate.

a very poor discriminator of truth and deception in comparisons to almost all of the other techniques. Interestingly, the average DEC for the RIT is the same as the DEC for the Arther Examination Procedure, a little used technique that was recently described in a case study published in this journal as lacking scientific validity and one whose use should be discontinued (Honts & Handler, 2013).

**Base Rate Considerations**

One concern when integrating PDD into a screening setting often centers on the base rate (Handler, et al., 2013; NRC, 2003). Base rate refers to the prior probability that a subject possesses the characteristic that is the target of the test, in this case deception. Base rate concerns are not unique to PDD testing; they occur anytime one employs an imperfect test and then has to make practical decisions that affect people in an applied setting.

While the base rate does not directly affect the accuracy of the PDD test, it does affect the confidence the end user has in the reported PDD test result (Honts, 1991). Table 3 illustrates the rates of correct and incorrect test result outcomes in situations with a low (10%) base rate of guilt. Although the accuracy of the test is high (90%), the confidence in a deceptive outcome is low as only 50% of those outcomes are correct.

**Table 3. A Conditional Probability Analysis of 900 Truthful and 100 Deceptive Applicants with a PDD Test that is Accurate 90 Percent with Both Guilty and Innocent Subjects**

| | Outcome | | |
|---|---|---|---|
| Actual Status | Truthful | Deceptive | Totals |
| Guilty | 10 | 90 | 100 |
| Innocent | 810 | 90 | 900 |
| Totals | 820 | 180 | 1000 |

The National Research Council (2003) expressed concerns that an unacceptable number of actually innocent people will be falsely found deceptive when testing subjects for low base of deception target issues. That valid concern is exacerbated as the base rate differential becomes more extreme. The National Research Council pointed out that if the base rate is low, then a large portion of positive test results will be false positive results with potentially adverse effects for those truthful subjects (also see Honts, 1991).

The usefulness of all diagnostic tests is related to the base rate. To be useful, test accuracy must exceed the accuracy rate achieved by simply predicting the base rate. The usefulness of a test is eliminated when the base rate exceeds the accuracy of test (Meehl & Rosen, 1955). In a seeming paradox, when the base rate exceeds the accuracy of a valid test, the use of that valid test may well reduce the accuracy of simply predicting the base rate.

There are metrics that define the useful range of diagnostic techniques. Wells and his colleagues (Wells & Lindsay, 1980; Wells & Olson, 2002) developed Information Gain (IG) analysis for use in assessing the information value of police eyewitness identification procedures. IG analysis uses Bayesian statistical analyses and empirical estimates of the accuracy of a test or technique to determine the information gain, if any, the test or technique provides over predicting the base rate of the target event. Honts & Schweinle (2009) adapted IG

analysis to polygraph testing. IG can thus be used to show the range of base rate over which polygraph testing can provide an increase in information to the end user. Moreover, IG analysis informs end-users and decision makers about the usefulness of using polygraph testing for various behaviors of varying base rate. Additionally, IG analysis can be prescriptive to a polygraph screening program by informing when not to use a particular technique. As we noted earlier, when the behavioral base rate estimates exceed the useful range of the polygraph, using the polygraph may actually increase the error rate of the decision maker over simply predicting the base rate.

Another consideration concerns how well we assess veracity without PDD testing (Handler, et al., 2013). Professional lie catchers often rely on their unassisted skills to tell when someone is lying to them. Base rate consideration notwithstanding, PDD testing would have to differentiate truthfulness from deception significantly better than the unassisted professional lie catcher to be considered an improvement. IG analysis can also be used to test if a polygraph technique outperforms the decisions of unassisted professional lie catchers across the range of base rates.

Honts and Schweinle (2009) calculated the range of base rates where PDD tests added incremental validity beyond the base rate and beyond the IG of unassisted assessments by lay persons and professional lie catchers in forensic and national security PDD settings. Handler, et al., (2013) used IG analysis to assess incremental gain of the Directed Lie Screening Tests (DLST; Handler, Nelson & Blalock, 2008), over the unassisted professional lie catcher. They found the DLST/ESS combination provided significant increases in IG over unassisted professional lie catchers across a range of base rates from 0.01 through 0.94 for deceptive outcomes and 0.07 through 0.99 for truthful outcomes.

The RIT technique has long been the subject of controversy (Podlesny & Raskin, 1977) due to having limited empirical support and near chance accuracy (APA, 2011; Raskin and Honts, 2002). Given a second peer-reviewed publication (Krapohl & Rosales, 2014) showed the RIT to have greater than chance unweighted accuracy, we sought to explore the usefulness of the RIT using IG analysis. Our goal is to contrast the RIT and the DLST using IG analysis to assess relative usefulness of the two techniques.

## Method

Using the accuracy data for the RIT from Krapohl and Rosales (2014) and the methods described by Honts and Schweinle (2009) we calculated IG for the RIT across a range of base rates from 0.01 to 0.99 inclusive. We then used Honts and Schweinle's (2009) calculations of the IG for professional lie catchers for comparison. Readers interested in the formulae and computational procedures for the calculation of IG are referred to Honts and Schweinle (2009).

# Results

Using the accuracy data for the RIT from Krapohl and Rosales, (2014; see Table 1) we calculated IG curves for test outcomes across the complete range of base rates. As seen in Figure 1, the IG for truthful outcomes peaked at a value of 0.268 at a base rate of deception = 0.63. The peak IG for guilty was .104 at a base rate of .45. Those results indicate that the RIT provides inferior IG across the range of base rates with both truthful and deceptive outcomes when compared to the Directed Lie Screening Test (Handler et al., 2013).

In the screening setting without PDD, it would be the interviewer who assesses veracity of the interviewee. Vrij (2008) has provided a compilation of the large amount of data assessing the professional lie catcher's ability to differentiate truth tellers from liars. Vrij (2008) reports on 31 studies of the unassisted professional lie catcher's ability to detect deception. The Vrij data were collected primarily from law enforcement, but some of the data were from immigration personnel who conduct entry interviews. Vrij (2008) reported 56.35% accuracy rate for recognizing truthful statements, and a 56.11% accuracy rate for recognizing deceptive statements. Using the Vrij (2008) data, Honts and Schweinle (2009) calculated the IG for

unassisted professionals and their results are illustrated in Figure 2. The IG for the unassisted professional lie catcher conclusion of truthful peaked at 0.06 at a base rate of deception = 0.53. The IG for the unassisted professional lie catcher conclusion of deception peaked at 0.06 at a base rate of deception = 0.47.

**Figure 1. Information gain curves for the RIT and the DLST for truthful, inconclusive and deceptive outcomes**

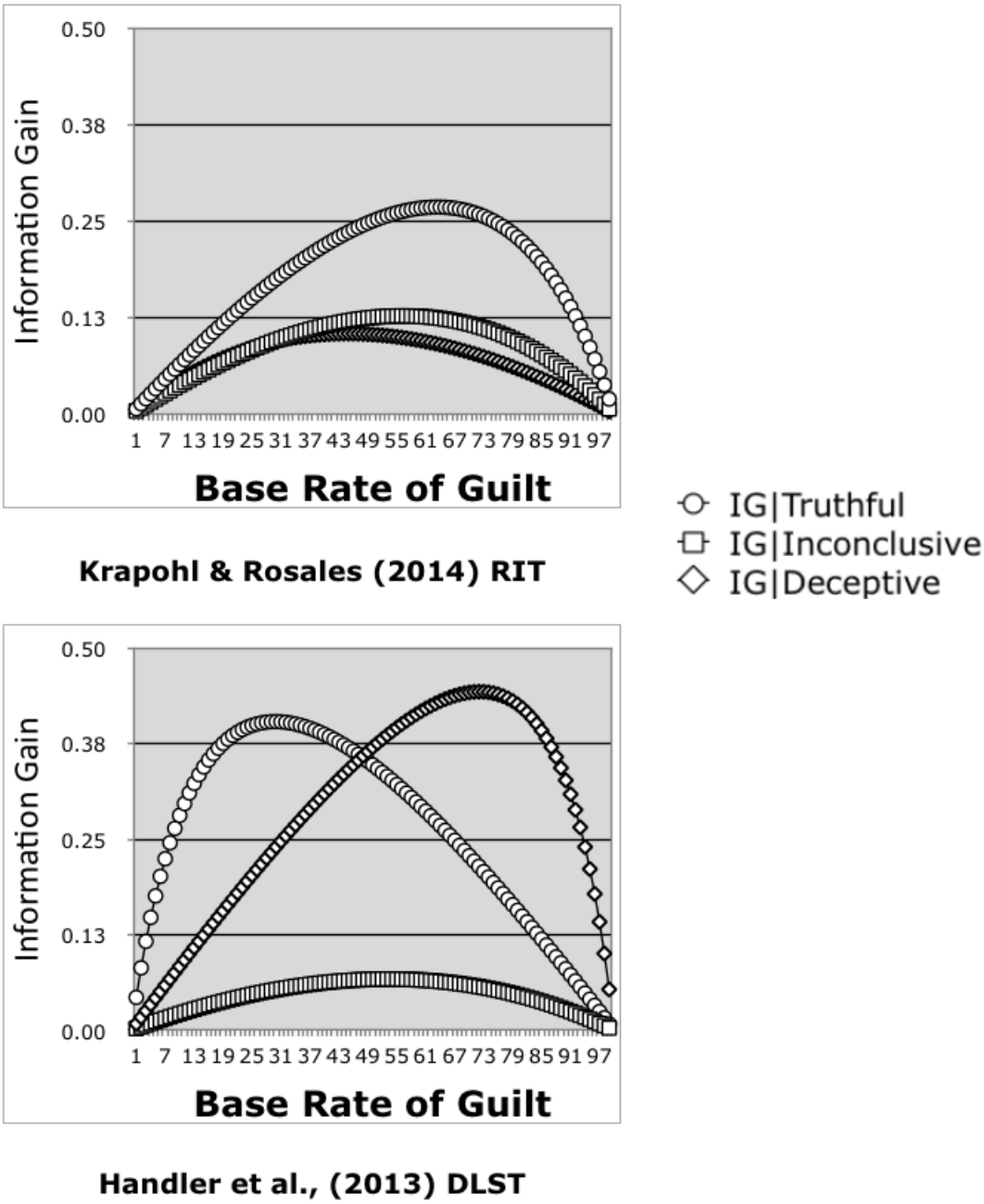
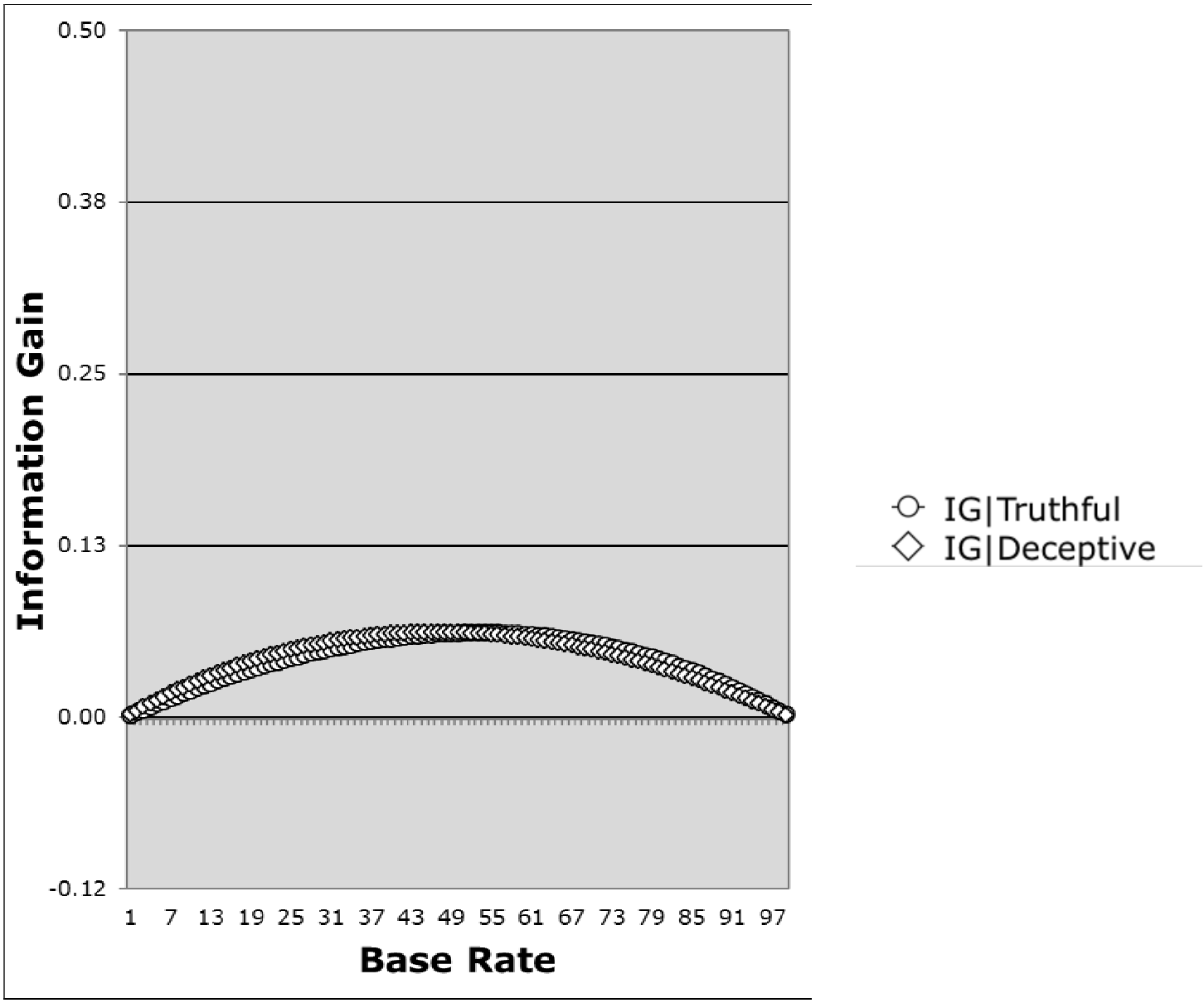


**Figure 2. Information gain for unassisted professionals for truthful and deceptive outcomes.**




We calculated improvement for the RIT over the unassisted professional lie catchers. In order to add incremental validity to the screening decision process, the RIT would have to provide more information than could be obtained by an unassisted professional lie catcher. Ideally, the RIT would provide IG over the unassisted professional that was as great as or greater than that of the DLST (Handler, et al., 2013). For truthful outcomes, there was a significant (1-tailed, $p < .05$) improvement over unassisted professionals for base rates ranging from 0.12 through 0.98 inclusive. At no base rate did deceptive outcomes from the RIT provide a significant improvement over an unassisted professional.

## Discussion

We adapted the Information Gain (IG) analysis procedures described by Honts and Schweinle (2009) to the Relevant – Irrelevant technique using data from Krapohl & Rosales (2014). We found that the RIT deceptive outcomes failed to provide IG over unassisted professionals at any base rate. Although truthful outcomes from the RIT were significantly more informative than truthful outcomes from unassisted professionals across a wide range of base rates, the amount of IG for the RIT was about half of the amount of IG provided by the DLST. The RIT data from Krapohl and Rosales (2014) provided inferior IG across the range of base rate with both truthful and deceptive outcomes when compared to the Directed Lie Screening Test (Handler, et al., 2013).

One concern is that uninformed readers of Krapohl and Rosales (2014) may misunderstand the high sensitivity to deception they reported for the RIT. Clearly to reach the conclusion that a deceptive outcome from an RIT is meaningful would be a mistake. What little information gain the RIT shows reflected a strong guilt bias. The result of the observed guilt bias was to inflate the correct hit (i.e., true positive) rate for guilty cases, with a corresponding reduction of false-negative errors. Although reduction of false-negative errors may be an important objective of screening programs, results of the RIT screening data from the Krapohl and Rosales (2014) study achieved this objective at the cost of a very weak specificity rate (36%) and a false-positive error rate that is so high (54%) that there is no improvement over unassisted lie detection or simply predicting the base rate. While there was some IG increase for innocent cases, when compared to unassisted lie detection or predicting the base rate, neither the observed false-negative rate nor the observed false-positive rate suggest any advantage over results using the DLST. The weak test accuracy of the RIT was clearly indexed by its low DEC as compared to other techniques. In practical terms this means that deceptive outcomes from a RIT are effectively uninterpretable or meaningless.

These results illustrate the potential importance and usefulness of the IG statistic for end users of polygraph test results. IG shows the usefulness of the test versus whether it is statistically better than chance. Paradoxically, the simple frequencies reported in the Krapohl and Rosales (2014) paper give the impression that a truthful result would be uninformative, because the rate of specificity to truth was low. However, IG curves explain a lot more about test performance expectations and test accuracy than simple percentages and frequencies. Our IG analysis shows that truthful outcomes added more useful information to the decision making process than an uninformed lie catcher could, although still half of that of the DLST.

The main limitations of the present analysis and also shared by the Krapohl and Rosales (2014) study, are a reliance on a non-random archival sample of field cases coupled with an analysis method that relies heavily on the subjective interpretation offered by expert practitioners. The effects of these factors on the generalizability of the results is unknown. An additional limitation is that our analysis does not consider the programatic ways in which the RIT screening test is used as an interview support tool. However, these results clearly indicated that it is not a good deception detection tool in any context.

These limitations notwithstanding, our findings suggest that program administrators and PDD examiners should endeavor to investigate and identify those techniques that actually add information to their decision making process. They should choose target behaviors that are likely to fall within the productive base rate range for PDD tests like the DLST/ESS. Careful and systematic consideration of both question targets and testing technique is warranted at level of the examiner and the program administrative level. Such consideration should be evidence-based and not impressionistic. Finally, both program administrators and field examiners should begin to address important questions surrounding the basis of an ethical justification for the use of a less effective technique when a more effective method is readily available with little or no difference in terms of physical, financial, or personnel resources.

The scientific community has expressed theoretical and evidence-based

concerns about the RIT since at least the 1970s (Podlesny & Raskin, 1977) while written concerns about the validity of the RIT in the polygraph profession go back much further (Reid, 1947). In all that time there has not been a single published study using field techniques that has shown high validity for the RIT. Two studies (Honts & Amato, 2007; Kircher, Woltz, Bell, & Bernhardt, 2006) that found moderate accuracy for the RIT, though statistical analysis have by all appearances been completely ignored by the profession.

During recent years there has been a growing awareness in the polygraph profession (for example, Hartwig, 2014) and the scientific community (Kassin et al., 2010) of the frequency of false-confessions, wrongful convictions, and the techniques that produced those miscarriages of justice. Along with the growing awareness in the scientific community, there has also been increased recognition by the courts of the importance of evidence-based practices in law enforcement. The courts have often responded with characteristically unsympathetic responses to professionals who chose methods known to be either suboptimal, highly flawed, or highly subjective.[5] Given the factors discussed above the usual recommendation for continued interest in, and additional research on, would seem to be folly with regards to the RIT.

Moreover, given the overwhelming scientific evidence of the inferior status of the RIT as compared to readily available and far more accurate alternatives, polygraph examiners and polygraph agencies that continue to make use of the RIT should be aware of their vulnerability to accusations of malpractice. Merriam-Webster (2014) defines malpractice as "a dereliction of professional duty or a failure to exercise an ordinary degree of professional skill or learning by one (as a physician) rendering professional services which results in injury, loss or damage." The selection of a highly subjective PDD testing method that provides inferior diagnostic accuracy and no IG over unassisted lie detection – in a era for which less subjective and more effective PDD methods are readily available - will undoubtedly be argued by some to fall within the scope of this definition as form of dereliction of duty to exercise adequate professional skill and learning.[6]

While the RIT was once regarded as a mainstay among PDD techniques, and although some found highly useful in the early years of PDD testing, that was prior to the development of far less subjective norm-referenced numerical and statistical analysis models. However considering the state of the science in 2014, the continued use of an inferior testing methodology is likely to become an increasing locus of professional and legal vulnerability. We anticipate a day, in the near future, when polygraph professionals who continue to make use of the RIT are frequently asked in a legal context to contended with the following question: How is the continued use of the RIT in 2014 for any substantive purpose not a form of malpractice?

[5] For an example of one such case involving flawed polygraph and interrogation techniques see Honts & Handler (2013). Two of the wrongfully convicted defendants in the case discussed by Honts and Handler were recently awarded 18 million dollars each in their wrongful conviction civil action against Nassau County New York (CBSNews/AP, 2014).

[6] For an excellent example of changing and evolving professional practice standards, consider that until the late 19th century physicians used bloodletting as a common accepted practice. Just before his death in 1799 George Washington was drained of 3.75 letters of blood as a treatment for a throat infection (Vadakan, 2004). While within the scope of reasonable professional practices at the time, today, because of scientific research it is known that bloodletting is not efficacious as a general medical treatment. Surely any physician today who treated a patient with a throat infection with bloodletting would be roundly denounced and subjected to litigation under the banner of malpractice.

## References

American Polygraph Association (2009a). Model policy for law enforcement/public service pre-employment screening examinations. [Electronic version]1/25/2012, available from http://www.polygraph.org.

American Polygraph Association (2009b). Model policy for post-conviction sex offender testing. [Electronic version]1/25/2012, available from http://www.polygraph.org.

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40, 196-305.

American Polygraph Association (APA). (2102, January 1). Bylaws and Standards of Practice. Retrieved May 23, 2014 from http://www.polygraph.org/section/about-us/laws-and-standards-practice.

CBSNewYork/AP, (2014). John Restivo and Dennis Halstead, Wrongfully Convicted Men, Awarded $36 Million. April 17, 2014 8:44 PM. Downloaded on June 7, 2014 from: http://newyork.cbslocal.com/2014/04/17/john-restivo-and-dennis-halstead-wrongfully-convicted-men-awarded-36-million/.

Consigli, J. (2002). Post-conviction sex offender testing and the American Polygraph Association, in M. Kleiner (Ed.), *Handbook of Polygraph Testing*, London, UK: Academic Press (pp. 237-250).

Department of Defense Polygraph Institute (2002). Law enforcement pre-employment test. Retrieved from: https://antipolygraph.org/documents/dodpi-lepet.pdf June 7, 2014.

Handler, M.D., Honts, C.R., & Nelson, R., (2013). Information gain of the directed lie screening test. *Polygraph*, 42, 192-202.

Handler, M. D., Honts, C. R., Krapohl, D. J., Nelson, R., & Griffin, S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Police and Criminal Psychology*, 24, 69-86. (Published online 20 May 2009.)

Handler, M., Nelson, R., & Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph*, 37, 100-111.

Hartwig, M. B., (2014, April). The psychology of interrogations. Invited lecture at the American Association of Police Polygraphists annual meeting, Las Vegas, Nevada.

Honts, C. R., (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports*, 4, 91-116.

Honts, C. R., & Amato, S. (2007). Automation of a screening polygraph test increases accuracy. *Psychology, Crime & Law*, 13, 187-199. (Available online in 2006).

Honts, C. R., & Handler M. (2013). A case study of the validity of the Arther Examination Procedures in a criminal case with DNA confirmation. *Polygraph*, 42, 61-71.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (2008). Scientific status: The case for polygraph tests. In, D. L. Faigman, M. J. Saks, J. Sanders, and E. Cheng (Eds.) *Modern scientific evidence: The law and science of expert testimony (Volume 5): 2008-2009 Edition*. Thompson West: Eagan, Minnesota.

Honts, C. R., & Schweinle, W. (2009). Information gain of psychophysiological detection of deception in forensic and screening settings. *Applied Psychophysiology and Biofeedback*, 34, 161-172. (Available online July 2009).

Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.

Horvath, F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16, 198-209.

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior*, 34, 3-38.

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.

Kircher, J. C., Woltz, D. J., Bell, B. G., & Bernhardt, P. C. (2006). Effects of audiovisual presentations of test questions during relevant-irrelevant polygraph examinations and new measures. *Polygraph*, 35, 25-54.

Krapohl, D., Senter, S., & Stern, B. (2005). An exploration of methods for the analysis of multiple-issue relevant/irrelevant screening data. *Polygraph*, 34, 47-61.

Krapohl, D., & Rosales, T. (2014). Decision accuracy for the relevant-irrelevant screening test: A partial replication. *Polygraph*, 43, 1-19.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216.

Merriam-Webster (2014). Malpractice. Retrieved from http://www.merriam-webster.com/dictionary/malpractice June 1, 2014.

National Research Council (2003). The polygraph and lie detection. Washington, DC: The National Academies Press.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and detection of deception. *Psychological Bulletin*, 84, 782-799.

Raskin, D. C. & Honts, C.R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London, UK: Academic Press.

Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law, Criminology and Police Science*, 37, 542-547.

Vadakan, V. V., (2014). The asphyxiation and exsanguinating death of President George Washington. *The Permanente Journal*, 8(2), 76-79.

Vrij, A. (2008). *Detecting Lies and Deceit, Pitfalls and Opportunities*, 2nd Edition. West Sussex, England: John Wiley & Sons.

Wells, G. L. & Lindsay, R.C.L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological*, 88, 776-784.

Wells, G. L. & & Olson, E.A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, 8, 155–167.