Information Gain of the Directed Lie Screening Test

Mark Handler, Charles R. Honts, and Raymond Nelson[1]

Abstract

We adapted the Information Gain (IG) analysis procedures described by Honts and Schweinle (2009) to the Directed Lie Screening Test (DLST; Handler, Nelson & Blalock, 2008), a comparison question technique (CQT) variant, based on the Test for Espionage and Sabotage (TES; Research Division Staff, 1995a; 1995b), evaluated using the Empirical Scoring System (ESS). IG provides an index of the increase in information gained by using a test result over predicting the base rate of the target event. The DLST is used in a number of screening settings, including public safety applicant screening and post-conviction sex offender monitoring. We found that 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. For NSR results, the IG exceeded 5% across the base rate range of 0.06 through 0.99. For SR results IG exceeded 5% across the base rate range of 0.02 through 0.93. Our analyses provide empirical support to the hypothesis that adding the polygraph component to a screening process adds incremental validity. The findings also remind us of the relative limitations of unassisted professional lie catchers in veracity assessment compared to using polygraph testing. These findings should be of interest to program managers, risk assessors and other consumers of polygraph testing.

Introduction

Psychophysiological Detection of Deception (PDD), also known as polygraph, screening tests are routinely included in the evaluation process of applicants for employment in the public safety setting (Handler, Honts, Krapohl, Nelson & Griffin, 2009; Honts, Raskin & Kircher, 2008; National Research Council, 2003) as well as post-conviction sex offender monitoring (Consigli, 2002). Screening tests are any tests conducted in the absence of a known or reported problem. Screening tests can consist of a single or multiple target issue(s), depending on situational needs. Recently the American Polygraph Association published a meta-analytic survey of the criterion accuracy of a number of polygraph techniques (APA, 2011). That analysis included data from diagnostic and screening applications of selected PDD techniques, including the Directed Lie Screening Test (DLST), evaluated using the Empirical Scoring System (ESS). The results from the APA (2011) review concur with previously reported evidence that PDD

[1] About the Authors

Mark Handler is the research chairperson for the American Association of Police Polygraphists, has published articles and studies on various aspects of the polygraph. Correspondence should be directed to polygraphmark@gmail.com.

Charles Honts received his Ph.D. in Experimental Psychology from the University of Utah in 1986, and joined the Boise State University Psychology faculty in 1995. Dr. Honts has current research interests in Psychology and Law that involve scientific credibility assessment with adults and children, specifically the psychophysiological detection, interrogation techniques and confession phenomena. Dr. Honts has testified as an expert witness in a number of US courts and in international venues. He has also authored numerous peer-reviewed publications and edited chapters (see http://charleshontsphd.com).

Raymond Nelson is a research specialist with the Lafayette Instrument Company and an elected member of the APA Board of Directors. The views expressed in this work are those of the author and not the LIC or the APA. Mr. Nelson is a psychotherapist, polygraph field examiner, developer of the OSS-3 scoring algorithm, and is the author of publications on various polygraph topics.

examinations can provide high levels of criterion validity (Abrams, 1989; Honts & Peterson, 1997; National Research Council, 2003; Office of Technology Assessment, 1983; Raskin & Honts, 2002). Comparison question technique (CQT) testing of multiple issues, for which the criterion states of individual test questions are assumed to vary independently, produced an aggregated decision accuracy rate of 0.850 (95% CI 0.773-0.926). The DLST/ESS produced an aggregated unweighted decision accuracy of 0.858 (95% CI 0.786-0.930). A complete description of the accuracy for the aggregated results of the four DLST/ESS studies is listed in Table 1.

The Base Rate Phenomenon

One concern when integrating PDD into a screening setting often centers on base rate phenomenon (NRC, 2003). The term base rate refers to the frequency with which the target issue appears in the testing population. In PDD screening, this would be the prior probability of a test subject having engaged in the behavior targeted by the test's relevant questions. Base rate concerns are not unique to PDD testing, they occur anytime one employs an imperfect test and then has to make decisions that affect people in a real world 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. Consider the following situation adapted from Honts (1991): A PDD test that is 90% accurate with both truthful and deceptive applicants, and a situation where exactly half of the applicants are deceptive (base rate of deception is 0.50). Table 2 illustrates the rates of correct and incorrect test result outcomes in this situation. With the base rate of truth and deception equal, interpretation of the outcome is balanced, 90% of the Significant Response outcomes are correct and 90% of the No Significant Response outcomes are correct.

Next consider a situation where very few of the applicants may be attempting deception, say an organization that asks about serious felony crimes. Table 3 illustrates the correct and incorrect test result outcomes in this situation using that same 90% accurate PDD test for both truthful and deceptive applicants, but with a base rate of deception of 0.10. Now 98.8% of those who produce No Significant Response on the test were in fact truthful but only 50% of those who showed Significant Response were deceptive.

An opposite pattern of results will occur when the base rate of deception is high: a high proportion of true-positive results and a high proportion of false negative errors among the persons who produce negative results. The high base rate produces quite a different situation than when the base rate was equal; base rate considerations are important for the persons who make decisions that include use of the PDD test result as part of their decision making process.

The National Research Council (2003) expressed concerns that an unacceptable number of actually innocent people will be falsely found deceptive when testing subjects with low base rate issues and this concern is exacerbated when the population tested is large. They point 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.

Two counter-arguments can be raised in answer to this criticism that can support the use of PDD testing in spite of the base rate phenomenon. Counterargument one can be made if the PDD test result were to add incremental validity to any decision making process at the base rate of interest for the target behavior. In other words, is a decision maker able to make more accurate judgments using the PDD test result for a given base rate then they would if they simply made their decision based on the base rate alone? One practical limitation of a simple actuarial decision making approach from base rates requires that the base rates be known. In reality, base rates for issues in polygraph screening tests are not known with precision. If there were a way to estimate the base rate where PDD tests are used, then a metric would be needed to assess the advantage of any diagnostic test. Information Gain (IG) analysis (Honts & Schweinle, 2009) can be used to provide such a metric.

Table 1. Dimensional Profile of Criterion Accuracy for DLST/ESS (APA, 2011)

Technique /TDA Method	DLST/ESS
Number of Studies	4
N Deceptive	149
N Truthful	149
Total N	298
Number Scorers	5
N of Deceptive Scores	174
N of Truthful Scores	173
Total Scores	347
Mean D	-2.131
StDev D	3.801
Mean T	3.412
StDev T	3.153
Reliability - Agreement	.840
Unweighted Average Accuracy	.858 (.037); {.786 to .930}
Unweighted Average Inconclusives	.090 (.026); {.039 to .142}
Sensitivity	.809 (.069); {.674 to .945}
Specificity	.751 (.031); {.691 to .811}
FN Errors	.112 (.057); {.001 to .224}
FP Errors	.146 (.027); {.093 to .2}
D INC	.078 (.052); {.001 to .180}
T INC	.102 (.014), {.075 to .130}
PPV	.848 (.041).; {.767 to .928}
NPV	.870 (.052); {.768 to .971}
D Correct	.878 (.067); {.746 to .999}
T Correct	.837 (.027); {.783 to .891}

(Standard deviations are in parentheses and 95% CI are in brackets.)

Table 2. A conditional probability analysis of 500 truthful and 500 deceptive applicants with a PDD test that is accurate 90% of the time.

PDD Test Outcome	Actual Status		
	Truthful	Deceptive	Totals
Significant Response	50	450	500
No Significant Response	450	50	500
Totals	500	500	1000

Table 3. A conditional probability analysis of 900 truthful and 100 deceptive applicants with a PDD test that is accurate 90% of the time.

PDD Test Outcome	Actual Status		
	Truthful	Deceptive	Totals
Significant Response	90	90	180
No Significant Response	810	10	820
Totals	900	100	1000

The second counterargument revolves around the methods for verifying subjects' veracity without PDD testing. Typically, veracity assessment professionals (i.e., police investigators, judges, employment specialists) rely on their unassisted skills to assign people to truthful or deceptive categories. 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 explore whether or not DLST/ESS outperforms the decisions of unassisted professional lie catchers across a range of base rates.

Information Gain Analysis

Wells and his colleagues (Wells & Lindsay, 1980; Wells & Olson, 2002) developed IG analysis for use in assessing the information value of police eyewitness identification procedures. IG analysis uses Bayesian statistical analyses[2] and empirical estimates of the accuracy of a test or technique to determine the information gain, if any, the test or technique provides when predicting the base rate of the target event. IG ranges from zero (no information gained above the base rate) and can approach but never achieve a value of 1. Usually the IG is presented as a curve of information gain plotted against the base rate of the target behavior, for example deception or guilt status.

The computational steps follow:

(1) Find an empirically or theoretically derived estimate of the BR of deception for suspects that are tested with PDD, i.e. p(deceptive).

(2) Find an empirically or theoretically derived estimate of the probability that a suspect will produce a significant response (SR) outcome given that he or she is deceptive. i.e. p(SR|deceptive). This estimate can be based on previous research on the decision model, and is known as test sensitivity.

(3) Calculate an estimate of the proportion of truthful subjects who are tested with PDD, (i.e. p(truthful) or 1-p(deceptive)).

(4) Find an empirically or theoretically derived estimate of the proportion of truthful subjects who can be predicted to produce "SR" PDD results, (i.e. p(SR|truthful)); this is the false-positive (FP) rate.

(5) Compute the conditional probability of interest, i.e. p(deceptive| PDD result), using Bayes' Theorem.

To calculate the IG, simply subtract the BR from the conditional probability (PPV). This reflects how much a given PDD result informs us about the truthfulness or deception of suspects above and beyond the base rate estimate of a suspect's truth or deception.

Honts and Schweinle (2009) adapted the IG statistic to the PDD testing situation. Honts and Schweinle (2009) calculated the range of base rates where PDD tests added incremental validity beyond the base rate and beyond the unassisted assessment ability of lay persons and professional lie catchers in forensic and national security PDD settings. They compared empirically derived estimates of how well lie catchers differentiate truth from deception without the aid of a polygraph instrument to polygraph testing results in both diagnostic and screening settings. They found that in forensic settings, PDD provided substantial improvements in information gain over unassisted laypersons across nearly the complete range of the base rate of guilt. Their analyses of screening PDD indicated that only deceptive outcomes provide useful information gain at relevant low base rates of guilt. Their results strongly support the use of PDD in diagnostic settings and raised concerns for how screening PDD results are used.

As the DLST/ESS is being presented at many national PDD training seminars we sought to extend the work of Honts and Schweinle (2009) to compare the IG using DLST/ESS screening exams to presently available evidence regarding the decision accuracy of unassisted professional lie catchers. The present study addressed two questions: 1) Is the information gain provided by the DLST/ESS more than zero? and 2) Does the DLST/ESS provide more IG as compared to unassisted professional lie catchers?

[2] Bayes' Theorem: [deceptive|SR) = p(SR|deceptive)(p(deceptive) ÷ p(SR|deceptive) p(deceptive) + p(SR|truthful)p(truthful)] is the proportion of true-positive results to all positive results (i.e., TP/(TP+FP)) and is sometimes referred to as the positive-predictive-value (PPV).

Method

Application of IG Analysis to PDD Screening

Using the accuracy data for the DLST/ESS from APA (2011), and the methods described by Honts and Schweinle (2009) we calculated IG for the DLST across a range of base rates from 0.01 to 0.99 inclusive. Calculations were made at intervals of 0.01. We then used Honts and Schweinle's calculations of the IG for professional lie catchers for comparison. Readers interested in the formulae for the calculation of IG are referred to Honts and Schweinle and to Wells and Olson (2002).

There are three possible outcomes of a screening PDD test: Significant Response (SR), No Significant Response (NSR) and Inconclusive, and there are two possible states of reality: the subject is either truthful or deceptive. These outcomes and states of reality can be depicted in a 2 x 3 matrix (see Table 4) that combines each PDD result with each state of reality.

Results

IG for the DLST/ESS

We calculated the IG for truthful and deceptive cases using data for the DLST/ESS from the four studies APA (2011; see Table 1). As seen in Figure 1, the IG for NSR outcomes had a negative skew and peaked at a value of 0.443 at a base rate of deception = 0.72. For NSR results, the curve exceeded 5% IG across the base rate range of 0.06 through 0.99. The curve for SR outcomes was positively skewed with a peak value of 0.404 at a base rate of deception = 0.30. For SR results IG exceeded 5% IG across the base rate range of 0.02 through 0.93. The IG curve for inconclusive outcomes was symmetrical and relatively flat exceeding 5% IG only across a narrow range at the middle of the curve. Those results indicate that with the DLST/ESS inconclusive outcomes provide little information.

IG for Professional Lie Catchers

To add incremental validity to the screening decision process, the DLST/ESS would have to provide more information than could be obtained by an unassisted professional lie catcher. In the screening setting 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 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 did entry interviews. Vrij reported 56.35% accuracy rate for recognizing truthful statements, a 56.11% accuracy rate for recognizing deceptive statements for an overall unweighted average accuracy of 56.23. Using the Vrij data, Honts and Schweinle (2009) calculated the IG for unassisted professionals. The results from Honts and Schweinle 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

Table 4. PDD outcomes and their associated probability notations.

PDD Test Outcome

Actual State

	SR	NSR	Inconclusive
Truthful	p(SR|truthful)	p(NSR|truthful)	p(inconclusive|truthful)
Deceptive	p(SR|deceptive)	p(NSR|deceptive)	p(inconclusive|deceptive)

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. For NSR results IG from the DLST/ESS provided significantly more information ($p < 0.05$) than the unassisted professional lie catcher across the base rate of deception from 0.07 through 0.99. For SR results, the IG from the DLST/ESS provided significantly ($p < 0.05$) more information than the unassisted professional lie catcher across the base rate of deception from 0.01 through 0.94.

Figure 1. Information gain for the DLST/ESS for three PDD outcomes.

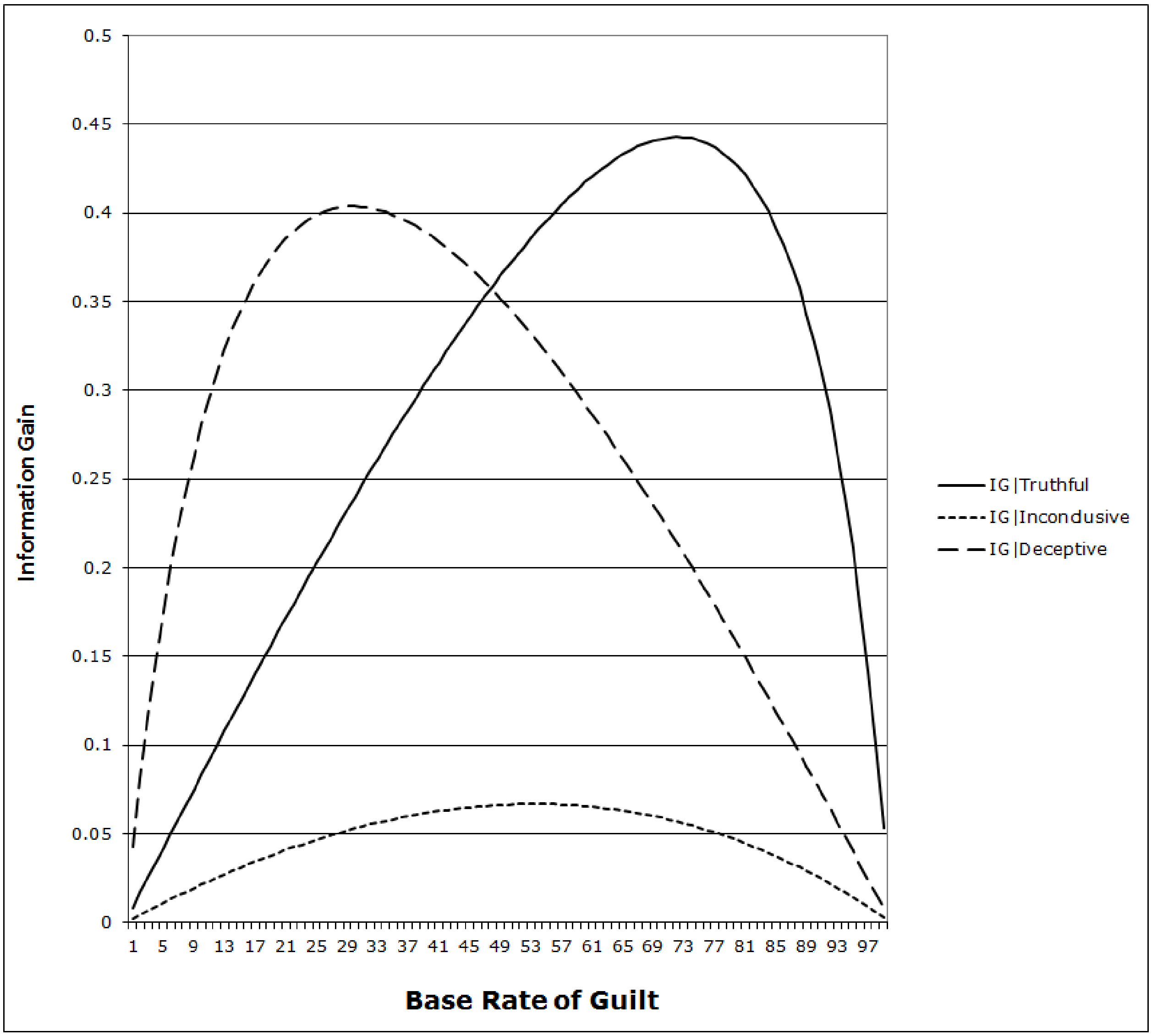
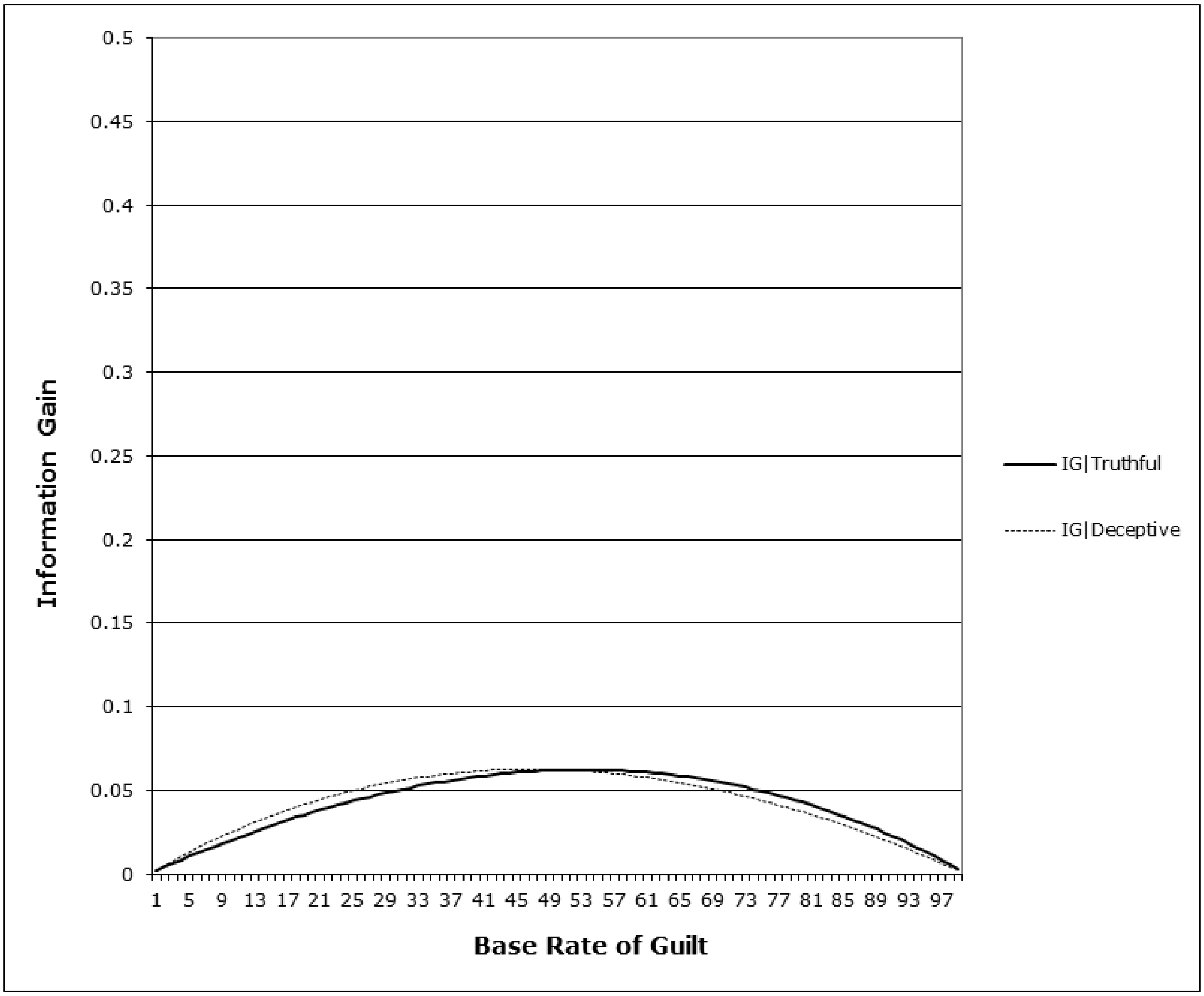


Figure 2. Information Gain of unassisted professional detection of deception with truthful and deceptive individuals across base rate of deception (adapted from Honts & Schweinle, 2009)



Discussion

IG analysis showed that the DLST/ESS provided greater (> 5%) information than simply predicting the base rate across a broad range. Moreover, the analysis indicates that the DLST/ESS provides significantly more information gain over unassisted professionals at all values except at extreme base rates where neither approach is able to outperform the simple prediction of the base rate. Those results clearly show that the DLST/ESS technique adds incremental validity above the ability of professionals trained in unassisted deception detection in a wide range of situations. Decision and policy makers tasked with incorporating deception detection results into their decision processes can assume that so long as the behavior targeted for credibility assessment falls within the broad range of significant IG for the DLST/ESS, their decision making process should be augmented by including the results of that test over the decisions of unassisted professionals. Our findings suggest that program administrators and PDD examiners should endeavor to investigate and identify target behaviors that are likely to fall within the productive base rate range for PDD tests like the DLST/ESS. Careful consideration of test question targets with regard to base rate is clearly warranted.

The null findings related to inconclusive test results are also informative. The Test for Espionage and Sabotage (Research Division Staff, 1995a; 1995b) data used by Honts and Schweinle (2009) reported no inconclusive outcomes for deceptive subjects and thus they did not calculate IG curves for inconclusive screening studies. However, they reported IG calculations for inconclusive results from several high quality field studies. Honts and Schweinle reported that for forensic CQT examinations, inconclusive outcomes provided significant IG that the person was in fact actually innocent with IG peaking at a value of 0.26 where p(deceptive) = 0.63. The authors noted this reflected the unbalanced inconclusive rates in the field studies where inconclusive results were nearly three times more likely to have come from truthful subjects (23.4%) as from the deceptive subjects (8.2%). The inconclusive results in the DLST/ESS studies were balanced and lower than those from the previously reported field studies with deceptive group estimate at 7.8% and truthful group inconclusive at 10.2%. Our IG analysis indicated that inconclusive outcomes did not provide any information gain with the DLST/ESS. However, our data came from controlled settings and it is equally important to collect validation data from field settings for the DLST/ESS. Until those data are collected and reported, the results of the present study should be considered with the knowledge that field replication has not yet been obtained. Nevertheless, our IG data suggest that it is unwise to make any inference following an inconclusive result with the DLST/ESS as such an outcome was just as likely to have come from a truthful individual as from a deceptive individual.

We suggest IG analysis is a useful metric for estimating those effective base rates across which a particular PDD technique is productive. Knowing this, one can make an effort at targeting behavior estimated to exist within the confines of those effective base rates. We recommend future studies of PDD techniques consider incorporating an effective range of base rates for information gain in their analysis. Arming consumers, decision makers and risk assessors with an IG analysis can improve the estimate of weight given any test result. Finally, the results suggest a continued interest in applying PDD testing to the screening milieu. Unassisted efforts at differentiating truth from deception are simply not supported empirically, while the use of PDD testing is. Relying on our human ability to differentiate truth from deception does not seem warranted, as painful as this may be to accept. We have a professional responsibility to our test subjects and consumers to provide them with evidence-based testing techniques and evidence-based evaluations of the test result. We believe IG analysis is a step in that direction.

References

Abrams, S. (1989). *The Complete Polygraph Handbook*. Lexington Books.

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40(4), 196-305.

Consigli, J. (2001). Post-conviction sex offender testing and the American Polygraph Association. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 237-250). New York: Academic Press.

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(2), 100-111.

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. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer. Available from the author.

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

National Research Council. (2003). The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. Office of Technology Assessment (1983). The validity of polygraph testing: A research review and evaluation. Washington, D. C.: U.S. Congress, Office of Technology Assessment.

Raskin, D. C. & Honts, C.R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. Academic Press.

Research Division Staff (1995a). A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats. Report number DoDPI94-R-0008. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 26(2), 79-106.

Research Division Staff (1995b). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. DoDPI94-R-0009. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 27(3), 171-180.

Vrij, A. (2008). *Detecting Lies and Deceit, Pitfalls and Opportunities*, 2nd edition. Wiley & Sons.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88(3), 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.