Analyzing Iacono's Thought Experiment About Polygraph Field Studies: Reason or Fantasy?

Article · October 2019

CITATIONS
0

READS
11

2 authors:

Charles Honts
Boise State University
73 PUBLICATIONS **1,177** CITATIONS
SEE PROFILE

Steven Thurber
Child and Adolescent Behavior Health Services Minnesota Department of Human Se…
118 PUBLICATIONS **705** CITATIONS
SEE PROFILE

Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice

VOLUME 48 **2019** **NUMBER 2**

Contents

Published Semi-annually

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice

Subscription information: *Polygraph* is published semi-annually by the American Polygraph Association. Editorial Address is Editor@polygraph.org. Subscription rates for 2019: One year $150.00 (Domestic). Change of address: APA National Office, P.O. Box 8037 Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice* DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

Analyzing Iacono's Thought Experiment About Polygraph Field Studies: Reason or Fantasy?

Charles R. Honts

Boise State University

and

Steven Thurber

Minnesota Department of Human Services

Abstract

We review and analyze a thought experiment first published in Iacono (1991) and reintroduced in Iacono and Ben-Shakhar (2019). The Iacono Thought Experiment (ITE) appears to have used backtracking methods to generate a series of assumptions and preconditions which would make it possible to have a polygraph test with chance accuracy that produces a confession-criterion field study with high accuracy. From this thought experiment, Iacono promulgated a hypothesis that all polygraph confession criterion studies produce exaggeratedly high estimates of accuracy. Our analysis of the assumptions and preconditions of the ITE found them to be unrepresentative and highly unlikely to be met in real world settings. We used a converging evidence approach that applied meta-analytic results, field studies that did not use a confession criterion, and data from wrongful conviction cases that involved polygraph examinations to test the Iacono hypothesis. We found strong falsification evidence to the Iacono hypothesis and conclude that it should be abandoned as a meaning description of field polygraph research.

Keywords: Ocular-motor, deception detection, eye tracking, reading

Analyzing Iacono's Thought Experiment About Polygraph Field Studies: Reason or Fantasy?

Polygraph tests represent an important and widespread application of a psychological test in law enforcement, national security, and employment around the world. Internationally, the American Polygraph Association shows members from 62 countries. Zhang (2011) estimated that there were as many as 8000 polygraph examiners operating in China alone. Despite the ubiquitous nature of polygraph testing, it has received relatively little attention in academic psychology and often, that attention has been in the form of negative commentary.

The most commonly used - and criticized - polygraph test is the Comparison Question Test (CQT). The CQT comes in sev-

Author Note

Correspondence should be addressed to Charles R. Honts, Ph. D., Department of Psychological Science, Boise State University, 1910 University Drive MS-1715, Boise ID 83725-1715. The authors would like to thank Adela Stephanescu for her help in editing the completed manuscript.

eral variants, but in all cases, it monitors the subject's autonomic physiology (usually, respiration, electrodermal activity, relative blood pressure, and often peripheral vasomotor activity) while the subject answers a series of questions. The question series contains two categories of critical questions (usually two or three of each). Relevant questions directly address the matter under investigation. Comparison questions are designed and presented in such a way that every subject lies, or is assumed to lie, in their response to them during the test. The subject's physiological responses are expected to show an interaction, so that subjects who are deceptive to the relevant questions show larger physiological responses to relevant questions as compared to their comparison questions. Subjects who are being truthful to the relevant questions are expected to show the opposite pattern, with physiological responses to comparison questions being larger than those to relevant questions.

There are a number of reviews of CQT research, typified by but not limited to the following examples: Raskin, Honts, and Kircher (1997), Iacono & Lykken (1997), The National Research Council (NRC, 2003), Honts (2004), Vrij (2008); American Polygraph Association (2011); Raskin, Honts, and Kircher (2014), and Iacono and Ben-Shakhar (2018). There is variation across the reviews, but nevertheless they generally produced overall accuracy estimates over 80%.

However, all of those reviews can be criticized for selective study choices and a lack of meta-analytic scrutiny. None attempted to test moderator variables, although they sometimes reached conclusions that hypothesized or even assumed powerful moderator effects. The NRC (2003) report was particularly egregious in that regard. NRC found objectively high discrimination estimating the area under the Receiver Operating Characteristic (ROC) curve (*AUC*) at 0.89. The use of ROC analysis and *AUC* as an effect size has been criticized as an inappropriate application of a technology developed to examine the performance of signal detection with technology (specifically an operator's ability to view RADAR screens and distinguish enemy ships, friendly ships, and noise; Tape, 2019) to psychology in general (Balakrishnan, 1999), and to polygraph testing in particular (Honts & Schweinle, 2009).

Nevertheless, the NRC used *AUC* as their index of effect size, but what does AUC actually mean and what does an *AUC* value of 0.89 imply about polygraph performance? The value of *AUC* can range from 0.50, which represents chance performance, to 1.00 which represents perfect performance (100% accuracy; Tape, 2019). Tape (2019) qualitatively characterizes *AUC* values between 0.80 and 0.90 as indicating a good discriminator and *AUC* values above 0.90 as excellent. Tables (Rice & Harris, 2005) and software (DeCoster, 2012) to convert between *AUC* and other measures of effect size are readily available. Reference to those tables and software show that an *AUC* value of 0.89 corresponded to a *Cohen's d* value of 1.74 and an r_{pb} of 0.66. Cohen (1969, 1988, 1992) described large effects in psychology as those with d values above 0.80 (corresponding $r_{pb} > 0.49$). Cohen famously said that, in applied psychology, effect sizes of d = 0.8 are "about as high as they come" (Cohen, 1988, p. 81). Thus, the *AUC* effect size reported by NRC (2003) indicates extremely high performance for the CQT as compared to other psychological tests and measures.

In spite of powerful empirical evidence of the usefulness of the CQT as a discriminator of truth and deception, the NRC discounted those findings, saying the research methods were substandard. To the present authors this seems to be an arrogant conclusion as the NRC substituted their judgment about the qualities of research published in first tier peer-reviewed journals of psychological science. Such a position is insulting to the editors of those first-tier journals and the working scientists who peer-review for them. The NRC's opinion is all the worse for the fact that none of the members of the NRC committee who wrote the report had ever published a study on deception detection.

Additionally, the NRC and others were notably critical of the use of experimental (laboratory) studies for assessing CQT validity. Iacono & Lykken (1997) completely dismiss the experimental research on the CQT, arguing that the real-world motivational contexts could not be modeled experimentally and therefore laboratory results were qualitatively different from those in real cases. However, the conclusions of the NRC and Iacono & Lykken (1997) about the generalizability of the

research methods in archival peer-reviewed journals of psychological science and the generalizability of experimental CQT research should all be viewed as opinion and not as fact, as none of those opinions were or are data-based.

The history of academic disagreement over the accuracy (criterion validity) of the CQT is long and has at times been polemic. Those disagreements are typified by, but are not limited to, published exchanges between researchers from the University of Utah and the University of Minnesota, beginning in 1978 in the journals *Psychophysiology* (Raskin and Hare, 1978; Lykken, 1978; Raskin, 1978) and *Psychological Bulletin* (Lykken, 1979; Raskin and Podlesny, 1979). Direct exchanges in the literature between these groups continued until 2002 (Honts, Raskin, and Kircher, 2002; Iacono and Lykken, 2002). Those disagreements were argued at a number of different levels on various topics. Throughout the disagreements, the Minnesota group radically rejected the notion that deception detection could be validly modeled in the laboratory and held that the results of laboratory studies were not useful for estimating criterion validity in field applications because they lacked external validity (generalizability). The Minnesota group holds that position until today despite the general rejection of such criticisms across the entirety of Psychological Science and specifically for deception detection. Hartwig and Bond (2014) provided a general discussion about generalizability of laboratory studies and provided a specific empirical rejection of differences between experimental and field settings, within a meta-analysis of the interpersonal deception detection research literature.

The scientific issues surrounding the contrast of experimental and field settings for research in interpersonal deception detection are nearly identical to those with the CQT. Recently, Honts and Thurber (2019) reported a comprehensive meta-analysis of the CQT that followed the analytic approach of Hartwig and Bond (2019). Honts and Thurber (2019) reported no statistically detectable effects for moderators of motivation, subject population or setting (experiment vs. field) in their comprehensive meta-analysis of the CQT.

The Minnesota group was initially supportive of field studies that fit their criterion for useful field studies. However, starting in the 1980s, field studies were published that produced high levels of accuracy with the CQT (Honts & Raskin, 1988; Raskin, Kircher, Honts, & Horowitz, 1988). Those studies were specifically designed to meet the Minnesota group's criteria. Subsequently, the Minnesota group rejected all field studies and took the radical position that valid research on the CQT could not be conducted. One keystone of that position was a thought experiment first reported by Iacono (1991) and then with some modification reintroduced in Iacono and Ben-Shakhar (2018)[1]. The Iacono (1991) thought experiment was originally presented as follows:

> Suppose that 800 crimes are being investigated using a polygraph technique that operates with exactly chance accuracy; i.e. half of both the guilty and innocent suspects will fail and half will pass. Because the polygraph is often used in crimes for which there are multiple suspects, let us assume, without loss of generality, that we are dealing with 800 two-suspect crimes, and that for each, one suspect is guilty and the other innocent. Let us assume further that (1) the guilty suspect is tested first 50% of the time, (2) the second suspect will not be tested if the result of the first test indicates deception, (3) neither innocent suspects nor those guilty suspects who pass the test will confess, and (4) 20% of the guilty who fail the test and are subsequently interrogated confess. (Iacono, 1991, pp. 202-203).

[1]In Iacono and Ben-Shakhar provided two simplified versions of the Iacono Thought Experiment with single subjects and with paired subjects. Assumptions 3, 4, and 5 do not apply to either the single subject or the paired tests as all subjects are tested regardless of the outcomes. Confession rates are not specified for either analysis thus Assumption 7 is not specific. The other assumptions are either explicit or implied in the latter version of the Iacono Thought Experiment.

Thought experiments are well known in philosophy and science. Thought experiments can be defined as "devices of the imagination used to investigate the nature of things" (Brown, 2014). One of the most famous scientific thought experiments was Galileo's reasoning that two objects of different weight must fall at the same speed. Galileo's thought experiment is easily validated from observations, such as the conclusive demonstration by Neil Armstrong on the moon when he dropped a feather and a hammer simultaneously and they landed on the surface at the same time (Pigllucci, 2006). Pigllucci further notes that thought experiments can also be wrong and be falsified by data. Had Galileo's thought experiment been invalidated by data, it would have been lost to history and forgotten.

Thought experiments can take on a number of forms or types. While a discussion of the multiple types of thought experiments is beyond the scope of this paper, it is worth noting that Iacono's thought experiment appears to be a type known as Backcasting (Robinson, 1982). In Backcasting one imagines a desired or possible state of the world and then reasons backward from that end-state to the necessary precursors. By definition, such logic necessarily does not provide a description of reality, it only provides a chain of precursors that might produce the desired end-state. Such thought experiments, like all thought experiments, are useful in the real world only to the extent that they can be tested and validated or falsified with data. We begin our analysis of the Iacono Thought Experiment (ITE) by defining the hypothetical precursors that he either invented or selected to reach the desired end-state where polygraph tests with chance accuracy could produce a field study with high accuracy rates.

Elucidation and Analysis of the Hypothetical Preconditions and Assumption of the Iacono Thought Experiment

Explicit Assumptions of the Iacono (1991) Thought Experiment.

Iacono (1991) makes a number of explicit assumptions that were used to create a possible path to the desired end state.

1.Eight hundred subjects are tested where 400 are Innocent and 400 are Guilty.

2.The polygraph preforms exactly at chance accuracy of 50% correct, 50% incorrect, and no inconclusive outcomes. This assumption is part of the overall desired end-state where a chance polygraph test could produce high accuracy outcomes. All of the other assumptions also serve the establishment of that end-state.

3.Each crime has only two suspects. (Iacono makes this assumption and states that it is made "without loss of generality" (p. 203).

4.The Guilty suspect is tested first in half of the cases

5.If the first suspect fails the polygraph test, the second suspect will not be tested.

6.Neither innocent nor guilty suspects who pass the test will confess.

7.Only 20% of the Guilty suspects who fail and are interrogated will confess.

Implicit Assumptions of the Iacono Thought Experiment. The following implicit assumptions are also necessary for the mathematics and logic of the Iacono Thought Experiment to reach the desired end-state.

8.The polygraph is the only source of information about who is guilty in a criminal case.

9.Guilty people only confess after polygraph examinations.

Analysis of the Explicit Assumptions of the Iacono Thought

Experiment

Assumption 1 is that the base rate of guilty to innocent subjects is equal. The base rate of guilt in a criminal case will vary greatly depending upon when the polygraph is used. If it is used early in an investigation, there are likely to be far more innocent than guilty sub-

jects; if it is used very late in an investigation, there may be many more guilty than innocent subjects. The assumption of equal base rates is acceptable for a thought experiment as long as one recognizes that variations in the base rate could dramatically alter the end state results and that a base rate of 50% will be unusual in actual practice.

Assumption 2 is that the polygraph performs exactly as a coin flip. This assumption is made as a premise of the thought experiment and it is a necessary component of the desired end state. However, this premise is without empirical support in the real world. To our knowledge, there are no studies that show any version of the CQT to perform at chance levels.

Assumption 3 is that each case has only two suspects. This premise simplifies the mathematics necessary to achieve the desired end state of the Iacono thought experiment, but it is a premise that is rarely met in the real world, and is not at all representative of the field at the time Iacono (1991) was written (for example, Honts & Raskin, 1988, Raskin, Kircher, Honts, & Horowitz, 1988 all contain many single and multiple suspect cases as do the more recent field studies). Iacono's assertion that this assumption is made without a loss of generality (for the Backcasting thought experiment) is clearly not supported by data.

Assumption 4 is that a guilty suspect is tested first in half of the cases. This assumption is tenable only if there are only two suspects and that the examiner has no reason to test one or the other suspect first. It is a convenient assumption for the thought experiment, but it is unlikely to be widely representative of field polygraph testing.

Assumption 5 states that if the first person is tested and fails the second suspect will not be tested. However, this is not the case in real investigations. If the first subject is tested and fails but does not confess, then the remaining suspect or suspects will likely be tested to assess their involvement in the crime. The likely logic of investigators would be that the additional suspect(s) would not be suspects unless there was a reason to suspect them, and they may well be involved. In our experience it is, in fact, common practice to test all suspects in a case during an investigation.

Assumption 6 states that neither innocent suspects nor guilty suspects who pass the test will confess. This is manifestly not true. Under certain circumstances, such as a wrongfully failed or deliberately misrepresented polygraph test result, innocent suspects will confess to crimes they did not commit. The White Paper of the American Psychology Law Society (Kassin, Drizin, Grisso, Gudjonsson, Leo, & Redlich, 2010) specifically notes that wrongfully failed or willfully misrepresented polygraph outcomes are a powerful false evidence ploy that puts the actually innocent at increased risk of false confession. Moreover, Bonpasse (2013), provides examples and discussion of actual cases where incorrect or misrepresented polygraph outcomes have contributed to miscarriages of justice through their role in eliciting false confessions. Assumption 6 is also incorrect for guilty suspects who pass polygraphs, as it ignores the fact that investigations rarely stop just because a polygraph has been passed. If the subsequent investigation continues and additional information is obtained, then the suspect will likely be interviewed a second time, despite the passed polygraph, and may provide a confession then or confess later as part of a plea bargain. At least one such case was included in Honts & Raskin (1988).

Assumption 7 this assumption states that only 20% of the guilty suspects who are interrogated will confess. The choice of 20% is arbitrary and has no empirical basis. The actual confession rate will depend upon the situation in which the tests were conducted. Polygraph tests conducted for defense attorneys, or by the police on subjects who have defense counsel, are unlikely to be followed up with interrogations regardless of the polygraph outcome. On the other hand, the U. S. Department of Defense (2002) has reported data indicating that in one polygraph program, more than 90% of the failed polygraph examinations resulted in relevant admissions. Clearly, the rate chosen for this assumption will have a major impact on the resultant outcomes of the ITE. Moreover, since all confession rate values are situationally specific, it is non-sensical to provide a single value for central tendency as such a value would be meaningless to any

specific applied setting.

Analysis of the Implicit Assumptions of the Iacono Thought Experiment

Assumption 8 asserts that the polygraph is the only source of information about who is guilty in a criminal investigation. This ignores that fact that individuals can confess in contexts other than polygraph examinations, or that other incontrovertible evidence of guilt or innocence may be obtained independent of the polygraph test. This was examined explicitly in Honts (1996) and no differences in numerical scores or accuracy rates were found between confession confirmed and evidence confirmed cases for guilty or innocent subjects. Interestingly, in Honts (1996) none of the innocent subjects were confirmed by a confession obtained in the context of a polygraph examination. Assumption 8 is preposterous on its face, but the ITE cannot work without it.

Assumption 9 is that guilty people only confess following polygraph tests. As covered in our discussion of Assumption 6, we noted that inaccurate and misrepresented polygraph tests can result in false confessions from innocent subjects. We also noted that even guilty subjects who pass polygraph tests will sometimes confess later, when faced with new or overwhelming evidence. Moreover, guilty suspects, and occasionally innocent subjects, will confess as part of a plea bargain. Thus, it is obviously true that guilty people who fail a polygraph test but either are not interrogated, or initially resist an interrogation, may confess later. At least three field studies have explicitly taken this into account and looked for confirming information in an exhaustive sample of cases within a particular period of time, and used all of the information available not only in the polygraph examination file, but in the complete police record of the case (Honts, 1996; Patrick & Iacono, 1991; Raskin et al., 2019).

Summary of the Analysis of Preconditions and Assumptions

Our analysis shows that the assumptions of the Iacono thought experiment were generally chosen without reference to data or professional practice, in the service of developing what became a highly improbable set of preconditions and assumptions leading to a specific solution showing that a polygraph test with chance accuracy could produce a field study with high accuracy rates. The Iacono thought experiment was then transmogrified into a normative statement that all field studies of the CQT were, are, and forever will be unreliable and overestimate actually accuracy. We do not believe that this normative conclusion is justified unless it can withstand empirical examination and falsification. For the remainder of this paper we will refer to the hypothesis derived from the Iacono thought experiment, that the CQT is no more accurate than chance and that all confession criterion studies are biased to dramatically overestimate the accuracy of the CQT as the Iacono Thought Experiment Hypothesis (ITEH).

Data That Could Falsify the Iacono Thought Experiment Hypothesis (ITEH)

Just like Galileo's thought experiment concerning falling objects, the ITEH survives the test of science based upon a lack of falsification data in the scientific research. This leads to the question of what data would falsify the ITEH? The remainder of this paper addresses several sources of converging data that do, in fact, lead to the conclusion that the ITEH is false.

Convergence of Experimental and Field Data Without Detectable Moderator Effects

Recent studies summarized by Hartwig and Bond (2014) have indicated strong convergence between experimental and field studies in psychological science and interpersonal deception detection. Hartwig and Bond explicitly rejected the notion that experiments and field research on interpersonal deception detection produced significantly different results. If the ITEH were true, then polygraph testing would have to be qualitatively different underlying mechanisms from interpersonal deception. Under such circumstances we would expect that laboratory studies of the CQT would produce dramatically lower ac-

curacies than the (according to the ITEH) exaggeratedly high accuracies produced by the supposedly unavoidable effects of the ITEH on field studies of the CQT. Existing reviews simply do not reveal dramatically more accurate results in field than in the laboratory (NRC, 2003; Honts & Thurber, 2019).

Lack of Differences in Accuracy Between Field Studies that Rely on the Confession Criteria and Those That Do Not.

Since the ITEH is critically bound to the use of confessions as a criterion of confirmation of Guilt and Innocence, and the ITEH predicts that the confession criterion critically biases field studies to show high accuracy, we should expect that field studies that included or used other methods of confirmation would produce accuracies that approach chance levels of accuracy. Empirically, this is simply not the case. Honts (1996) directly tested this hypothesis, rating strength of confirmation on a scale that ranged from confessions with the generation of new evidence at one end of the scale to no confirmation at the other. Honts (1996) tested that scale against decision accuracy and against numerical scores. In direct opposition to the predictions of the ITEH, Honts found no effects for the level of confirmation. That is, confession confirmed cases did not have higher accuracy levels than cases that were confirmed by methods other than confession (physical evidence and/or witness statements).

Similarly, there are two field studies that use paired testing and mathematics to estimate accuracy (Ginton, 2013; Mao, Liang, & Hu, 2015). This paired testing approach, while not without problems (Iacono & Ben-Shakhar, 2018), is not dependent upon confessions and so is outside the scope of the ITEH. Estimated accuracy rates from the paired subjects studies converge with data from both laboratory and field studies and thus provide support for both.

Lack of Concurrence Between Wrongful Convictions and Failed CQT Polygraph Tests.

If the ITEH is correct that CQT polygraphs are no more accurate than chance, we would expect that, on average, half of the innocent subjects tested in criminal justice settings would produce false positive errors. In the criminal justice settings, innocent subjects who failed the polygraph would be exposed to interrogation and thus put at risk of false confession. Under such circumstances, we would expect there to be a relatively large number of false positive outcomes among the ranks of the wrongfully convicted. Bonpasse (2013) reviewed the case files of the National Registry of Exonerations[2] which was founded in 1989 as a joint project of the University of Michigan and Northwestern University Law School. Bonpasse reported finding 215 exoneration cases where polygraph tests were involved. Of those 215 cases only 23 (10.7%) contained information that an Innocent subject had been tested before trial and had failed the polygraph. However, there were 44 (20.5%) Innocent subjects who had been tested with the polygraph before trial, produced truthful outcomes, but those favorable outcomes did not help them avoid wrongful conviction. Although the ITEH predicts that false positive errors should be common among the wrongfully convicted, they occurred at only half the rate of true negative outcomes. Bonpasse also reported that across all testing, before and after trial and including tests of the immediate suspect and others (co-defendants and witnesses), 135 (62.9%) of the polygraph test outcome were favorable to the wrongfully convicted person while only 31 (14.4%) produced unfavorable outcomes. Data from the wrongfully convicted strongly contradicts and thus falsifies ITEH.

Discussion

To our knowledge, there is not a single study of the CQT, either laboratory or field, that produced chance accuracy rates. While there is a substantial amount of variability between studies of the CQT, no review has found

[2]http://www.law.umich.edu/special/exoneration/Pages/about.aspx

that laboratory studies are dramatically less accurate than field studies (NRC, 2003; Honts & Thurber, 2019). Thus, the ITEH completely lacks empirical substantiation. Moreover, data from the Honts & Thurber (2019) meta-analysis, field studies that do not use the confession criterion, and the wrongfully convicted all provide evidence that the ITEH is false. The results of the ITE are therefore seen as a failed thought experiment that is completely without empirical support, and which should be relegated to the trash heap of history's failed ideas.

References

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph, 40,* 194-305.

Balakrishnan, J. D. (1999). Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1189–1206.

Bonpasse, M. (2013). Polygraph and 215 wrongful conviction exonerations. *Polygraph, 42,* 112-127.

Brown, J. R. (2014). Thought experiments. *Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/entries/thought-experiment/

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Edition). Hillsdale NJ:Erlbaum.

Cohen, J., (1992). *A power primer. Psychological Bulletin*, 122, 155-159.

DeCoster, J. (2012). Converting effect sizes 2012-06-19-4.xls Retrieved from http://www.stat-help.com/spreadsheets.html

Ginton, A. (2013). A non-standard method for estimating accuracy of lie detection techniques demonstrated on a self-validating set of field polygraph examinations. Psychology, Crime, & Law, 19, 577-594. DOI: http://dx.doi.org/10.1080/1068316X.2012.656118

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology,* 28, 661-676.

Honts, C. R. (1996). Criterion development and validity of the control question test in field application. *The Journal of General Psychology, 123,* 309-324.

Honts, C. R. (2004). The psychophysiological detection of deception, in P. Granhag and L. Strömwall (Eds.) *Detection of deception in forensic contexts.* London: Cambridge University Press 103-123.

Honts, C. R., & Raskin, D. C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16, 56-61.

Honts, C. R., & 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)

Honts, C. R., & Thurber, S. (2019, March). A Comprehensive Meta-Analysis of the Comparison *Question Polygraph Test.* Paper presented at the annual meeting of the American Psychology Law Society, Portland, Oregon.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (2002). The scientific status of research on polygraph techniques: The case for polygraph tests. In, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) Science in the Law: Social and Behavioral Sciences Issue, American Casebook Series (pp. 598-634). West Group: St. Paul, Minnesota.

Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In J. R. Jennings, P.K.

Ackles & M. G. H. Coles (Eds.) *Advances in psychophysiology* (pp. 201-207). London, UK: Jessica Kingsley Publishers.

Iacono, W. G., & Ben-Shakhar, G. (2019). Current Status of forensic lie detection with the comparison question test: An Update of the 2003 National Academy of Sciences report on polygraph testing. *Law and Human Behavior, 43*, 86-98.

Iacono, W. G., & Lykken, D. T. (1997). The scientific status of research on polygraph techniques: The case against polygraph tests. In, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) *Science in the Law: Social and Behavioral Sciences Issue, American Casebook Series* (pp. 582-618-). West Group: St. Paul, Minnesota.

Iacono, W. G., & Lykken, D. T. (2002). The scientific status of research on polygraph techniques: The case against polygraph tests. In, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) *Science in the Law: Social and Behavioral Sciences Issue, American Casebook Series* (pp. 634-688). West Group: St. Paul, Minnesota.

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.

Lykken, D. T. (1978). The psychopath and the lie detector. *Psychophysiology*, 15, 137–142. http://dx.doi.org/10.1111/j.1469-8986.1978.tb01349.x

Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin,* 86, 47–53. http://dx.doi.org/10.1037/0033-2909.86.1.47

Mao, Y., Liang, Y., & Hu, Z. (2015). Accuracy rate of lie-detection in China: Estimate the validity of CQT on field cases. *Physiology & Behavior,* 140, 104-110.

National Research Council (2003). *The Polygraph and Lie Detection.* Washington, DC: The National Academies Press.

Patrick, C. J., & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology, 76,* 229-238.

Piglucci, M. (2006). What is a thought experiment, anyhow? *Philosophy Now: A magazine of Ideas, 58,* Retrieved from

https://philosophynow.org/issues/58/What_is_a_Thought_Experiment_Anyhow

Raskin, D. C. (1978). Scientific assessment of the accuracy of deception of detection: A reply to Lykken. *Psychophysiology, 15,* 143-147.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15,* 126-136.

Raskin, D. C., & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. P*sychological Bulletin, 86,* 54–59. http://dx.doi.org/10.1037/0033-2909.86.1.54

Raskin, D. C., Honts, C. R., & Kircher, J. C. (1997). The scientific status of research on polygraph techniques: The case for polygraph tests. Chapter in, D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.) Modern scientific evidence: The law and science of expert testimony (pp. 565-582).

Raskin, D. C., Honts, C. R., & Kircher, J. C. (2014). *Credibility assessment: Scientific research and applications.* Oxford, UK: Academic Press. ISBN: 978-0-12-394433-7 (ebook version available online 17 December 2013).

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (2019). A study of the validity of polygraph examinations in criminal investigations. Final report to the National Institute of Justice, Grant Number 85-IJ-CX-0400. *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice, 48,* 10-39.

Rice, M. E. and Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior,* 29, 615-620.

Robinson, J. B. (1982) Energy backcasting: A proposed method of policy analysis. *Energy Policy*, December.

Tape, T. G. (2019). Interpreting Diagnostic Tests. University of Nebraska Medical Center. Retrieved from, http://gim.unmc.edu/dxtests/Default.htm also referencing subordinate web pages on the same topic at this URL.

U. S. Department of Defense (2002). *Department of Defense Polygraph Program Annual Report to Congress, Fiscal Year 2002.* Office of the Assistant Secretary of Defense (Command, Control, Communications, and Intelligence.

Vrij, A. (2008). Detecting Lies and Deceit: Pitfalls and Opportunities, Second Edition. Chichester, UK: Wiley.

Zhang, X. (2011). The evolution of polygraph testing in the People's Republic of China. *Polygraph,* 40, 181-193.