

Polygraph

VOLUME 37

2008

NUMBER 2

Contents

A Focused Polygraph Technique for PCSOT and Law Enforcement Screening Programs Mark Handler, Raymond Nelson, & Ben Blalock	100
Optimal Decision Rules for Evaluating Psychophysiological Detection of Deception Data: An Exploration Stuart M. Senter & Andrew B. Dollins	112
Degrees of Deception: Diploma Mills and the Polygraph Examiner Community – A Recommendation for Change Timothy J. Weber & Frank Horvath	125
Improving the Detection of Physical Countermeasures with Chair Sensors Jack Ogilvie & Donnie W. Dutton	136
Exploration of a Two-Stage Approach Stuart M. Senter & Andrew B. Dollins	149
An Introduction to the APA's Panel on International Developments in Polygraphy Frank Horvath	165
The Use of the Polygraph in Lithuania Anonymous	166

A Focused Polygraph Technique for PCSOT and Law Enforcement Screening Programs

Mark Handler, Raymond Nelson and Ben Blalock

The real voyage of discovery consists not in seeking new lands, but in seeing with new eyes.
- Marcel Proust, French novelist

Abstract

Testing techniques most commonly used in polygraph screening programs were adapted from protocols originally developed for event-specific investigative polygraph testing, including the examination structures and decision rules. Screening examinations are being increasingly recognized for providing a unique and powerful tool for decision-makers and with the widening demand for polygraph screening services there is a commensurate obligation for polygraph professionals to give attention to oft-neglected questions regarding the validity and reliability of the methods they employ. In that vein, the authors propose a focused approach for polygraph screening, derived from a validated polygraph screening technique developed at the Department of Defense Polygraph Institute (now the Defense Academy for Credibility Assessment). In addition, we suggest selecting investigation targets that are informed by risk prediction and risk management research, and are consistent with our present understanding of the psychological and physiological mechanisms upon which the polygraph technique depends. An example of this approach is provided.

Background

Screening polygraph examinations are those conducted where there is an absence of a known event or known allegation. Polygraph screening has been used since as early as the 1930's when Leonarde Keeler signed an agreement with the insurance firm Lloyds of London to periodically test bank employees for embezzlement (Alder, 2007). Krapohl and Stern (2003), however, provided an overview of the challenges inherent in screening polygraph programs in their discussion of the "successive hurdles" approach (Meehl & Rosen, 1955). Research by Barland, Honts, and Barger (1989) and Honts (1992) revealed potential inadequacies existed in polygraph screening methods employed at the time.

One early screening test, the Counterintelligence Screening Test (CIST) was developed in about 1971 by US Army military intelligence examiners using directed-lie comparison (DLC) questions (Barland, 1981). DLC questions are those which the examiner instructs the examinee to answer falsely (Honts & Raskin, 1988; Raskin & Honts, 2002). Studies using DLC techniques (DoDPI Research Division Staff, 1997; Research Division Staff, 1998) suggested that a DLC approach and other improvements in test administration structure and decision policies contributed significantly to polygraph testing program objectives of sensitivity to deception and specificity to truthfulness.

There are undoubtedly fewer field and laboratory studies that address validity of the

Authors' Note

The authors are grateful to Paul Menges, Don Krapohl, Dale Austin, Donnie Dutton, George and Paula Baronowski, Dr. Bonnie Harrison, Dr. Stephen Harrison, Dr. Charles Honts, Jerry Thomas, Dr. Stuart Senter and Dr. Tim Weber for their thoughtful reviews and comments to drafts of this paper. The views expressed in this article are solely those of the authors, and do not necessarily represent those of the Department of Defense, American Association of Police Polygraphists, or the American Polygraph Association. Questions and comments are welcome at polygraphmark@gmail.com.

DLC than the PLC. However, the results of existing studies (Barland, 1981; Barland et al., 1989; DoDPI Research Division Staff, 1997; DoDPI Research Division Staff, 1998; Honts & Raskin, 1988; Horowitz, Kircher, Honts & Raskin, 1997; Kircher, Packard, Bell & Bernhardt, 2001; Reed, 1994; Raskin & Kircher 1990) have shown the DLC to perform as well as the probable-lie comparison (PLC) questions. DLCs require less complex administration practices than those associated with the PLC approach and offer greater potential for standardization.

The Research Staff at DoDPI undertook an effort to address the perceived inadequacies of the currently used screening tests and eventually created the Test for Espionage and Sabotage (TES). The design specifications of their improved screening technique included the standardization of the pretest portion of the examination, as well as standardization and reduction of investigation targets to two primary issues. The two target issues are usually presented three times each in a single examination chart. Test protocols allow for the inclusion of additional investigative targets in a separate series of questions, again conducted within a single examination. Variability in test administration is reduced through the presentation of each test question in a standardized sequence. The testing protocol includes a standardized acquaintance test, a standard rationale and explanation of the DLC questions, a standard explanation of instrumentation and psychophysiological responses and a standardized in-test chart presentation. Decision policies require that the examinee is regarded as responding significantly to the examination as a whole, rather than to individual questions, if the observed responses are significant or consistent with those expected from deceptive persons. The National Research Council (2003) reported the accuracy index (A) of the improved screening test to be 0.90.

One relevant study (DoDPI Research Division Staff, 1997) compared the TES to the CSP using PLCs and the CSP using DLCs. There was no significant difference in the overall accuracies in identifying programmed innocent participants; 89%, 95% and 95% for the TES, CSP-PLC and CSP-DLC, respectively. However, for *programmed guilty* participants,

the TES format outperformed both versions of the CSP; 83% for the TES versus 56% for CSP-PLC and 59% for CSP-DLC. These accuracy estimates are given excluding inconclusive rates which were 21%, 23% and 20% for the TES, CSP-PLC and CSP-DLC, respectively.

The DoDPI Research Division Staff (1998) conducted a second study as a formal replication of the first study using data from the first TES study to evaluate new scoring criteria in an effort to improve upon the technique. In this second study innocent examinees were identified with 98% accuracy and guilty examinees with 83% accuracy. The researchers reported an initial average inconclusive rate of about 15% but this was later reduced to an inconclusive rate of 2% for the innocent and 0% for the guilty after continued testing to resolve inconclusive tests.

Reed described a third TES study (Reed, 1994) addressed examiner subjective opinion bias, an expanded comparison question list, question formatting and wording, and a "team approach" to the administration of the TES. The "team approach" portion of the study explored an approach where one examiner administered half of the examinations including the pre-test, in-test, and data analysis phases. The remaining half of the examinations was performed by two examiners; one conducted the pre-test and one conducted the in-test. Both examiner-members of the team evaluated the test data individually. The overall combined accuracies (excluding inconclusive results) were 85% for the innocent, 78 % for the guilty with an average initial inconclusive rate of about 13%. In summary the three studies indicated that the TES could produce accuracy rates that were significantly above chance levels.

Standardization of any technique can serve to increase inter-rater and test-retest reliability and both measures constrain the potential validity of a technique. Excessive variability in test administration or interpretation will necessarily compromise the reliability and validity of any test method. Inter-rater reliability is a concern that will remain of paramount importance to questions about polygraph validity. When standardized practices are based on principles that are

consistent with validated constructs and data obtained through the objective study of data, we can more reasonably anticipate that improvements will contribute meaningfully to the test design goal of criterion validity.

Test Description

Following existing practices we defined a screening technique we call the Directed-Lie Screening Test (DLST) that contains two neutral questions (N1, N2), a sacrifice relevant question (SR), two separate relevant questions (R1, R2) and two comparison questions (C1, C2).

The sequence is as follows:

- N1- Neutral or Irrelevant question
- N2- Neutral or Irrelevant question
- SR- Sacrifice Relevant question
- 1C1- First presentation of DLC#1
- 1R1-First presentation of R1
- 1R2-First presentation of R2
- 1C2-First presentation of DLC#2
- 2R1-Second presentation of R1
- 2R2-Second presentation of R2
- 2C1- Second presentation of DLC#1
- 3R1-Third presentation of R1
- 3R2-Third presentation of R2
- 2C2-Second presentation of DLC#2

Presentation of the question sequence is intended to be standardized except when it is necessary to present an additional neutral question before proceeding with the next test question. Additional presentations are allowed when three artifact-free presentations of each have not been obtained. In the latter case, examiners are permitted to present the question sequence a fourth time. This can take place as a fourth presentation of the test stimuli within the single examination chart or through the completion of a second shorter chart, consisting of the following sequence (N1, N2, SR, 3C1, 4R1, 4R2, 3C2).

Test Data Analysis and Decision Criteria

The test data are hand-scored with validated scoring criteria by comparing the relevant question response to the stronger response of an adjacent comparison question per each component sensor. In consideration of the cautions expressed by Bell, Raskin, Honts, and Kircher (1999) regarding artifacted

or uninterpretable data, examiners should be careful to assign scores only to pneumograph data of arguable authentic quality and interpretive value. One cautionary issue exists in scoring DLC exams. Kircher and Raskin (2002) and Kircher et al. (2001) have reported that the data collected from pneumographs in DLC examinations do not appear to have diagnostic value.

Test data analysis can be automated by dividing the single examination chart of three presentations of each test stimulus into three virtual charts, using the following sequences:

- Chart 1: (1C1, 1R1, 1R2, 1C2)
- Chart 2: (1C2, 2R1, 2R2, 2C1)
- Chart 3: (2C1, 3R1, 3R2, 2C2)

If a fourth presentation of the test stimulus is completed, the sequence will be (2C2, 4R1, 4R2, 3C1) or (3C1, 4R1, 4R2, 3C2) depending on whether the fourth presentation of the stimuli was completed as part of the single examination chart sequence or as a separate short examination chart respectively.

Hand-scored results for each relevant question are totaled along with the grand total for the examination as a whole. A spot total of -3 or lower at either spot, or a grand total of -4 or less results in an opinion of Significant Response (SR). No Significant Response (NSR) opinions are the result of a grand total of +4 across the two relevant targets, as long as there is a positive numerical subtotal for each target. If the result is neither SR nor NSR it is Inconclusive or a No Opinion (NO) can be rendered. Many examiners will recognize these cutting scores as identical to those for the "You-Phase" two-question Zone Comparison Technique (Department of Defense, 2006).

These rules differ from the common spot scoring rules for MGQT examinations (Ansley, 1998; Department of Defense, 2006), which require a +3 or greater at every relevant question. Existing practices are based on the belief that each question is related to a separate issue and therefore should be treated separately. However, none of the existing cut scores for spot scoring decisions has been subject to statistical analysis and examiners cannot presently calculate a p-value for the

significance of hand-scored results. Empirical studies of spot scoring practices suggest that present values may not be optimal (Capps & Ansley, 1992). Research suggests when an SR decision is rendered, the strongest physiological responses are not always to the question to which the examinee is being deceptive (Barland, 1981; Barland et al., 1989; Correa & Adams, 1981; Kircher, Raskin, Honts & Horowitz, 1988; DoDPI Research Division Staff, 1998). In general, accuracy tends to decrease when examiner opinions are made on a per-question basis. An examinee may be practicing deception to one relevant question on the test and have more arousal to another relevant question on the test. It is clear the existing polygraph methods can alert an examiner when an examinee is practicing deception. Data do not yet support the notion that existing polygraph screening methods can advise an examiner regarding an exact test question to which an examinee is practicing deception.

Several studies of polygraph scoring (Krapohl, 2005; Krapohl and Cushman, 2006; Senter, 2003) have shown that two-stage scoring rules maximize decision accuracy by using spot scores to resolve inconclusive results. Mathematical expectations that spot scoring rules may inflate false positive and inconclusive results are supported by Nelson, Handler and Krapohl (2007), who found that alternative decision policies, based on statistical theory, can help to optimize the specificity and sensitivity of screening examinations. The "test as a whole" decision rule applies when assessing for NSR results. Nelson et al., (2007) used a Kruskal-Wallis equation, as a one-way analysis of variance to evaluate differences between different investigation targets before rendering an NSR result. This process procedurally approximates the requirement for a positive sign value for all spots when hand-scoring the DLC screening exams according to procedures described by Department of Defense (2006). Another consideration for empirical inquiry involves the potential advantages of sequencing decision rules in various ways. Senter (2003) found no significant differences in sequencing of decision rules in hand-scoring experiments. Nelson et al., (2007) achieved optimal balance of sensitivity and specificity by sequencing decision rules that parse NSR results ahead of decision rules for

SR results with event specific investigative polygraphs involving ZCT and MGQT techniques. They were able to maximize sensitivity to deception in screening exams by executing rules for SR classifications before those for NSR results.

Target Selection

While reviewing exact details for each investigation question will always remain a task for the examiner and examinee at the time of the examination, the selection of investigation targets is an important consideration prior to the examination. It would be a simplistic and naive assumption to suggest that polygraph examiners themselves know what questions or targets to investigate on behavior of an investigation or risk assessment process. In investigative polygraph programs, examination targets are specified by the details of a crime or investigation. Investigation targets in polygraph screening programs are properly informed by data from risk prediction and risk management research.

Post Convicted Sex Offender Testing (PCSOT) polygraph monitoring programs should emphasize behaviors that provide supervision and treatment professionals with early warning of an escalating risk level, and allow for corrective intervention prior to a new assault. Possible behavioral indicators include the unauthorized use of pornography, unauthorized physical contacts with children or being alone or unsupervised with minors, masturbation behaviors involving fantasies of children or violence, and secretive or undisclosed sexual partners. Other investigation targets may address concerns about use of alcohol or illegal drugs while under supervision. By emphasizing investigation targets pertaining to safety and compliance behavior that is a precursor to re-offense activities, supervision and treatment professionals will avert the costs to individuals, families, and communities associated with new sexual assaults. Polygraph questions regarding noncompliance with supervision and treatment will also not create secondary problems involving offenders' rights against self-incrimination regarding new crimes. Additionally, noncompliance behaviors might be expected to occur at higher base-rates than re-offense activities,

which serve to simplify any error estimation methods based on Bayesian models.

PCSOT risk assessment polygraph programs will also be interested in investigation targets pertaining to unknown or unreported sexual offense history behaviors that have a direct role in actuarially derived risk-prediction schemes. Polygraph has been shown to increase the amount of useful disclosure (Ahlmeyer, Heil, McKee & English, 2000; Grubin et al., 2004) as well as deterring unwanted or risky behavior (Kokish, Levenson & Blasingame, 2005) among sex offenders. Target behaviors for sexual history disclosure polygraphs may include an offender's history of incest activities or sexual contact with relatives, adult sexual contact with underage persons (e.g., minors under age 15 in Colorado and other states, or children four or more years younger than juvenile offenders). Risk assessment targets involving historic victim access behaviors may include questions about forced or violent sexual assault, including implied or threatened violence. Additional targets of interest may be questions about sexual contact with persons who were unconscious from alcohol or illegal drugs, or while sleeping. Risk assessment targets involving historic sexual compulsivity behaviors, may include voyeurism (sexual peeping), exhibitionism (indecent exposure), frottage (unwanted rubbing or touching of strangers in public), theft of underwear or undergarments, or public masturbation activities.

Countermeasures

No consideration of a new approach would be complete without the discussion of countermeasures. Countermeasures have become a highly discussed topic among polygraph professionals. One can hardly attend a national conference and not expect to be afforded an opportunity to attend a lecture that includes a discussion on countermeasures. Several well designed scientific studies have assessed the vulnerability of polygraph to countermeasures (for a thorough discussion see Honts & Amato, 2002). The findings of most polygraph countermeasure studies suggest that under very specific conditions, countermeasures can reduce sensitivity to deception. These findings suggest that effective countermeasure training

must include; educating examinees on testing procedures; teaching them to evoke physiological arousal through physical movements and/or mental arousal, and coaching them by attaching them to a polygraph to practice. Absent the last key element, most published research suggests countermeasures will be ineffective at producing negative polygraph outcomes. One can only hope that sex offender access to a polygraph instrument and an examiner willing to train them is very limited.

Some research suggests the use of countermeasures by examinees is actually counterproductive. Innocent and programmed guilty examinees who engage in countermeasures have been found to produce polygraph test data more indicative of lying (Honts & Alloway, 2007). This was consistent with the finding that innocent examinees engaging in spontaneous countermeasures are more likely to fail a test and the guilty that so engage enjoy no benefit as a result of their attempts (Honts, Amato & Gordon, 2001).

Some may argue a DLC approach is an invitation to employ countermeasures. However, a review of *The Lie Behind the Lie Detector* (Maschke & Scalabrini, 2007) finds PLC testing is addressed with equal (if not greater) depth than DLC testing. DLC polygraph formats task the examinee with simply saying "no" to a personally significant question when they and the examiner both know the answer is not true. During the question review portion of the pre-test interview, the examinee is encouraged to recall a minor past transgression unrelated to the issue(s) at hand. The examinee is not instructed to recall this transgression while answering the DLC during the test data collection. They are instructed to answer all test questions in an equally timely manner. The DLC acquires salience from the task demands, not from the recall effort. We know of no research suggesting that examinees use different countermeasure strategies depending on whether they are targeting PLCs or DLCs. In other words, whatever countermeasures examinees would use against DLCs they would also use against PLCs. Perceptions that DLCs are more vulnerable to countermeasures than are PLCs are not supported in the published literature. As a practical matter, examiners unable to detect

or deter countermeasures with DLCs would probably not fare better if they used PLCs.

One Example of the Application of the DLST

One examiner who tests subjects for a probation department agreed to conduct some tests using a DLST and graciously share those data with us. The examiner was conducting a periodic maintenance polygraph examination on an offender on probation for a sexual offense. Maintenance polygraphs target non-compliance behaviors that reveal the early onset of an escalating risk level. This offender's treatment provider requested the target areas include viewing pornographic material, being alone or unsupervised with anyone under age ten, and the use of alcohol or illegal drugs. This offender's last polygraph test was a sexual history disclosure and had taken place about six months prior to this exam.

The examiner formulated and reviewed the following relevant questions for the first chart:

R1 Since your last polygraph test, did you drink any alcoholic beverages?

R2 Since your last polygraph test, did you use any illegal drugs?

The examiner conducted the first sub-test which included three iterations of these two relevant questions. After hand-scoring the examination using the federal 7-position numerical evaluation scoring system (Department of Defense, 2006) the examiner rendered an opinion of *No Significant Response*. The test data were evaluated by the OSS3 algorithm (Nelson *et. al.*, 2007), excluding the respiration channel. OSS3 reported a probability that the data were produced by a deceptive person was 0.042 or approximately 4%. (It should be noted, however, the OSS3 tool has not yet been validated with DLC testing.) The chart one is presented below in three sections.

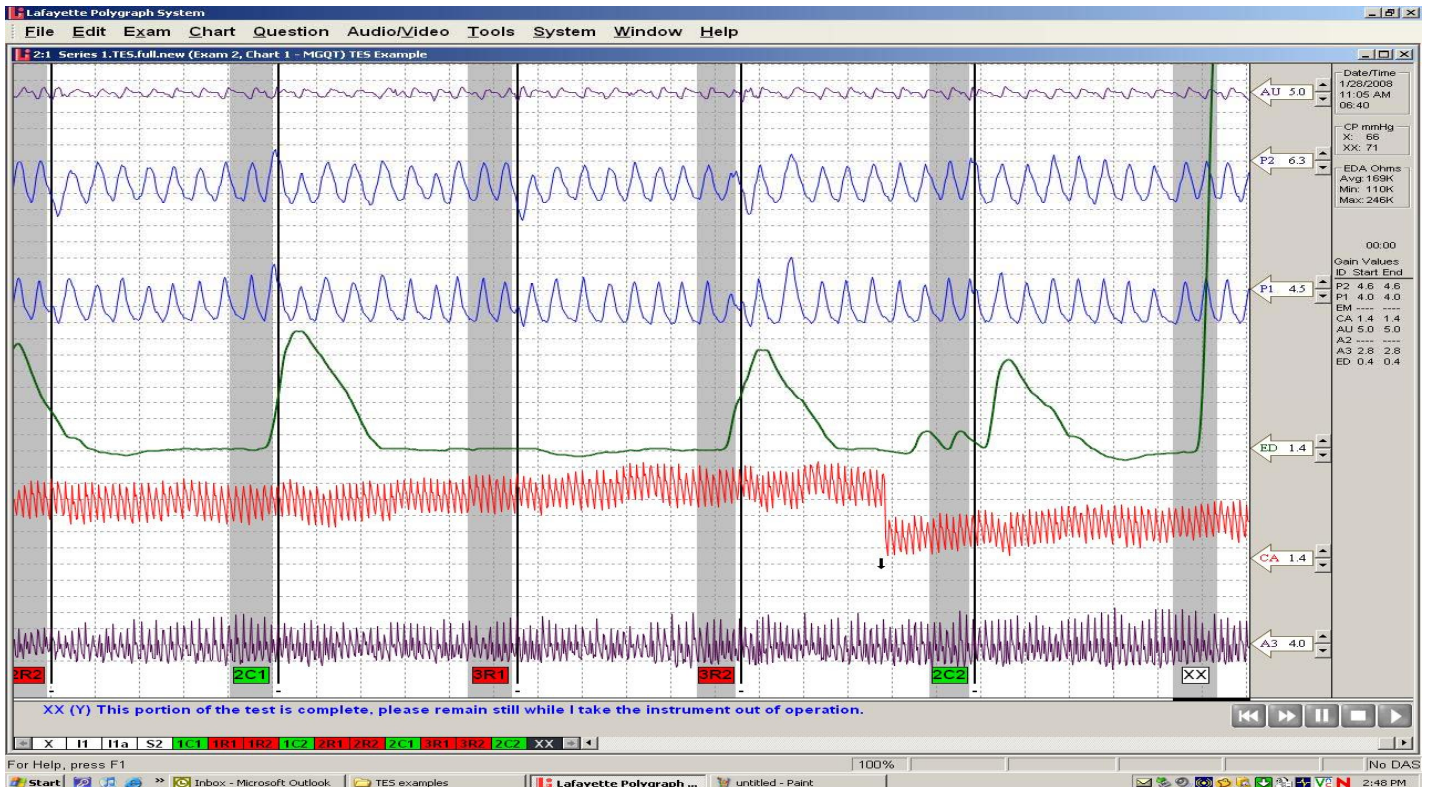
Chart one, section one of three.



Chart one, section two of three.



Chart one, section three of three.



Next the examiner formulated and reviewed the following relevant questions for the second chart:

R1 Since your last polygraph test, other than that one time, have you viewed any pornography?

R2 Since your last polygraph test, have you been alone or unsupervised with anyone under the age of ten?

The examiner conducted the first sub-test which included three iterations of these

two relevant questions. After hand-scoring the examination using the federal 7-position numerical evaluation scoring system (Department of Defense, 2006) the examiner rendered an opinion of *Significant Response*. The test data were evaluated by the OSS3 algorithm, excluding the respiration channel. OSS3 reported a probability that the data were produced by a truthful person was 0.020 or 2%. (Again we remind readers the OSS3 tool has yet to be validated with DLC testing.) The chart two is presented below in three sections.

Chart two, section one of three.



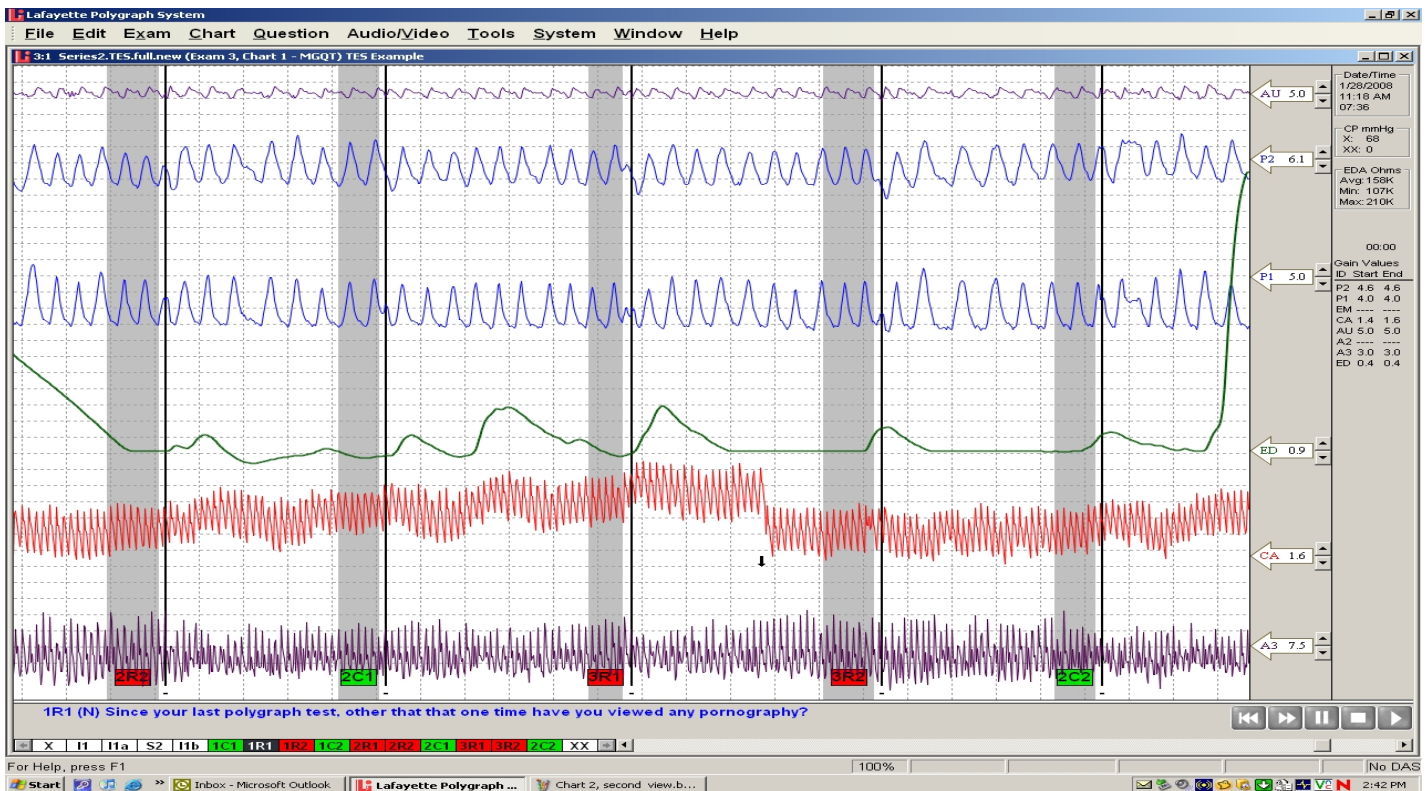
During a post-test interview, the examinee admitted to multiple viewings of pornography. He told the examiner he had downloaded pornography from an I-Pod music playing device onto his own handheld device

capable of viewing video. The examinee admitted the infraction to his probation officer and was expected to be confronted with the issue during his next group therapy session.

Chart two, section two of three.



Chart two, section three of three.



Summary

We propose here one alternative screening polygraph method. We suggest it may prove to work well for PCSOT for offenders who are tested regularly. It is a modification of a well-researched technique, the TES, which in the laboratory has been shown to be effective as the initial method for polygraph screening in the counterintelligence realm. Combined with the “successive hurdles” approach (Krapohl & Stern, 2003; Meehl & Rosen, 1955) it can be a powerful tool to assist treatment providers and supervisory officials in the treatment and containment of sex offenders. Though we focused our discussion primarily around PCSOT, we feel this approach may also result in increased sensitivity and specificity in other polygraph screening milieus.

It might be argued by those skeptical to this approach that DLST has never been researched in the PCSOT setting, certainly a legitimate observation. The only generalization of the validity of DLST is the replicated research on the TES, the method after which DLST is modeled. It is important to note that the only difference between DLST and TES are the test questions. To our knowledge polygraph techniques are not designed for only one type of test question (otherwise, we might have to have as many techniques as there are crimes to investigate). For this reason we are confident that DLST can be used effectively in the PCSOT setting. To those who would persist that DLST has not been validated for PCSOT, we would simply point out the obvious: such a standard would

eliminate all other polygraph techniques, as well.

There are certain caveats that attend the use of DLST. First, examiners with no familiarity with DLCs should receive formal instruction in their proper development and introduction. Second, the pneumograph for any DLC technique cannot be analyzed using the same criteria as are used for PLC testing. Consequently, scoring rules must be adjusted, and there are currently no algorithms available that have been trained on DLST data. Third, DLST is a one-chart test, and can only accommodate two relevant questions per series. In using the DLST examiners depart from the more familiar PLC techniques which can accommodate larger numbers of questions but demand larger numbers of charts. Finally, like the TES, DLST may not be a standalone technique but may be only the first step in a successive hurdles approach.

Advantages to DLST are obvious. DLCs do not require the type of manipulation seen necessary for PLCs, resulting in better time management in the examination room. DLCs have the added benefit of remaining useful over repeated examinations. This is important given that many offenders are tested every few months, a circumstance which poses a challenge to examiners to keep PLCs salient over time. Lastly, DLCs reduce the intrusiveness of polygraph testing over PLC methods, thereby eliminating one source of complaints against the examiner or the examination. The DLST may serve as a primary or secondary screening technique, in both general and unique screening cases.

References

- Ahlmeyer, S., Heil, P., McKee, B. & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sex Abuse*, 12, 123-138.
- Alder, K. (2007). *The History of an American Obsession- The Lie Detectors*, (p. 124), New York, NY, Free Press.
- Ansley, N. (1998). Question Formulation. *Polygraph*, 27(3), 181-187.
- Barland, G. H. (1981). *A validation and reliability study of counterintelligence screening test*. Unpublished manuscript, Security Support Battalion, 902nd Military Intelligence Group, Fort Meade, Maryland.
- Barland, G. H., Honts, C. R. & Barger, S.D. (1989). Studies of the accuracy of security screening polygraph examinations. *DTIC AD Number A304654*.
- Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.
- Capps, M. H. & Ansley, N. (1992). Analysis of federal polygraph charts by spot and chart total. *Polygraph*, 21, 110-131.
- Correa, E. J. & Adams, H.E. (1981). The validity of the pre-employment polygraph examination and the effects of motivation. *Polygraph*, 10, 143-155.
- Department of Defense. (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retrieved on 03/13/2008 from <http://antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf>
- DoDPI Research Division Staff, (1997). A Comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats. *Polygraph*, 26, 79-106.
- DoDPI Research Division Staff. (1998). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. *Polygraph*, 27, 68-73.
- Grubin, D., Madsen, L., Parsons, S., Sosnowski, D. & Warberg, B. (2004). A prospective study of the impact of polygraphy on high-risk behaviors in adult sex offenders. *Sex Abuse*, 16, 209-222.
- Honts, C. & Raskin, D. (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. (1992). Counterintelligence scope polygraph (CSP) test found to be poor discriminator. *Forensic Reports*, 5, 215-218.
- Honts, C. R. & Amato, S.L. (2002). Countermeasures. In Murray Kleiner (Ed.), *Handbook of Polygraph Testing*. (pp. 251-264). Academic Press.
- Honts, C., Amato, S. & Gordon, A. (2001). Effects of spontaneous countermeasures used against the comparison question test. *Polygraph*, 30(1), 1-9.
- Honts, Charles, R. & Alloway, Wendy , R. (2007). Information does not affect the validity of the comparison question test. *Legal and Criminal Psychology*, 12, 311-320.

- 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.
- Kircher, J. C. & Raskin, D.R. (2002). Computer Methods for the Psychophysiological Detection of Deception. In Murray Kleiner (Ed.), *Handbook of Polygraph Testing*. (pp. 287-326). : Academic Press.
- Kircher, J. C., Packard, T., Bell, B. G. & Bernhardt, P.C. (2001,). *Effects of prior demonstration of polygraph accuracies on outcomes of probable-lie and directed-lie polygraph tests*. DoDPIO2-R-0002.
- Kircher, J. C., Raskin, D. C., Honts, C. R. & Horowitz, S.W. (1988). Generalizability of mock crime laboratory studies of the control question polygraph technique. *Psychophysiology*, 25, 462-463.
- Kokish, R., Levenson, J. S. & Blasingame, G.D. (2005). Post-conviction sex offender polygraph examination: client-reported perceptions of utility and accuracy. *Sex Abuse*, 17, 211-221.
- Krapohl, D. J. (2005). Polygraph Decision Rules for Evidentiary and Paired Testing (Marin Protocol Applications). *Polygraph*, 34, 184-192.
- Krapohl, D. J. & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph*, 35(1), 55-63.
- Krapohl, D. J. & Stern, B., A. (2003). Principles of Multiple-Issue polygraph screening: A model for applicant, post-conviction offender, and counterintelligence testing. *Polygraph*, 32, 201-210.
- Maschke, G. W. & Scalabrini, G.J. (2007). *The Lie Behind the Lie Detector*, 4th Edition. Internet: AntiPolygraph.org, Retrieved on 03/13/2008 from <http://antipolygraph.org/lie-behind-the-lie-detector.pdf>
- Meehl, P. E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychology Bulletin*, 52(3), 194-216.
- National Research Council. (2003). *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press.
- Nelson, R., Handler, M. & Krapohl, D. (2007). *Development and Validation of the Objective Scoring System, version 3*, Poster presentation at the 2007 annual meeting of the American Polygraph Association in New Orleans, LA.
- Raskin, D. C., & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic (1-49).
- Raskin, D.C. and Kircher, J.C. (1990). Development of a Computerized Polygraph System and Physiological Measures for Detection of Deception and Countermeasures: A Pilot Study. A preliminary report under contract 88-L655300-000, Scientific Assessment Technologies, Inc. Salt Lake City, Utah.
- Reed, S. (1994). Briefing on the Test of Espionage and Sabotage (TES), Expansion and Validation Study: Retrieved on 03/13/2008 from <http://antipolygraph.org/documents/tes-expansion-validation.shtml>
- Senter, S. M. (2003). Modified general question test decision rule exploration. *Polygraph*, 32, 251-263.

Optimal Decision Rules for Evaluating Psychophysiological Detection of Deception Data: An Exploration

Stuart M. Senter and Andrew B. Dollins

Abstract

This project was undertaken to improve decision accuracy for Zone Comparison Test format polygraph examinations. The data were collected during two previous studies. Decisions were rendered using eight different decision rules. The decision rules varied in the number of question series used in the decision (three, five, or 'three or five'), and the type of decision rule used (total cutoff rule, spot score rule, or both). Across both data sets, approaches that used 'three or five' question series did better (79% correct decisions) than those which used only three question series or all five question series (71% correct decisions). The best results across both data sets were produced by decision rules that implemented 'three or five' question series and used both the total cutoff and spot score rules (81% correct decisions). Other decision rules may, however, be preferable depending on the parameters of the tested population. Thus, strong variations in polygraph accuracy were produced as a function of decision rule.

Introduction

Previous studies have shown that decision accuracy for laboratory based psychophysiological detection of deception examinations is mitigated by the rules used to derive decisions (Podlesny & Truslow, 1993; Senter & Dollins, 2004; Senter, Dollins, & Krapohl, 2004). No differences in decision accuracy were found as a function of principles used to assign relevance to physiological responses (Bell, Raskin, Honts, & Kircher, 1999; Swinford, 1999). Most of the variance in decision accuracy was a function of two factors. The most influential of the two is the question series (a set of approximately 10 questions) rule, which addresses whether data from three or five question series will be evaluated when making a decision. The second is the decision approach regarding cutoff criteria; whether to use a spot score or a total cutoff rule (these are explained in detail below). The objective of this archival study is to evaluate the efficiency of psychophysiological detection of deception decision rules designed using variations and combinations of question series usage and decision rules.

Comparison question psychophysiological detection of deception examinations are composed primarily of relevant and comparison questions (Federal Psychophysiological Detection of Deception Examiner Handbook, 1999). Relevant questions are those which pertain to the examinee's involvement in the issue in question (e.g., Did you steal any of that money?). Comparison questions are related to the issue, but are separated from that issue, typically using time bars (e.g., Prior to this year, did you ever steal anything?). In theory, deceptive examinees produce larger responses to the relevant questions, and truthful examinees produce larger responses to the comparison questions. Each relevant question is paired with one or more comparison questions for evaluative purposes. The greatest response elicited by a comparison question is compared to the response elicited by a relevant question. Positive values are assigned when the responses to the comparison question are greater than those to the relevant question, and negative values are assigned when responses to the relevant question are greater. Scores are assigned to each data channel for each relevant-comparison pair on each question series. The

Dr. Senter and Dr. Dollins are research psychologists at the Defense Academy for Credibility Assessment. Comments and questions can be sent to Senters@daca.mil. The conclusions expressed in this paper do not necessarily represent the views of the U.S. Government or the American Polygraph Association.

Zone Comparison Test format includes three of these relevant-comparison question pairs. Assigned scores may be totaled in different ways to produce decisions, as described below.

Senter and Dollins (2004), and Senter et al. (2004) examined the impact of using different decision rules after scores had been assigned to relevant-comparison question pairs. Both of these previous works used physiological data from laboratory mock crime polygraph studies (Department of Defense Polygraph Institute Staff, 2001 and Kircher & Raskin, 1988, respectively). These studies required deceptive participants to steal items from a secretary's desk, while truthful participants did not complete a crime. In each study, physiological data were recorded as five question series were presented to each participant. Scorers assigned values to each of three relevant-comparison question pairs for each data channel (respiration, cardiovascular, and electrodermal), and for each question series. The data used in the Senter et al. also included scores assigned to data collected using a photo-plethysmograph, which provides a measure of peripheral blood pressure. Scores assigned to this channel were not included in the current study to maintain consistency with the data reported by Senter and Dollins, which did not include the photo-plethysmograph channel. The decision accuracy obtained using two rules currently implemented in laboratory studies was also compared during these previous studies.

The first factor examined, the question series rule, addresses the number of question series presented and evaluated when making veracity decisions. The Defense Academy for Credibility Assessment supports a three question series rule, which is that data recorded during three question series should be evaluated when making veracity decisions (Defense Academy for Credibility Assessment, 2008). Data from a fourth question series may be evaluated if one of the earlier series contains artifacts which preclude evaluation of an earlier response. Investigators at the University of Utah support a 'three or five' question series rule, which is that data are recorded during five question series (at least in laboratory studies). If a veracity decision cannot be made using the first three question

series, then data from all five question series are used to make a decision. These researchers have found this procedure to be effective, producing high levels of accuracy (Honts & Raskin, 1988; Honts, Raskin, & Kircher, 1994; Horowitz, Kircher, Honts, & Raskin, 1997; Kircher & Raskin, 1988).

The second rule examined, the decision rule, addresses numerical decision criteria. The spot score rule is one of the decision rules examined. The spot score process involves adding the assigned scores for each relevant-comparison question pair across each data channel and across each question series. This produces a spot score for each relevant question. A decision of truthful is made if the three spot scores sum to +6 or more and if each spot score is +1 or greater. A decision of deceptive is made if the three spot scores sum to -6 or less or if any spot score is -3 or less. A no opinion decision is made if a truthful or a deceptive decision is not made (Department of Defense Polygraph Institute, 1992). The total cutoff rule (Bell, et. al., 1999) produces a truthful decision if all assigned scores sum to +6 or more and a deceptive decision if all assigned scores sum to -6 or less. A no opinion decision is produced if the total falls between -5 and +5, inclusive.

We found significant differences in the number of correct decisions obtained when different question series rules were used (Senter & Dollins, 2004; Senter et al. 2004). More correct decisions were produced when the 'three or five' question series rule was used than when the three question series rule was used. Senter and Dollins (2004), and Senter et al. (2004) found an average increase of 8% (69% vs. 77%) and 12% (67% vs. 79%), respectively, in the number of correct decisions produced using the 'three or five' question series rule, relative to using the three question series rule.

The impact of using the total cutoff rule versus the spot score rule was also explored (Senter & Dollins, 2004; Senter et al. 2004). Negligible differences were produced (73% vs. 73%, and 72% vs. 74%, respectively) in the frequency of correct decisions. However, both studies did show asymmetries across participant veracity in the frequency of correct decisions. Combining data from the two studies, more correct decisions were produced

for deceptive participants using the spot score rule (83% vs. 69%) and more correct decisions were produced for truthful participants using the total cutoff rule (62% vs. 78%).

In the present study, we examined decision approaches and question series rule combinations that are not currently taught or, to our knowledge, used. This was done to determine optimal combinations of decision approaches and question series rules, or what we will refer to as decision rules. We selected this finite combination of decision rules based on the existing real-world use of their component factors. That is, each given factor of each decision rule is used in real-world polygraph applications, though many of the resulting combinations were novel in nature. We defined an optimal decision rule as one that produces an overall high number of correct decisions, and high numbers of correct decisions for both deceptive and truthful participants. However, we note that the optimal criteria for a decision rule may change with the purpose of a given examination. For example, in certain situations, it may be more important to avoid false negative errors, and in others it may be more important to avoid false positive errors. In these situations, an optimal decision rule would be one that produces high numbers of correct decisions for deceptive and truthful participants, respectively, but not necessarily for both.

In light of the previous work by Senter et al. (2004) and Senter and Dollins (2004), we expected to find substantial differences as a function of decision rule, especially across deceptive and truthful participants. More specifically, we expected those decision rules that included the spot scoring rule in some capacity to perform well with deceptive participants, and perhaps not as well with truthful participants. This prediction is derived from the simple fact that, as described previously, it is simply easier to produce a decision of deceptive using the spot score rule, as opposed to a decision of truthful. Finally, we expected to find a higher proportion of correct decisions and fewer no opinion decisions with those decision rules that evaluated five charts in some capacity. The foundation for the latter prediction lies not only in evidence gained from previous works, but also from basic statistical principles which suggest that more reliable research results

can be obtained with the inclusion of greater amounts of data.

Method

The assigned scores collected by Senter and Dollins (2004) and Senter et al. (2004) were used to produce decisions using eight different decision rules (listed in Table 1). The Senter and Dollins data set included scores produced by four evaluators of 16 deceptive and 16 truthful participants. Each of the four evaluators were trained at the Defense Academy for Credibility Assessment and had served as polygraph examiners for the federal government for at least 10 years. The Senter et al. data set included scores produced by five different evaluators of 50 deceptive and 50 truthful participants. Two of the evaluators in this data set were trained at the Defense Academy for Credibility Assessment, had served as polygraph examiners for the federal government, and were on staff at the Defense Academy for Credibility Assessment at the time of the study. One of these examiners was trained at the Backster School of Lie Detection (San Diego, CA), had served as polygraph examiners for the federal government, and was on staff at the Defense Academy for Credibility Assessment at the time of the study. Each of these three evaluators had at least 10 year of polygraph experience. The other two evaluators were trained at the University of Utah, a highly-regarded center of polygraph research, with one being a Professor of Psychology at the time of data evaluation and the other being a graduate student. It is worth noting that the evaluator trained at the Backster School was not included in the final Senter et al. report, as the comparison of Utah versus Defense Academy for Credibility Assessment data evaluation training was a primary emphasis of that work.

Analyses of the Senter and Dollins and Senter et al. data sets were conducted separately as data sets 1 and 2, respectively. All decision rules used total cutoffs of +6 and -6 for truthful and deceptive decisions, respectively. Spot score rules also used a -3 cutoff for a single question pair. Two decision rules used three question series, with one using the total cutoff rule (3T) and one using the spot score rule (3S). Two decision rules

Table 1
Decision Rules

Model	Description
3T	three question series rule, total cutoff rule
3S	three question series rule, spot score rule
5T	five question series rule, total cutoff rule
5S	five question series rule, spot score rule
3T5T	three or five question series rule, total cutoff rule
3S5S	three or five question series rule, spot score rule
3T5S	three or five question series rule, total cutoff rule used on first three question series, if no opinion, spot score rule used on all five question series
3S5T	three or five question series rule, spot score rule used on first three question series, if no opinion, total cutoff rule used on all five question series

used five question series, again with one using the total cutoff rule (5T) and one using the spot score rule (5S). The four remaining decision rules used the 'three or five' question series rule and a variation of the spot score and total cutoff rules. Two of these decision rules used either the total cutoff rule (3T5T) or the spot score rule (3S5S). The 3T5T decision rule used the total cutoff rule to produce decisions after three question series, and if a no opinion decision was produced, the total cutoff rule was used with all five question series. The 3S5S decision rule used the spot score rule to produce decisions after three question series, and if a no opinion decision was produced, the spot score rule was used with all five question series. The 3T5S and 3S5T decision rules used both the spot score rule and total cutoff rule along with the 'three or five' question series rule. The 3T5S decision rule used the total cutoff rule for the first three question series and the spot score rule for all five question series if a no opinion decision was made after the first three question series. The 3S5T decision rule used the spot score rule for the first three question series and the total cutoff rule for all five

question series if a no opinion decision was made after the first three question series.

Based on previous findings (Senter et al., 2004; and Senter & Dollins, 2004), we expected the decision rules that included the 'three or five' question series rule (3T5T, 3S5S, 3T5S, and 3S5T) to produce more correct decisions, fewer incorrect decisions, and fewer no opinion decisions than those that included the three question series rule (3T and 3S). We predicted that those decision rules using the five question series rule (5T and 5S) would fall somewhere in between. We also expected decision rules that included the spot score rule (3S, 5S, and 3S5S) to produce more correct decisions, fewer incorrect, and fewer no opinion decisions with deceptive participants than with truthful participants. We predicted the reversed trend with decision rules that included the total cutoff rule (3T, 5T, and 3T5T). Finally, we expected the two decision rules that included both the spot score rule and the total cutoff rule (3T3S, 3S5T) to fall somewhere in between these two extremes.

Separate 2 (Veracity [deceptive vs. truthful]) X 8 (Decision rule [3T, 3S, 5T, 5S, 3T5T, 3S5S, 3T5S, 3S5T]) repeated measures analysis of variance (ANOVA) statistics (Keppel, 1991) were calculated on the percentages of correct, incorrect, and no opinion decisions, with evaluator ($n = 4$ for data set 1, and $n = 5$ for data set 2) used as the unit of analysis. It is important to note that the three decision types are not independent of each other; large values in one category will result in smaller values in another, as they are all produced from the same pool of possibilities.

Results: Data Set 1

Table 2 shows the mean percentage of correct, incorrect, and no opinion decisions by participant veracity for the eight decision rules using assigned scores from data set 1. As Table 2 shows, there were differences across cells in the number of correct, incorrect, and no opinion decisions. Homogeneity of variance violations were indicated for all three analyses ($F_{\max} = 21.8, 9.0, \text{ and } 9.0$, for correct, incorrect, and no opinion decisions, respectively). We followed Keppel's (1991) suggestion and reduced the significance criterion to the more conservative .01 level for all analyses to compensate for variance homogeneity violations.

A Veracity X Decision rule repeated measures ANOVA was calculated for the proportion of correct decisions produced by the four scorers. The main effect of Decision rule was significant, $F(7,21) = 21.3, p < .01$. There was an increasing percentage of correct decisions from the 3T decision rule to the 3S5T decision rule for total decisions. The main effect of Veracity was not significant. The Veracity X Decision rule interaction was significant, $F(7,21) = 20.4, p < .01$. As Table 2 shows, the accuracy for deceptive participants was higher than for truthful participants for the three decision rules that used the spot score rule exclusively (e.g., 3S, 5S, 3S5S) ($\Delta M = 8.0\%$, where ΔM indicates the value for deceptive examinations minus the value for truthful examinations). The decision rules that included both the spot score rule and the total cutoff rule (e.g., 3S5T, 3T5S), produced higher accuracy for truthful participants than for deceptive participants ($\Delta M = -11.5\%$).

Finally, the decision rules that included only the total cutoff rule (e.g., 3T, 5T, 3T5T) produced a much higher accuracy for truthful participants than for deceptive participants ($\Delta M = -30.7\%$).

For incorrect decisions, only the Veracity X Decision rule interaction was significant, $F(7,21) = 6.2, p < .01$. The percentage of incorrect decisions across participant veracity increased from those decision rules using the spot score rule exclusively ($\Delta M = 4.0\%$), to those using both the spot score and total cutoff rules ($\Delta M = 8.5\%$), to those using only the total cutoff rule ($\Delta M = 13.7\%$).

For no opinion decisions, the main effect of Decision rule was significant, $F(7,21) = 18.9, p < .01$. Table 2 shows the decrease in no opinion decisions from the 3T decision rule to the 3S5T decision rule. The main effect of Veracity was not significant for the percentage of no opinion decisions. The no opinion decision Veracity X Decision rule interaction was significant, $F(7,21) = 5.4, p < .01$. As shown in Table 2, fewer no opinion decisions were produced for deceptive participants than for truthful participants by decision rules using the spot score rule exclusively ($\Delta M = -4.0\%$), and this tendency reversed with decision rules using both decision approaches ($\Delta M = 3.0\%$), and grew more extreme with decision rules that use only the total cutoff rule ($\Delta M = 17.0\%$).

Finally, pairwise comparisons indicated that the 3T5S decision rule produced a significantly greater percentage of correct decisions than the 3S, 3T, and 5T decision rules, $p < .01$. In addition, the 3T5T decision rule produced a significantly greater percentage of correct decisions than the 3S decision rule, and the 3S5T decision rule produced a significantly greater number of correct decisions than the 3T decision rule. No other pairwise comparisons were significantly different in terms of percentage of correct decisions.

Table 3 shows the proportion of agreement of each decision rule when applied to the decisions produced by the four scorers. This value represented the mean proportion of cases for which each pair of scorers made the same decision. There were no significant differences among the proportions of agreement.

Table 2

Mean proportion of correct, incorrect, and no opinion decisions as a function of decision rule and participant veracity (Data Set 1)

Rule	Deceptive			Truthful			Total		
	Cor.	Incor.	NO	Cor.	Incor.	NO	Cor.	Incor.	NO
3T	51.6	10.9	37.5	82.8	0.0	17.2	67.2	5.5	27.3
3S	70.3	6.3	23.4	62.5	7.8	29.7	66.4	7.0	26.6
5T	54.7	17.2	28.1	81.2	1.6	17.2	68	9.4	22.7
5S	73.4	7.8	18.8	60.9	14.1	25.0	67.2	10.9	21.9
3T5T	56.2	17.2	26.6	90.6	1.6	7.8	73.4	9.4	17.2
3S5S	78.1	9.4	12.5	73.4	12.5	14.1	75.8	10.9	13.3
3T5S	73.4	14.1	12.5	85.9	4.7	9.4	79.7	9.4	10.9
3S5T	73.4	15.6	10.9	82.8	7.8	9.4	78.1	11.7	10.2

Note. Cor. = Correct, Incor. = Incorrect, NO = No Opinion.

Discussion

The analyses of data set 1 reflected the impact of both question series and decision rules. First, the use of the 'three or five' question series rule produced significantly more correct decisions, and significantly fewer no opinion decisions than when 'three or five' question series were used. Thus, the use of the 'three or five' question series rule appeared to increase the percentage of correct decisions primarily through the correct resolution of more no opinion decisions. Second, decision rule had a large impact on the frequency of correct, incorrect, and no opinion decisions as a function of participant veracity. The spot score rule produced more correct decisions, and fewer incorrect and no opinion decisions for deceptive participants versus truthful participants. In contrast, the total cutoff rule produced more correct decisions, and fewer incorrect and no opinion decisions for truthful participants versus deceptive participants.

The two decision rules that employed both decision rules and the 'three or five' question series rule (e.g., 3T5S and 3S5T) appear to be the most effective for two reasons. First, these approaches produced the greatest number of total correct decisions. Second, the percentages of correct decisions for the 3T5S and 3S5T decision rules were among the highest for both deceptive and truthful participants. While the 3S5T and 3T5S decision rules produced higher accuracy for truthful versus deceptive participants, this may reflect the overall trend of correct decisions for the data set ($M = 77.5\%$ vs. 66.4% , for truthful and deceptive participants, respectively). These rules do not appear to be heavily biased by participant veracity, while maintaining a high total accuracy. However, conclusions regarding the effectiveness of any decision rule should be made with caution when considering a single data set. Thus, the same analysis was conducted on the Senter et al. (2004) data to determine the generalizeability of these effects.

Table 3**Mean Proportion of Scorer Agreement as a Function of Eight Decision Rules
(Data Set 1)**

Rule	Proportion of Agreement
3T	.69
3S	.69
5T	.78
5S	.69
3T5T	.71
3S5S	.69
3T5S	.69
3S5T	.71

Results: Data Set 2

Table 4 shows the mean percentage of correct, incorrect, and no opinion decisions by participant veracity for the eight decision rules using assigned scores from data set 2. The previous analyses were replicated on the proportion of correct, incorrect, and no opinion decisions. Because of the homogeneity violation indicated by the highly variant standard deviations for correct, incorrect, and no opinion decisions ($F_{max} = 169, 49, \text{ and } 225$, respectively), the alpha level was reduced from .05 to .01.

For correct decisions, the main effect of Veracity was significant, $F(1,4) = 26.2, p < .01$. As Table 4 shows, the mean percentage of correct decisions for deceptive cases was higher than that for truthful cases (83% vs. 70%, respectively). The main effect of Decision rule was also significant, $F(7,28) = 19.7, p < .01$. Table 4 shows an increase in correct decisions from the 3T decision rule to the 3S5T decision rule. The Veracity X Decision rule interaction for correct decisions was also significant, $F(7,28) = 22.6, p < .01$. Those decision rules that included the spot score

rule exclusively produced more correct decisions for deceptive cases than for truthful cases ($\Delta M = 26.0\%$), versus the two decision rules that included both decision approaches ($\Delta M = 13.0\%$), and those that included only the total cutoff rule ($\Delta M = -0.7\%$).

For incorrect decisions, the main effect of Decision rule was significant, $F(7,28) = 9.4, p < .01$. Those decision rules which used the spot score rule exclusively produced more incorrect decisions ($M = 8.7\%$) than those which used both decision rules ($M = 8.5\%$), and those which only used the total cutoff rule ($M = 5.0\%$). The linear polynomial trend was not significant, $p > .01$. The main effect of Veracity was not significant. The Veracity X Decision rule interaction was significant for incorrect decisions, $F(7,28) = 10.3, p < .01$. As Table 4 shows, those decision rules which used only the spot score rule produced fewer incorrect decisions for deceptive than for truthful participants ($\Delta M = -8.7\%$), as did those which used both decision rules ($\Delta M = -5.0\%$). Decision rules which used only the total cutoff rule produced more incorrect decisions for deceptive than for truthful participants ($\Delta M = 2.3\%$).

Table 4

Mean proportion of correct, incorrect, and no opinion decisions as a function of decision rule and participant veracity (Data Set 2)

Rule	Deceptive			Truthful			Total		
	Cor.	Incor.	NO	Cor.	Incor.	NO	Cor.	Incor.	NO
3T	62.8	5.6	31.6	70.8	2.0	27.2	67.2	5.5	27.3
3S	78.4	4.0	17.6	56.4	10.8	32.8	66.4	7.0	26.6
5T	81.6	5.2	13.2	78.0	4.4	17.6	68	9.4	22.7
5S	90.0	4.0	6.0	60.0	13.2	26.8	67.2	10.9	21.9
3T5T	81.6	6.8	11.6	80.0	4.8	15.2	73.4	9.4	17.2
3S5S	89.6	5.2	5.2	64.4	14.8	20.8	75.8	10.9	13.3
3T5S	88.8	6.0	5.2	72.8	10.0	17.2	79.7	9.4	10.9
3S5T	86.0	5.6	8.4	76.0	11.6	12.4	78.1	11.7	10.2

Note. Cor. = Correct, Incor. = Incorrect, NO = No Opinion.

A significant no opinion decision Veracity main effect, $F(1,4) = 59.6$, $p < .01$, indicated more no opinion decisions for truthful ($M = 17.0\%$) than for deceptive participants ($M = 12.4\%$). The main effect of Decision rule was significant, $F(7,28) = 30.1$, $p < .01$. Table 4 shows that the three question series rule produced the most no opinion decisions ($M = 27.0\%$), the 'three or five' question series rule produced the least ($M = 11.8\%$), and the five question series rule fell in the middle ($M = 15.5\%$). Decision rules using only the spot score rule produced more no opinion decisions for truthful than for deceptive participants ($\Delta M = -17.3\%$), than did decision rules involving both rules ($\Delta M = -8.0\%$), and those using only the total cutoff rule ($\Delta M = 1.7\%$).

Finally, pairwise comparisons indicated that all decision rules (with the exception of the 3T decision rule) produced a significantly greater percentage of correct decisions than the 3S decision rule (all $ps < .01$). In addition, the 3T5S decision rule produced a significantly greater percentage of

correct decisions than the 3T decision rule. No other pairwise comparisons were significantly different for the percentage of correct decisions.

Table 5 shows the proportion of agreement of each decision rule when applied to the decisions produced the four scorers. There were no significant differences across decision rules in the proportion of agreement.

Discussion

The pattern of results from data set 2 resembled that of data set 1. Decision rules that included the 'three or five' question series rule produced more correct decisions and fewer no opinion decisions than decision rules using three question series. Thus, the increased percentage of correct decisions again appeared to be due to the resolution of no opinion decisions, rather than the conversion of incorrect decisions into correct decisions. One difference between data sets 1 and 2 was that the percentage of correct decisions using the five question series rule

Table 5**Mean Proportion of Scorer Agreement as a Function of Eight Decision Rules
(Data Set 2)**

Rule	Proportion of Agreement
3T	.75
3S	.78
5T	.82
5S	.79
3T5T	.81
3S5S	.79
3T5S	.81
3S5T	.81

was equivalent to that obtained using the 'three or five' question series rule.

The use of the spot score rule again produced more correct decisions, and fewer incorrect and no opinion decisions for deceptive participants than for truthful participants. The total cutoff rule was equally accurate for deceptive and truthful participants in data set 2.

General Discussion

Analysis of the results produced by the eight decision rules have revealed useful information regarding the impact of question series rules and decision rules across two different data sets. First, the most effective decision rules are those that employ the 'three or five' question series rule. The three question series rule is clearly inferior to the 'three or five' question series rule across both data sets, but the difference between 'three or five' and five question series is not as stable. While the difference in the percentage of correct decisions was not large for data set 2 ($M = 78.5\%$ vs 80.0% for decision rules using

the five and 'three or five' question series rules, respectively), it was quite large for data set 1 ($M = 67.5\%$ vs 76.8% for the five and 'three or five' question series rules, respectively). Thus, although using five question series may be an effective approach in some circumstances, the use of the 'three or five' question series appears to produce consistently higher decision accuracy. Second, the differential impact of the two decision rules was clear and consistent across both data sets. The spot score rule produced more correct decisions, and fewer incorrect and no opinion decisions for deceptive participants than for truthful participants. In contrast, the total cutoff rule tended to produce the reversed pattern of accuracy.

All decision rules considered in this study were effective in the sense that they diagnosed deception at above chance levels, and all were highly accurate (all greater than 86%) when no opinion decisions were excluded. As stated earlier, the optimal rule may depend upon the purpose of the polygraph examination. This study has shown that decision rules using the spot score rule

exclusively (i.e., 3S, 5S, and 3S5S) are more effective at identifying deceptive individuals, decision rules using the total cutoff rule exclusively (i.e., 3T, 5T, 3T5T) are more effective at identifying truthful individuals, and the 3T5S and 3S5T decision rules are effective at identifying both deceptive and truthful individuals. We suggest that the optimal decision rules are those that are equally effective in identifying deceptive and truthful individuals, but acknowledge that decision rules that are more effective in the detection of either deceptive or truthful individuals may be preferable in some circumstances.

One difference between data sets 1 and 2 was in the percentage of correct decisions as a function of participant veracity, though the total percentage of correct decisions did not differ significantly ($M = 72.0\%$ vs 76.2% for data sets 1 and 2, respectively, $z = -1.0$, $p > .05$). In data set 1, the mean accuracy across all eight decision rules was higher for truthful ($M = 77.6\%$) than for deceptive individuals ($M = 66.3\%$). In data set 2, mean accuracy was higher for deceptive ($M = 82.5\%$) than for truthful individuals ($M = 69.8\%$). For truthful participants, the accuracy rates for data sets 1 and 2 was not significantly different, $z = 1.2$, $p > .05$. However, for deceptive participants, accuracy was significantly higher in data set 2 than in data set 1, $z = 2.9$, $p < .05$. The source of this difference may be the experimental procedures used in the two studies. Data set 1 was collected during the development of a mock crime scenario to be used for the study of the psychophysiological detection of deception. In contrast, data set 2 was collected during a study using an established mock crime paradigm (Kircher & Raskin, 1988). Thus, it is possible that these differences occurred because the mock crime paradigm in one study was more effective than in the other. While the differences in these two data sets may appear to be problematic, they served as a contrast to determine the robustness of these decision rules across different data sets. In spite of the differences between the two data sets, common patterns were produced across participant veracity as a function of decision rule.

There were some limitations associated with this study. First, the number of scorers for both data sets 1 and 2 was small only four

and five, respectively. This raises some concerns as to how well these findings will generalize to the larger population of polygraph scorers. Second, while data set 2 included a large number ($N=100$) of participants, we can be somewhat less confident in the results for data set 1, as there were only 32 participants. Finally, the comparisons between the 'three or five' and 'three or five' question series decision rules were not completely clear, and this study did not include a straightforward comparison between the three and five versus the 'three or five' question series decision rules, as they were explored in combination with other rules.

One issue related to the collection of data beyond three question series is habituation, or the diminishing of responses with repeated questioning, and hence a reduction of the diagnostic value of additional question series. Some work has provided evidence supporting the existence of this phenomenon (Balloun & Holmes, 1979; Suzuki & Hikita, 1981), while other studies have shown that responses do not degrade with repeated presentations (Dollins, Cestaro, & Pettit, 1998; Elaad & Ben-Shakhar, 1997; Nakayama & Kizaki, 1990; Yankee & Grimsley, 1987). In the present study, the use of additional question series in the decision process produced substantial increases in accuracy, further suggesting that diagnostic responses continue to occur after the first three question series.

Two conclusions can be drawn from this study. First, the use of a 'three or five' question series rule is the most effective method to maximize the percentage of correct decisions, relative to making decisions from only three question series or only five question series, regardless of the specific decision rule used in conjunction with this approach. Second, the spot score and the total cutoff rules are biased towards the detection of deception, and truthfulness, respectively. In other words, the spot score rule produced more correct decisions for deceptive participants than for truthful participants and the total cutoff rule produced more correct decisions for truthful participants than for deceptive participants.

Clearly, some caution must be taken when evaluating these conclusions. These

results were produced using laboratory data, and not data collected in real world field situations. It would be prudent to examine field data to determine the generalizeability of these findings, not only to additional data, which is always necessary, but to determine the applicability of these decision rules to real world examinations. Finally, the results of the current study do not address the potential for

no opinion decisions to be resolved during subsequent psychophysiological detection of deception examinations. Future studies should compare the performance of decision rules which tend to resolve a greater percentage of no opinion decisions in single session, with the performance of procedures that use additional sessions to resolve no opinion decisions.

References

- Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology*, 64, 316-322.
- Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. *Polygraph*, 28, 1-9.
- Defense Academy for Credibility Assessment. (2002). *Zone comparison test*. Fort Jackson, SC: Author.
- Department of Defense Polygraph Institute Research Division Staff. (2001). Test of a mock theft scenario for use in the psychophysiological detection of deception: IV. *Polygraph*, 30(4), 244-253.
- Dollins, A. B., Cestaro, V. L., & Pettit, D. J. (1998). Efficacy of repeated physiological detection of deception testing. *Journal of Forensic Science*, 43, 1016-1023.
- Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology*, 34, 587-596.
- Federal psychophysiological detection of deception examiner handbook*. (1999, May). Fort Jackson, SC: Department of Defense Polygraph Institute.
- 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., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79(2), 252-259.
- 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.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.
- Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73(2), 291-302.
- Nakayama, M., & Kizaki, H. (1990). Usefulness of the repeated presentation of questions on the psychophysiological detection of deception. *The Japanese Journal of Psychology*, 60 (6), 390-393.
- Podlesny, J. A.; Truslow, C. M. (1993). Validity of an expanded-issue (Modified General Question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78(5), 788-797.
- Senter, S. M., & Dollins, A. B. (2004). Comparison of three versus three or five question series contingency rules: A replication. *Polygraph*, 33(4), 223-233.
- Senter, S. M., Dollins, A. B., & Krapohl, D. J. (2004). Comparison of Utah and DoDPI scoring accuracy: Equating veracity decision rule, chart rule, and number of data channels used. *Polygraph*, 33(4), 214-222.

- Suzuki, A., & Hikita, Y. (1981). An analysis of response on polygraph: A diminution of responses. *Polygraph*, 10(1), 1-7.
- Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph*, 28, 10-27.
- Yankee, W. J., & Grimsley, D. L. (1987). The effect of a prior polygraph test on subsequent polygraph tests [Abstract]. *Psychophysiology*, 24, 621-622.

Degrees of Deception: Diploma Mills and the Polygraph Examiner Community — A Recommendation for Change

Timothy J. Weber and Frank Horvath

In the past half-decade the U.S. government has made a strong effort to investigate the extent to which federal employees have used academic degrees from unaccredited colleges and universities to meet established criteria for employment or promotion. (These institutions are often referred to as “diploma mills;” we use that term interchangeably with “unaccredited” institutions, though they may not necessarily be the same.) Surprisingly, an initial inquiry by the General Accounting Office (GAO) discovered that 463 government employees had been students in two unaccredited schools. An additional 28 employees in high-level civil service positions were also found to have been enrolled in such schools. Moreover, it was shown that the government spent over \$169,000 in tuition fees at these two schools alone and the GAO identified 43 other, similar academic programs attended by government employees.

The GAO reported that the number of government employees holding degrees from unaccredited institutions was likely to be underreported because such schools tend to use names that closely resemble legitimate academic institutions. Even the highest levels of government were affected by this concern when it was revealed that Laura Callahan, a member of the senior executive service at the Department of Homeland Security, was forced to resign. A standard security clearance background investigation revealed that her bachelor’s, master’s and doctorate degrees were all received from the same institution, the unaccredited Hamilton University (Cramer, 2004). In her defense, Callahan claimed that she was unaware that Hamilton was not accredited until after it was reported in the media (Ezell & Bear, 2005, p.295).

The GAO report resulted in greater vigilance in the government’s hiring process. Vacancy announcements from the US Department of Labor, for example, now warn

potential applicants of the implications of submitting degrees from unaccredited institutions:

The Department of Labor does not recognize academic degrees from schools that are not accredited by an accrediting institution recognized by the Department of Education. Any applicant falsely claiming an academic degree from an accredited school will be subject to actions ranging from disqualification from federal employment to removal from federal service (US Department of Labor, 2005).

Because of the attention now being given to the problem of illegitimate academic degrees in the federal government, it was of interest to examine this issue in the field of polygraphy, a field composed of persons engaged in “seeking the truth” for both public and private sector purposes. There are several reasons for doing this, and this paper is organized, in part, based on those. First, a description of the nature of the problem of “diploma mills” and the manner in which such institutions differ from those that are accredited is presented. Second, persons who are informed about the first issue would be more likely to initiate open and vigorous discussion of “diploma mills” and their effect on the community of polygraph examiners. Third, a full discussion of the concerns regarding “diploma mills” will, hopefully, lead to policy changes within organizations representing polygraph examiners that ultimately will improve the field. The paper concludes with recommendations for changes in the current policy of the American Polygraph Association.

What is a “Diploma Mill”?

In their review of online resumes and news media reports, the famous diploma mill hunters, Ezell and Bear (2005, p.268),

revealed a number of “unexploded time bombs” in the resumes of prominent individuals. One of these was a “Polygraph expert with a fake Ph.D.” This “expert,” claimed to hold a doctorate degree from LaSalle University; the degree was acquired shortly before the FBI raided that institution and its founders pled guilty to mail fraud.

One of the reasons such occurrences arise is due to unscrupulous and widespread advertising. This has certainly deceived some who enroll in diploma mills believing that their prospective degree will actually be legitimate. Others who enroll in such institutions understand in advance the fraudulent nature of the institutional credentials. Regardless, both categories of students are negatively affected.

The advent of Internet, as well as the increase in legitimate distance education programs, has enabled diploma mills to flourish as an estimated 20 million dollar a year industry (Ezel & Bear, 2005 p.15). Distinguishing between “diploma mills” and accredited institutions has become more difficult because of changes in modes of educational delivery. Although those in academia traditionally have not been enthralled with distance or internet-based education, that situation has changed in recent years (Merrill, 1999). Now, a large number of accredited colleges and universities engage in what has come to be known as e-learning, or online education. The practice is now widely recognized as an accepted and emerging trend in higher education. For example, in the 1999-2000 academic year, eight percent of undergraduate students participated in some form of “distance education.” Of these, 60% received their education entirely through Internet courses. Detractors of this practice point to the higher dropout rate for undergraduate on-line students of 33% versus 22% for those attending brick and mortar institutions (National Center for Educational Statistics, 2002). Conversely, proponents applaud the convenience of asynchronous on-line courses that can fit into a work schedule and can be accessed worldwide, allowing students tremendous flexibility in pursuit of an education.

Notwithstanding the trend in educational delivery, it is unfortunately true

that not all schools that advertise distance or internet-based programs are interested in the academic needs of the student. What distinguishes a diploma mill from a reputable American institution of higher learning is its lack of accreditation by an accrediting organization recognized by the U.S. Department of Education. The importance of such accreditation cannot be overemphasized. Accreditation provides prospective students with the assurance that the academic institution’s faculty, curriculum, facilities, and administration have been inspected and are in compliance with the highest educational standards. Unaccredited institutions may provide some course work, but they do not provide a diploma that is recognized as legitimate by accredited institutions or the U.S. Department of Education.

In most countries, the State, or “central” government, accredits educational institutions directly. In the United States the process differs in that colleges and vocational schools are licensed (to practice) by individual states and may be “accredited” by non-governmental organizations. The U.S. Department of Education (DOE) recognizes only some of these regional and national accreditation organizations as legitimate.

Regional agencies are generally the most commonly recognized accrediting bodies in academia. They include the Middle States Association of Colleges and Schools, New England Association of Schools and Colleges, North Central Association of Colleges and Schools, Northwest Commission on Colleges and Universities, Southern Association of Colleges and Schools, and the Western Association of Schools and Colleges. National accrediting agencies, such as the Distance Education and Training Council, or the Accrediting Council for Independent Colleges and Schools (ACICS) may be more limited in some ways than regional bodies. Students graduating from nationally accredited programs may face restrictions when applying for graduate school or transfer of credit.

In addition to regional and national accrediting bodies, some specific programs of instruction within an institution of higher education also may be accredited by organizations that are recognized by the U.S. Department of Education. The American Bar

Association's accreditation of law schools is an example of this process (Merrill, 1999).

Because they don't conform to traditional academic and institutional standards that the DOE provides through its recognized accrediting organizations, some diploma mills simply create their own "accrediting" bodies or they join with other institutions to establish such a body. These organizations then "accredit" the diploma mills, who proudly announce themselves as "fully accredited" on their web sites. For example, Columbus University is accredited by the World Association of Colleges and Universities (WACU), said to be a "global accreditation association founded to establish and promote global standards in higher education among colleges and universities worldwide" (Columbus University, 2000). However, the WACU is not recognized as an accrediting organization by the DOE, and therefore its evaluation of a school's status is of no value. Some of the unrecognized accreditation organizations also claim, wrongly, to accredit prominent universities. They do this to make it appear, to those who are naïve about the accrediting process, that they are legitimate accrediting bodies. Some "diploma mills" capitalize on this by claiming accreditation from the unrecognized body, misleading those who might apply for admission to their programs (Ezell & Bear, 2005,).

Diploma mills may also deliberately attempt to confuse prospective students by stating that they are recognized or licensed by a state government agency. This is merely an attempt to gain legitimacy even though it has no relationship to "official" accreditation. This argument has been expressed by at least one person who received a doctoral degree from Columbia Pacific University (CPU) in 1989. According to this person, CPU had been granted "Full Institutional Approval" as a "California Degree-Granting Institution by the California State Department of Education, after the Superintendent of Public Instruction impaneled a qualified visiting committee and conducted a comprehensive on-site qualitative review and assessment of the institution and all programs offered (Matte, 2005). But, such state approval is not the same as DOE recognized accreditation. Those who are unaware that state regulatory agencies are not accrediting organizations and are only

responsible for the licensing of the school as a business fail to understand the accrediting process (State of California, 2005).

In the case of CPU it is worth noting that it eventually lost its "Full Institutional Approval" and license to practice in the State of California. In October 2000, the California Bureau for Private Postsecondary and Vocational Education's denial of CPU's re-approval application was upheld and affirmed by the California Supreme Court. CPU was never at any time an accredited degree-granting institution. CPU is now located in another state.

What is also confusing is that there are different levels of accreditation. Colleges and universities are accredited at four levels: associate's, bachelor's, master's and doctorate, based upon the highest degree that they confer. Non-degree granting vocational programs may be accredited at the certificate level such as the Defense Academy for Credibility Assessment (DACA) which is accredited by ACICS. The accreditation at the certificate level means that the DACA may not grant a degree, but may provide students with a certificate of program completion. The DOE recognizes many vocational accrediting organizations such as the National Accrediting Commission of Cosmetology Arts and Sciences which accredits "postsecondary schools and departments of cosmetology arts and sciences and massage therapy" (US Department of Education, 2005). There are also legitimate vocational schools that are unaccredited. The student must determine if the curriculum meets his or her needs and whether potential employers will recognize the school's diploma.

Identification of diploma mills that use the Internet as their sole delivery vehicle is further confounded by deceptive advertising practices. Some of these organizations have professional quality web sites, featuring pictures of a beautiful campus, classrooms and administration facilities, and smiling "students." The fact that the photos are actually taken from travel brochures or other similar sources is not, of course, noted in the advertisements (Ezell & Bear, 2005, p.99). Some diploma mills even produce professional quality video productions that describe in great detail their degree granting programs (Columbus University, 2000). Not

uncommonly, prospective students are drawn in by advertisements that boldly announce: “No need to take admissions tests, no need to study. Receive a college degree for what you already know” (Ashwood University, 2005).

Regardless of their claims, there are several elements that serve to identify diploma mills. Reputable colleges and universities typically charge tuition by the credit hour or, in some cases, by the semester. Diploma mills typically charge by the degree. Diploma mills are also relatively inexpensive compared to the average university. Tuition for a doctoral degree from Ashwood University, which is unaccredited by any accrediting organization recognized by the DOE, is only \$599 dollars (Ashwood University, 2005). This sum is not realistic; one of the authors, for instance, spent over \$33,000 in tuition payments alone in a doctoral program at the regionally accredited Nova Southeastern University.

Diploma mills sell “paper” degrees and, because that is their real source of income, ostensible but meaningless tuition bargains can be found. The unwary prospective student may be offered a bargain if multiple degrees are purchased at the same time. An example of this occurred when one of the authors failed to respond to the literature he requested from Columbus University. After that first request he received a coupon for \$100 dollars off the standard cost of tuition. When this failed to produce a positive response, another letter from that university was sent. This time he was offered a \$200 dollar tuition reduction.

The time required to complete programs in diploma mills is also a telltale sign. Lacrosse University may require only 14 months to complete a doctoral program, while the time spent by students enrolled in accredited institutions is measured in years (Lacrosse, 2004). Some diploma mills actually provide degrees within days or weeks by granting academic credit for “life experiences.” While some life experiences may deserve legitimate academic credit, this is usually rarely done and only when the experiences can be documented to be equivalent to accepted college course work. For example, a student who submits official documentation of military service to an accredited university may be granted a limited number of credits in military

science, physical education or a related area. A diploma mill, however, will grant physical education credit to someone who merely claims to play golf on the weekends; indeed, some of them grant degrees based completely on self-reports of life experiences. Accredited institutions will not do this.

To appear legitimate, some diploma mills require completion of course work. Rest assured, however, that the course work will not have the depth or breadth comparable to that found in accredited programs. Pacific Western University, as one example, has only one required text for each course in its doctoral program (Pacific Western University, 2004). Columbus University is even less strenuous with one text for the entire doctoral program. Students may elect to complete chapter summaries or take 45 days to complete an open book examination in order to complete courses (Columbus University, 2000). At the unaccredited Novus University, a doctoral dissertation, typically an original, scholarly contribution of varying length, is required to be 55 pages long, as if the length of the work is indicative of its scholarly merit (Bartlett & Smallwood, 2004).

If the prospect of participating in any academic effort is unappealing it is possible to obtain a diploma in another way. Web sites are available on the Internet that provide “duplicate diploma services.” These sites will print a realistic diploma from an accredited university for “novelty purposes.” There is no legitimate need for such services; any graduate of an accredited university can easily obtain a duplicate diploma. Falsification of an academic diploma may be difficult to detect by superficial means. However, unlike diploma mills that provide a “backstopped cover” for their “graduates” in that they produce an “official” transcript and a mailing address to verify a degree, duplicate diploma services send their product directly to the purchaser. In order to detect such a scheme it is necessary to request that an official transcript be sent directly from the university to the requester (Ezell & Bear, 2005).

Investigation and Prosecution of Diploma Mills

The investigation and prosecution of unaccredited institutions who offer degrees is

not a high law enforcement priority. One exception to this rule was the FBI Diploma Scam (DIPSCAM), an investigation that started in 1980 and concluded in 1991. In this case, Special Agent Allen Ezell initiated an investigation after a major sports commentator announced during a televised football game that a player had graduated from a college in Greenville, South Carolina; the college was not legitimate. Special Agent Trace Kirk was assigned to investigate the matter in an undercover capacity. He was able to obtain fraudulent medical degrees via the postal service thus establishing mail fraud against the owner of the college. After his arrest, however, the owner committed suicide rather than face prison (Kirk, personal communication, 2004). Though DIPSCAM was initially aimed at what were perceived to be the worst offenders, those that provided fraudulent medical degrees, eventually the investigation resulted in the closing of dozens of institutions, a plethora of arrests, and the seizure of millions of dollars. When Ezell, who initiated the DIPSCAM investigation, retired from the FBI the inquiry came to an end (Ezell & Bear, 2005).

Efforts in the States to Deal with Diploma Mills

Just as the federal government has begun to seriously address the problem of illegitimate academic credentials so too some state governments also have acted in this regard. Recently enacted state statutes, for example, have criminalized not only the operation of institutions, but also the claim that one holds a degree if it is from an unaccredited institution. Four states have adopted statutes to protect the public. The state of Florida's criminal law is the most comprehensive. That statute reads in part as follows:

Making false claims of academic degree or title.

(1) No person in the state may claim, either orally or in writing, to possess an academic degree, as defined in s. 1005.00 or the title associated with said degree, unless the person has in fact, been awarded said degree from an institution that is:

(a) Accredited by a regional or professional accrediting agency

recognized by the United States Department of Education and the Commission on Recognition of Postsecondary Accreditation;

(b) Provided, operated, and supported by a state government or any of its political subdivisions or the Federal Government.

(c) A school, institute, college or university chartered outside the United States, the academic degree from which has been validated by an accrediting agency approved by the United States Department of Education as equivalent to the baccalaureate or post baccalaureate degree conferred by a regionally accredited college or university in the United States;

(d) Licensed by the Commission for Independent Education pursuant to ss. 1005.01-1005.38 or exempt from licensure pursuant to chapter 1005; or

(e) A religious seminary, institute, college or university which offers only educational programs that prepare students for a religious vocation, career, occupation, profession, or lifework, and the nomenclature of whose certificates, diplomas, or degrees clearly identifies the religious character of the educational program.

(2) No person awarded a doctorate degree from an institution not listed in subsection (1) shall claim in the state either orally or in writing the title "Dr." before the person's name or any mark appellation, or series of letters, numbers, or words, such as, but not limited to, "Ph.D.," "Ed.D.," "D.N." or "D.Th.," which signifies, purports, or is generally taken to signify satisfactory completion of the requirements of a doctoral degree, after the person's name.

(3) (a) A person who violates the provisions of subsection (1) or subsection (2) commits a misdemeanor of the first degree, punishable as provided in s.775.082 or s. 775.083

(b) In addition to any penalty imposed under paragraph (a) a violator shall be subject to any other penalty provided by law, including, but not limited to, suspension or revocation of the violator's license or certification to

practice an occupation or profession (Florida State Senate, 2006).

In effect, what the Florida statute means is that merely introducing oneself as “Doctor” could result in an arrest if one does not hold a doctoral degree from an accredited institution. The penalty in Florida includes imprisonment of up to a year and a fine of up to \$1,000 dollars (Florida State Senate, 2006).

Other states have taken action similar to that of Florida. In North Dakota it is a misdemeanor to use an illegitimate degree in any business dealings or to obtain employment, promotion or acceptance into an educational institution. The relevant statute states:

[It is] Unlawful to use degree or certificate when coursework not completed- penalty

1. An individual may not knowingly use a degree, certificate, diploma, transcript, or other document purporting to indicate that the individual has completed an organized program of study or completed courses when the individual has not completed the organized program of study or the courses as indicated on the degree, certificate diploma, transcript, or document:

- a. to obtain employment;
- b. to obtain a promotion or higher compensation in employment;
- c. to obtain admission to an institution of higher learning, or
- d. in connection with any business, trade, profession, or occupation.

2. An individual who violates this section is guilty of a class A misdemeanor (Ezell and Bear, 2005).

The state of Oregon has enacted civil penalties that range from \$300 when the public is deliberately deceived and no economic advantage is taken, to \$1,000 in fines for utilization of an illegitimate degree to entice business (Oregon Office of Degree Authorization, 2006). In the state of Washington recent legislation makes it a felony to operate an unaccredited college and a misdemeanor to use a fraudulent degree in advertising to solicit business (Washington State, 2007).

Diploma Mills and Polygraph Examiners

Aside from legal complications, there are also ethical considerations involved in unjustified claims of academic achievement. This would seem to be an obvious issue of concern in the community of polygraph examiners. A polygraph examiner who uses fraudulent academic credentials to work for a government agency denies qualified applicants access to that position. In the private sector, examiners using the title of “Doctor” illegitimately are likely to be inappropriately influencing consumers and examinees. The business advantage that may accrue from such a misimpression may be considerable and, as noted by some, could be seen as a deceptive business practice:

I think the question that the polygraph community, and especially the American Polygraph Association (APA), needs to come to terms with is whether it is acceptable for a member to place the title "Doctor" in front of his or her name when in fact that member has not earned any such degree from any accredited university. I think that this is a deceptive business practice and should be roundly condemned (Maschke, 2005).

While a business advantage may be leveraged by misuse of illegitimate academic credentials, the problem is exacerbated when a person who holds a degree from an unaccredited institution provides expert testimony in court or in another official forum. Here, exposure of fraudulent credentials discredits not only the individual examiner but also brings undeserved dishonor to those in the field. This is not a problem unknown in other specialty areas in the forensic sciences (Midkiff, 2005; Starrs, 2005)

There are indeed reported occasions when academic credentials from a “diploma mill” have been offered to the U. S. courts by polygraph examiners. In one case, for instance, an examiner’s signature on an amicus brief submitted to the US Supreme Court was claimed as proof of this practice (Maschke, 2003). While an academic degree is not required for the courts to recognize an individual as an expert, academic credentials are typically a part of a resume used to

establish a witness's expertise (Holten & Lamar, 1991). If an expert swears, as is typically required, to the truthfulness of the information contained in a resume, it raises a question about the trustworthiness of the expert's testimony. If the expert's credentials are challenged, and the use of a degree from a "diploma mill" is revealed under cross examination, there certainly would be questions raised regarding the value of that expert's testimony, as well as possible charges of perjury. Such a revelation might not only negatively impact the immediate court case, but it could also dishonor the individual expert (examiner) and the polygraph examiner community.

The APA is not the only polygraph organization affected by problems related to diploma mills. One of the founding members of the National Polygraph Association claims an advanced degree but without clearly stating that the degree was obtained from an unaccredited university (McGavock, 2006). In 2000, the Defense Academy for Credibility Assessment (DACA) implemented a policy requiring its faculty to have a master's degree or to be enrolled in a graduate program at an accredited college or university. The DACA transferred one of its instructors (assigned from another agency) after it was discovered that his master's degree, and the doctorate that he was in the process of obtaining, were both from unaccredited programs. Despite the apparent lack of academic rigor in this doctoral program, the instructor maintained that he was completely unaware that he had been and was currently enrolled in an academic program that was not legitimate. Faced with disciplinary action, he resigned his government position to become a contract examiner. At last report this person was enrolled in a master's degree program at an accredited university. In another case, one federal polygraph program rejected the services of a contract examiner after an investigation revealed that he had received his bachelor's degree from a "diploma mill." It was also revealed that this individual, for ten years prior to his application for the contractor position, had been conducting polygraph examinations under contract to another federal agency (Gaines, personal communication, December 12, 2004).

The Scope of the Problem in the Polygraph Examiner Community

Because there is no central registry for polygraph examiners in the U.S., it is impossible to identify and distinguish between those who hold academic credentials from accredited institutions and those who are "graduates" of diploma mills. It certainly would be of interest to determine the number of persons in each group at varying levels of academic achievement. The only statistics available on this issue, however, are those derived from two surveys carried out by Horvath, one in 1995 and the other more recently, in 2006. In 1995, a mailed questionnaire was sent to all identifiable polygraph examiners worldwide. In 2006, all members of the APA were surveyed by requesting each of them to access a specified Internet site and complete a questionnaire online. In both surveys, respondents were asked to specify their academic background; the highest level of education received and, if a college degree was held, the major area of study. The 1995 data showed that 1% reported only a high school diploma, 27% said they had "Some college," another 27% reported holding a bachelor's degree and 45% said they had a bachelor's degree and had done some graduate work (Horvath, 1996). In 2006, 1% again reported having only a high school diploma, 26% reported "some college," 29% said they held a Bachelor's degree, 20% reported some graduate work and 24% reported having either an LLB, a Master's degree, a Doctorate or another graduate-level degree. It is noted that only two of the respondents reported having a doctoral degree (Horvath, 2007).

Because the statistics derived from these surveys are based only on the data reported by the respondents, that is, those who were surveyed and cooperated by providing information, there is reason to believe that the data may not represent the entire polygraph examiner community. Equally important is that there is no way to determine if the academic credentials reported by the respondents were legitimate. In an effort to deal with this latter issue the authors surveyed the membership directory of the APA and used other sources to determine how many members held doctoral degrees and how many of these were identifiable as being from

an accredited or a non-accredited university or college. In this search we were able to identify twenty-seven APA members who reportedly held doctoral degrees. Of these, six were verified by transcript to hold a doctoral degree from an accredited institution. Nine members held a doctoral degree from a university or college known to be unaccredited by any accrediting agency recognized by the DOE in accordance with accepted academic standards. In the remaining twelve cases, eight claimed to be graduates of regionally accredited institutions, two claimed to be graduates of foreign universities, and in two cases the degree granting college or university could not be identified based upon the information available.

Resolution of the “Diploma Mill” Problem

Paragraph 4.8.1 of the APA’s “Code of Ethics” states, in part: “A member shall not knowingly make, publish, or cause to be published any false or misleading statements or advertisements relating to the Association or the polygraph profession.” (APA, 2005). This statement is the only one in the Code of Ethics that might be construed to cover the problem raised when a member uses academic credentials from a diploma mill to advertise services as a polygraph examiner. But it is clear for a number of reasons that the statement may not justify action against a member who claims to hold a doctoral or other degree from an unaccredited institution.

Perhaps because of the ambiguity in the APA’s Code of Ethics, and a concern raised by some members about the use of degrees from diploma mills, the Board of Directors discussed and attempted to deal with the problem by implementing a resolution in which the APA’s “official” position on members’ use of academic credentials was more directly expressed. That policy states:

Be it resolved that, the American Polygraph Association does not endorse or recognize diplomas or degrees from post secondary courses of education not accredited by an accrediting body routinely accepted within the academic community or appropriate government entity at the time of the award.

The APA will not accept advertisements, publications or directory listings which include a reference or title representing a diploma (sic) or degree from a post secondary course of education not so accredited.

The APA Board of Directors directs that should there be a question of whether a reference to a title or degree, a member may be required to produce evidence that their diploma or degree meets the requirements of this resolution before acceptance of any advertisement, publication or directory listing which include a reference to such diploma or degree (APA, 2004).

Although this policy statement certainly suggests an effort by the APA to deal with the ethical and other issues raised by the use of diploma mill degrees, it is, like the phrase in the Code of Ethics, inadequate. The policy is vague in that it permits diploma mill degree holders to argue that their degree is legitimate because their institution was licensed by a state and was “accredited” by an agency that is “accepted” by other similar institutions. The section of the policy that states “not accredited by an accrediting body routinely accepted within the academic community or appropriate government entity at the time of the award” should be altered to read: “not accredited by an accrediting organization recognized by the US Department of Education.”

The APA’s policy statement does prohibit members from citing fraudulent academic credentials in APA sponsored advertising and publications; yet, it does not prohibit the practice in other forums which may be related directly to the polygraph field. For instance, it does not extend to the use of diploma mill degrees in testimonial situations, whether for criminal, civil or administrative matters. In addition, though the APA neither “endorses” nor “recognizes” diploma mill degrees, it inadvertently facilitates claims made regarding such degrees by linking its web site to the sites of members who advertise polygraph services along with a conspicuous display of an advanced “degree.”

The APA can easily and effectively remedy the shortcomings in its current policy

statement. This can be done by incorporating in its Bylaws or Code of Ethics an expression similar to the following:

A member of the APA shall not in his or her polygraph-related presentations, publications, advertising, testimony or other aspects of professional practice claim to possess an academic degree, or the title associated with said degree, unless that person has in fact, been awarded that degree from an institution that is:

(a) Accredited to grant that degree by a national, regional or professional accrediting agency recognized by the United States Department of Education.

(b) A school, institute, college or university chartered outside the United States, that is approved by the Ministry of Education or similar government agency of that country to grant the degree. If the academic degree or title is to be used in the United States, it must be validated by a credential evaluation service that is a member of the National Association of Credential Evaluation Services as equivalent to the baccalaureate or post-baccalaureate degree conferred by a regionally accredited college or university in the United States. The provisions of this section shall not apply to persons who are residents of countries outside of the United States who are invited participants/speakers or attendees at meetings of the APA.

In this recommended revision it can be seen that all academic degrees, whether at the bachelor's, master's or doctoral level, or their equivalent, will be covered. All claims of holding a bachelor's degree used by a member to advance to full membership need to be evaluated and verified by evaluation of an official transcript. Applications for membership by those whose highest academic degree is from an American institution not accredited by an accreditation agency recognized by the DOE will be rejected. Due to the number of off shore diploma mills, applications for membership by U.S. citizens who are graduates of foreign universities will be required to provide an evaluation of their degree by a credential evaluation organization that is a member of National Association of Credential Evaluation Services. These agencies charge a reasonable fee for their services but their activities ensure that degrees earned from non-U.S. programs are indeed equivalent to those from an accredited U.S. program.

One of the most important and perhaps most difficult tasks that a professional organization must address is the identification and adjudication of unethical conduct, and implementation of actions to deal with such conduct by its members. These issues are among those which are of greatest concern to the membership of the APA (Horvath, 1995; 2007). Given the concerns of the membership and the importance of the issue as reflected by the actions of the federal and state governments, it is time for the polygraph community in general, and the APA specifically, as the world's largest and most influential association of examiners, to take appropriate action on the educational credentialing problem. This would be a significant advance forward in the interest of all who desire to enhance the field.

References

- American Polygraph Association (2004). *Academic resolution*. Retrieved June 1, 2005 from <http://www.polygraph.org/academicresolution.htm>
- American Polygraph Association (2005). *APA Bylaws*. Retrieved May 7, 2005 from <http://www.polygraph.org/bylaws.htm>
- Ashwood University (2005). *Get accredited life experience online degrees for what you already know*. Retrieved May, 9 2005 from <http://www.ashwooduniversity.net/>
- Bartlett, T. & Smallwood, S. (2004). Psst, wanna buy a Ph.D.? *Chronicle of Higher Education* Vol 1 N.42
- Cramer, R. J. (2004). *Diploma Mills*. Washington, D.C., US General Accounting Office
- Columbus University (2000). *Columbus University*. Retrieved May 7 2005 from www.ColumbusU.com 2000
- Ezell, A. & Bear, J. (2005). *Degree mills: The billion dollar industry that has sold over a million fake diplomas*. Amherst, NY, Prometheus Books
- Florida State Senate. (2006). *Making false claims of academic degree or title*. Retrieved July 9 2007 from <http://www.flsenate.gov/Welcome/index.cfm?CFID=45297364&CFTOKEN=97090285>
- Holten, N. G. & Lamar, L. L. (1991). *The Criminal Courts*. New York, McGraw-Hill, Inc.
- Horvath, F. (1995, August). *Professionalism in polygraphy: Organizational membership and its relationship to views, attitudes and practices in the field*. Presented at the American Polygraph Association, Las Vegas, NV.
- Horvath, F. (2007, August). *Polygraphy and Polygraphists—1996-2006: A Decade of Change or Only a Change in Decades*. Presented at the American Polygraph Association, New Orleans, LA
- Maschke, G.W. (2003). *Polygraph Operator "Dr." Edward I. Gelb Exposed as a Phony Ph.D.* Retrieved March 2005 from <http://antipolygraph.org/articles/article-036.shtml>
- Maschke, G.W. (2005). *James Allan Matte*. Retrieved July 9, 2007 from <http://antipolygraph.org/cgi-bin/forums/YaBB.pl?board=Policy;action=display;num=1117728910>
- Matte, J. A. (2005). *Matte Polygraph Service*. Retrieved May 15, 2005 from <http://www.mattepolygraph.com/matte/cpu.html>
- McGavock, J. (2006). Dr. Chris Gugas: The early pioneer in our field. *Truthseekers*.
- Merrill, R. L. (1999). *Accrediting distance degrees*. Retrieved from www.degreefinders.com/accreditation.html on June 30, 2000
- Midkiff, C. R. (2005). More mountebanks. In: Saferstein, R. *Forensic Science Handbook*, Volume II (pp. 35 – 80). Upper Saddle River, N.J.: Pearson-Prentice Hall, 35-80.
- National Center for Educational Statistics. (2002). Retrieved July 18, 2004 from <http://nces.ed.gov/programs/coe/2002/analyses/nontraditional/sa05.sp>

- Starrs, J. E. (2005). Mountebanks among forensic scientists. In: Saferstein, R. (Ed.) *Forensic Science Handbook*, Volume II (pp. 1 -34). Upper Saddle River, N.J.: Pearson-Prentice Hall.
- State of California (2005). *Columbia Pacific University ordered to close permanently*. Retrieved June 1, 2005 from <http://www.bppve.ca.gov/> on
- US Department of Education (2005). *Accreditation*. Retrieved May 7, 2005 from http://www.ed.gov/admins/finaid/accred/accreditation_pg8.html#det
- US Department of Labor (2005). *Position Announcement*. Retrieved from <http://jobsearch.usajobs.opm.gov/getjob.asp?Job1K=30562184&brd=3876&AVSDM=2005>
- Washington State. (2007). *Legislature passes phony academic degree measure*. Retrieved July 9 2007 from <http://src.leg.wa.gov/media2006/2006archives/Schoesler/schoeslerPhonyDegree030806.htm>

Improving the Detection of Physical Countermeasures With Chair Sensors

Jack Ogilvie and Donnie W. Dutton

Abstract

We conducted a reanalysis of unpublished data collected by V. Cholan Kopparumsolan to investigate whether specialized sensors would improve the performance of polygraph examiners in the detection of covert physical countermeasures. Five blind scorers evaluated the physiological data in two conditions. In the first condition they looked at 68 conventional polygraph cases for indications of countermeasures. In the second condition at least two months later they saw the same charts, this time with a channel that displayed data from a sensor designed to detect physical countermeasures. The addition of the countermeasure sensors significantly improved examiners' performance in the detection of physical countermeasures. The presence of the sensor information did not affect the scores or countermeasure ratings of examinees that had not been programmed to perform countermeasures.

keywords: countermeasures, polygraph, sensors

Countermeasures have been a pervasive and persistent challenge in the psychophysiological detection of deception. The idea that a deceptive individual could avoid detection by the polygraph is not new. Even in the earliest exploration of pre-polygraph instruments for detecting deception, examinees were already using strategies in an attempt to defeat them (Benussi, 1914). As the modern polygraph was becoming established in the US during and just after World War II there was no shortage of published advice on how to fool it (Blakemore, 1953; Masserman & Jacques, 1952; Stewart, 1941). In more recent years the advent of the Internet has made countermeasure information easily obtainable.

Though the information proffered on the Internet may be more detailed than in times past, the principal strategy recommended to potential examinees has not changed significantly for more than 50 years:

Simply induce reactions on non-relevant questions. The underlying premise is that the manufactured reactions will confound the examiner's ability to interpret reactions on the relevant questions. If the examiner sees reactions to other questions, especially if they are large reactions, he might discount or misinterpret real reactions. Creating reactions is a fairly simple process for most examinees, leading to its popularity as a countermeasure approach.

By necessity polygraph examiners have been forced to pay attention to the effect of Internet advice on their examinees, and as a consequence have come to understand countermeasures better. This has led to methods to detect, deter and defeat them. From an operational perspective, there are four general categories of countermeasures: pharmacological/chemical, behavioral, mental and physical (Krapohl, 1996). Each of these approaches entails particular strategies, but

Acknowledgements:

We are especially grateful to V. Cholan Kopparumsolan, who graciously allowed us to use the polygraph charts he had collected in an unpublished study he conducted at Michigan State University, and whose generous donation of charts and background made this project possible. We also relied on the dedication of the scorers: Rick Kurtz, John Fyffe, Det. L. Johnson, Bill Clifford, and Billy Wingo. We'd also like to thank Rose Swinford for managing the scoring data, and Dr. Stuart Senter for his helpful comments and suggestions to an earlier draft of this paper. The views expressed in this article are those of the authors, and do not necessarily represent those of the US Government, Department of Defense, the Kentucky State Police, or the Phoenix Police Department. Request for reprints should be directed to the first author at jack.ogilvie@phoenix.gov.

all share the common aim of causing a polygraph decision error. The last category, physical countermeasures is the focus of the present paper.

To appreciate how physical countermeasures might be used, it is important to examine how polygraph decisions are made. In modern polygraphy there are three principal types of test questions. First are the relevant questions, which cover the topic(s) that brought the person to the examination. Second, there are irrelevant questions, which are neutral questions added for technical purposes. Last, there are comparison questions. Comparison questions are included as benchmarks, against which reactions to relevant questions are gauged.

In an oversimplification of the actual process, if the greatest reactions are on the relevant questions, the report of the examination is Deception Indicated (DI). Conversely, if the reactions to the comparison questions are the greatest, the results are No Deception Indicated (NDI). Equal reactivity to both categories of questions produces a result of Inconclusive or No Opinion (NO).

Because comparison question techniques (CQT) rely on differential reactions to relevant and comparison questions, any countermeasure that augments the reactions to comparison question or diminishes reactions to relevant questions could be useful to a deceptive examinee attempting to evade detection of his lie. Strategies that increase or decrease reactivity to both types of question simultaneously, such as in the use of drugs or biofeedback, do not lead to a false negative error (calling a deceptive examinee truthful) (Honts, 1987). If an effective method exists for dampening reactions only to relevant questions it has not yet appeared in the literature.

Self-initiating reactions to comparison questions, a second approach, could be effective under the right conditions. Honts and his collaborators (Honts, 1987; Honts, Amato & Gordon, 2001; Honts & Hodes, 1983; Honts, Hodes & Raskin, 1985; Honts, Raskin & Kircher, 1983; Honts, Raskin & Kircher, 1987; Honts, Raskin, Kircher, & Hodes, 1988) have conducted a very thorough investigation into what makes for successful

countermeasures. They found that, in order to make physical countermeasures viable, the examinee had to be made aware of the principles underlying the CQT, advised on how to identify comparison questions, instructed in methods for manufacturing reactions, and given feedback from a polygraph examiner or psychophysicologist while being recorded with a polygraph. Their research determined that this last step was essential. Without real-time feedback, countermeasures were not effective for producing a false negative decision. Despite an overabundance of countermeasure instruction guides, the general unavailability of countermeasure training with feedback severely limits the number of successful countermeasure users.

It is interesting to note that recent findings from Honts and Alloway (2007) indicate that the countermeasure instruction proffered on the Internet can affect the polygraph outcomes of truthful examinees who employ them. Their data revealed a clear shift of polygraph scores of truthful examinees toward the deceptive direction when they used countermeasure strategies available on the Internet. Deceptive examinees using the same methods were not successful in moving their polygraph scores in a positive direction. In other words, the Internet advice only hurt truthful examinees. The Honts and Alloway finding regarding the ineffectiveness of countermeasure instruction for deceptive examinees would be predicted from earlier studies, but the negative impact on truthful examinees was unexpected.

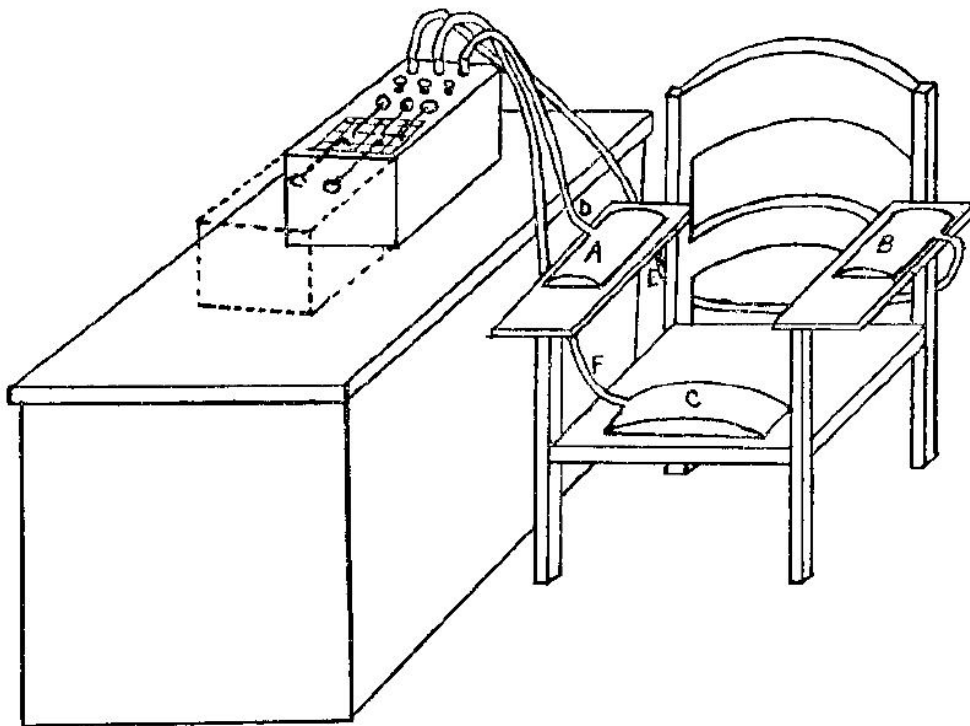
Detection of physical countermeasures by polygraph examiners without the aid of special sensors has proven problematic (Honts, 1984; Honts & Hodes, 1983; Honts, Hodes & Raskin, 1985). Examiners do not perform above chance levels when relying on only the conventional polygraph channels to uncover countermeasures. Detection of physical countermeasures is important for a number of reasons. First, they allow for the examiner to gauge the degree of examinee cooperation. Second, detection of countermeasures allows one to determine how much reliance to place on the physiological data. Finally, the ability to detect physical countermeasures provides greater evidence with which to defend the polygraph results.

The inability of examiners to identify physical countermeasures limits confidence that can be placed on truthful polygraph decisions. Any method that improves examiner abilities in this regard would be welcomed.

Reid (1945) published the first design for a device to detect covert physical movements during polygraph testing (Figure 1). Air bladders were placed in the seat and arms of a polygraph test chair, and changes in the bladder's air pressure were communicated to the polygraph and recorded on the strip chart. Reid suggested that detection of movements could be used to gauge the level of

the examinee's cooperation. His was not the only technical approach to be suggested in the detection of physical countermeasures. Since Reid's time, methods have included electromyography (Honts et al., 1987) and strain gauges mounted near the rear leg of the test chair (Stephenson & Barry, 1986). Most commercial systems in existence have relied on pneumatic or electric sensors attached to or imbedded in the test chair. All approaches attempt to reveal tactical movements by examinees that might influence the physiological data recorded during polygraph testing.

Figure 1. Design for a system for detecting concealed muscular movements. Letters A and B denote bladders on which examinees place their arms, and C is a bladder positioned below the examinee's thighs. From Reid (1945). Reprinted by special permission of Northwestern University School of Law, *The Journal of Criminal Law and Criminology*.



Evidence for the efficacy of these sensors has been inconsistent (Abrams & Davidson, 1988; Murray, 1989; Ohnishi, Tanaka, & Matsuno, 1968; Stephenson & Barry, 1986; also see Honts, 1987). A general rule, as one might expect, is that the value of the sensor depends on whether the examinee used a physical countermeasure that produces a signal in the sensor. Because examinees may try any number of countermeasure strategies, only a subset of those methods may be detected by the sensors. Even among physical countermeasure strategies, only those executed within the range of the sensor would be detected. A movement sensor-approach to the detection of mental countermeasures, pharmacological countermeasures, or behavioral countermeasures should be ineffective, and other means must be brought to bear against these strategies.

The present research had the limited goal of assessing whether physical countermeasures that are advocated on the Internet can be detected with a commercially available array of sensors. Using a sample collected by Cholan Kopparumsolan in 2002, we evaluated the contribution of the Lafayette activity sensors to the detection of physical countermeasures.

Methodology

Subjects

A total of 96 subjects were recruited from an undergraduate introductory course in criminal justice at Michigan State University in 2002. All had been offered extra credit for volunteering for the study, and told that they could earn a small cash reward at the end of the study. Half of the volunteers were female.

Polygraph Examiner

One polygraph examiner conducted all of the examinations. He received basic and advanced training at an American Polygraph Association (APA) accredited polygraph school, and had more than three years of field experience conducting examinations for both criminal and security applications in Singapore.

Blind Evaluators

Six law enforcement polygraph examiners were recruited to evaluate the polygraph charts. Three were on staff with the Phoenix Police Department, and three were examiners with the Kentucky State Police. One evaluator from the Phoenix Police Department did not complete all of the chart analyses by the close of the study, and his data were excluded from the project. All evaluators received their polygraph training at schools accredited by the APA. The evaluators were not provided any information other than the polygraph charts.

Apparatus

All subjects were seated in a polygraph test chair which was outfitted with motion sensors. The Lafayette 76875S Activity Sensors are designed to detect an examinee's physical movements during testing (see Figure 2.) The seat and arm sensors are urethane air bladders with nylon covers. The sensors for the subject's feet were bladders built into a set of feet plates. Changes in the subject's posture along with increases and decreases in muscle tension cause changes in the air pressure in the affected bladders. The pressure waves are conducted via tubing to a central point where the mechanical energy is converted to electronic signals, which in turn, are represented as a moving line on a computer screen. The line does not identify which of the bladders was affected by the examinee's movement.

A Lafayette LX 3000 computer polygraph (Lafayette Instrument Company, Lafayette, IN) was used to test all of the subjects in this study. During testing the polygraph recorded and displayed breathing patterns from the examinee's thoracic and abdominal areas, electrodermal responses from sensors placed on two fingers of the subject's left hand, and cardiovascular activity using an occlusion cuff placed about the subject's upper left arm. Cuff pressure was maintained between 60 to 70 mm Hg.

Design

The subjects were divided by gender and then randomly assigned to six treatment

Figure 2. Lafayette 76875S Activity Sensors configured with polygraph testing chair. Photo courtesy of the Lafayette Instrument Company, Inc.



groups: innocent, guilty control, guilty practice cognitive, guilty practice physical, guilty experience cognitive, and guilty experience physical countermeasures. Each of the six treatment groups consisted of eight males and eight females. For the present purposes we used only three groups: the innocent, the guilty control, and a single combined group (labeled here as the physical countermeasure group) made up of the guilty practice physical and guilty experience physical groups. The scorers in the present study analyzed all of the cases, but because of the focus on physical countermeasures, only the innocent, guilty control, and physical countermeasure groups are considered here.

Procedure

As stated earlier there were six treatment groups. Only those relevant to the current analysis are reported.

Innocent Group

Subjects assigned to the innocent group did not commit the mock crime,

received no countermeasure training or instructions, and were not told to use countermeasures during their polygraph tests.

Guilty Control Group

In this group subjects committed the mock crime, but were not given countermeasure training or instructions nor told to engage in countermeasures during their polygraph tests.

Physical Countermeasure Group

In the physical countermeasure group, the subjects committed the mock crime, received (at a minimum) training in countermeasures as advocated on anti-polygraph.org, and were instructed to try to “beat” their polygraph testing using the strategies offered at the anti-polygraph website.

Instructions to Innocent Subjects

Innocent subjects listened to the following taped instructions:

“You have been randomly assigned to participate in this study as an innocent person. Your task, once this tape is completed, is to leave the building and go for a short walk returning here in approximately 15 minutes. During the time you are out walking, there will be a crime committed, but you will have no knowledge of what transpired.”

“Within the next four days, you will be given a lie detection test as a possible suspect in the crime because of you being in the area. You are to speak to no one about your participation in this study and to appear, as you are, innocent. If you pass the test, that is, if the polygraph examination shows that you are innocent, you will earn a small cash reward in addition to your course credit. Good luck, now carry out your instructions.”

Instructions to Guilty Subjects

Subjects in the guilty control and the physical countermeasure group were given written instructions on the mock crime they were to commit. The instructions were as follows:

“You have been randomly assigned to be a guilty subject. Your task, if you choose to participate, is to proceed from this location to the elevator lobby in Baker Hall and take the elevator to the fifth floor. Upon your arrival on the fifth floor proceed to room 541, which is the mailroom of the School of Criminal Justice. Once you are there, go inside the mailroom and look for mail slots under the column marked 6 and look for a mail slot marked with an orange tag bearing the name Professor Frank Horvath. Thereafter, carry out a quick systematic search of all the mail in the said mail slot to locate a business size ‘airmail’ envelope with red and blue markings around the edge and a large ‘X’ marked across on each side. Once you locate the said envelope, take it out of the mail slot and hide it on your person and quickly leave the mailroom. Should anyone walk into the room while you are carrying out the theft, make whatever excuse you think is necessary and continue on your way. After the commission of the theft, quickly leave the floor and return here by whatever means or route you choose. Remember, you have only 15 minutes to complete this portion of your assignment, so do not waste time. You will be

given additional instructions when you return.”

“You will be given a polygraph examination (lie detection test) within the next four days. It is very important that you remember that during this examination you are not to tell the person conducting the examination under any circumstance or any reason whatsoever what you have done. Any questions related to the envelope, where it was taken from or contents of the envelope must not be answered truthfully. Deny all involvement with the ‘airmail’ envelope. If you can successfully accomplish this task, that is, if the polygraph examination shows you to be truthful, you will be rewarded with the contents of the ‘airmail’ envelope you took.”

“Lastly, it is very important that you do not speak with other students about your participation in this study and that you appear truthful at all times. Good luck, now carry out your instructions.”

Following the mock crime, guilty subjects received instructions to open the “airmail” envelope where they found two \$5 bills. The bills were turned over to one of the trainers. The money would be dispensed to those examinees who obtained a decision of NDI based on the manual scoring of the testing examiner. Subjects were scheduled for a polygraph examination from 1 to 4 days after the programming.

Countermeasure Training

Subjects received a paper copy of the internet publication *The Lie Behind the Lie Detector* (Maschke & Scalabrini, 2002) downloaded from www.anti-polygraph.org website, and studied the instructions for 20 minutes. The publication discusses in detail the polygraph instrumentation, techniques, protocols, and methods of analysis. It also encourages and informs readers how to covertly augment reactions to comparison questions in order to defeat the examination. After the subjects reviewed the material, trainers administered a 10-item written quiz to assess whether the subjects understood the countermeasure instructions. They then provided the correct answers to the quiz items, and answered any of the subjects’ questions about the countermeasure

instructions. A subset of the countermeasure group received real time feedback on their countermeasure efforts from an experienced polygraph examiner during a practice test with the polygraph. The polygraph examiner providing the feedback was trained at an APA school, and was a member of the Michigan State Police. The physical countermeasure was consistent with the advice given in *The Lie Behind the Lie Detector*: toe curling during the comparison questions. Subjects were all told to practice their countermeasure methods during their free time before their polygraph examinations.

The Polygraph Examination

V. Cholan Kopparumsolan was the testing polygraph examiner for all examinations. Though he knew the base rates and the types of countermeasure instruction, he was blind to ground truth and countermeasure programming for individual subjects.

The test questions and sequence was as follows.

Q1 Irrelevant - Are you now in Indiana? Answer: No

Q2 Overall Truth - Have you told me the absolute and complete truth about every single thing we discussed here today? Answer: Yes

Q3 Secondary Relevant - Are you now lying to me in any way about the missing airmail envelope from Dr. Horvath's office at 122 Baker Hall? Answer: No

Q4 Irrelevant - Are you now in Michigan? Answer: Yes

Q5 Relevant - Did you remove that airmail envelope from Dr. Horvath's office? Answer: No

Q6 Comparison - Not connected with this case, did you ever take something that did not belong to you, even one time in your entire life? Answer: No

Q7 Irrelevant - Are you now in the United States? Answer: Yes

Q8 Relevant - Did you remove five dollars from an airmail envelope taken from Dr. Horvath's office? Answer: No

Q10 Comparison - Not connected with this case, other than what you told me, have you ever told an important lie, even one time in your entire life? Answer: No

Q13 Irrelevant - Are you now in Canada? Answer No

The first test was a Silent Stimulation Test (SST) in which the subjects were instructed to listen but not answer the test questions (Horvath & Reid, 1972). The SST was followed by a Card Stimulation Test, sometimes called an Acquaintance Test. The third and fourth tests in the series were Verbal Answer Tests. The fifth and last test was the Yes Test, a countermeasure detection method also described in *Truth and Deception* (Reid & Inbau, 1977). In the present analysis only charts 1, 3 and 4 were used. The Yes Test had been excluded because it is designed to prompt countermeasure attempts in a form that are more easily discernable in the charts, and also because in field practice it is not regularly administered. The Card Stimulation Test was similarly excluded because it is a type of approach used by a minority of field examiners, and it too was not relevant to the study.

Blind Analyses

All of the cases were analyzed by the five scorers, but only 68 cases are reported here. This is the number of innocent, deceptive control, and physical countermeasure group, minus four cases which were lost due to file corruption. The 68 cases consisted of 15 innocent, 15 deceptive control, and 38 physical countermeasure cases.

The polygraph charts were printed twice: once with the motion sensor data removed and once with those data present. The scoring of each of these sets of charts was separated by at least two months. The case numbers and order were changed between the first and second evaluation by the examiners. The examiners separately scored the charts using 7-position numerical analysis and they also assessed the likelihood of countermeasures for each case on a five-point

continuum: certainly, probably, unsure, probably not, certainly not. These terms were converted to whole numbers from -2 to +2, respectively. See Appendix A for the score sheet provided to the examiners.

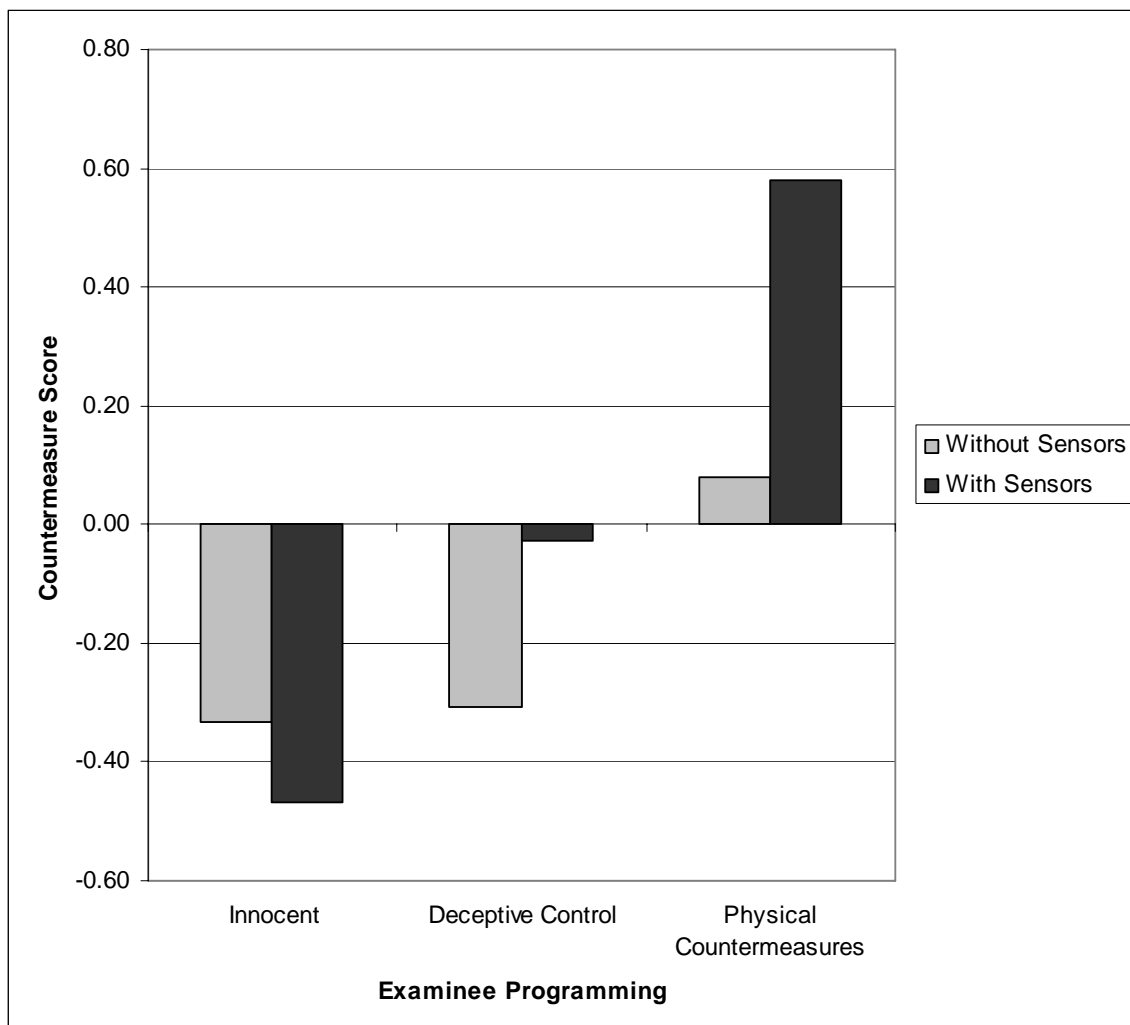
Data reduction

The countermeasure likelihood ratings for each case were averaged across evaluators for each of the three conditions; innocent, deceptive control and physical countermeasure groups. Because of multiple analyses, the Bonferroni correction was used (Miller, 1991). Calculation of alpha of .05 using the Bonferroni correction of α / n was .017 ($.05/3 = .017$).

Results

Using a two-tail t test there were no significant differences in the mean countermeasure scores for the innocent cases or deceptive control cases between the sensor and no-sensor conditions (innocent $t[14] = 0.65$, ns; deceptive control $t[14] = 1.20$, ns). Mean countermeasure scores for the physical countermeasure group between the sensor and no-sensor conditions were statistically different ($t[37] = 2.60$, $p < 0.017$). In other words, the addition of the motion sensor data on the charts did not cause the polygraph examiners to increase their estimates of concealed movements in any condition except when examinees were programmed to use physical countermeasures. See Figure 3.

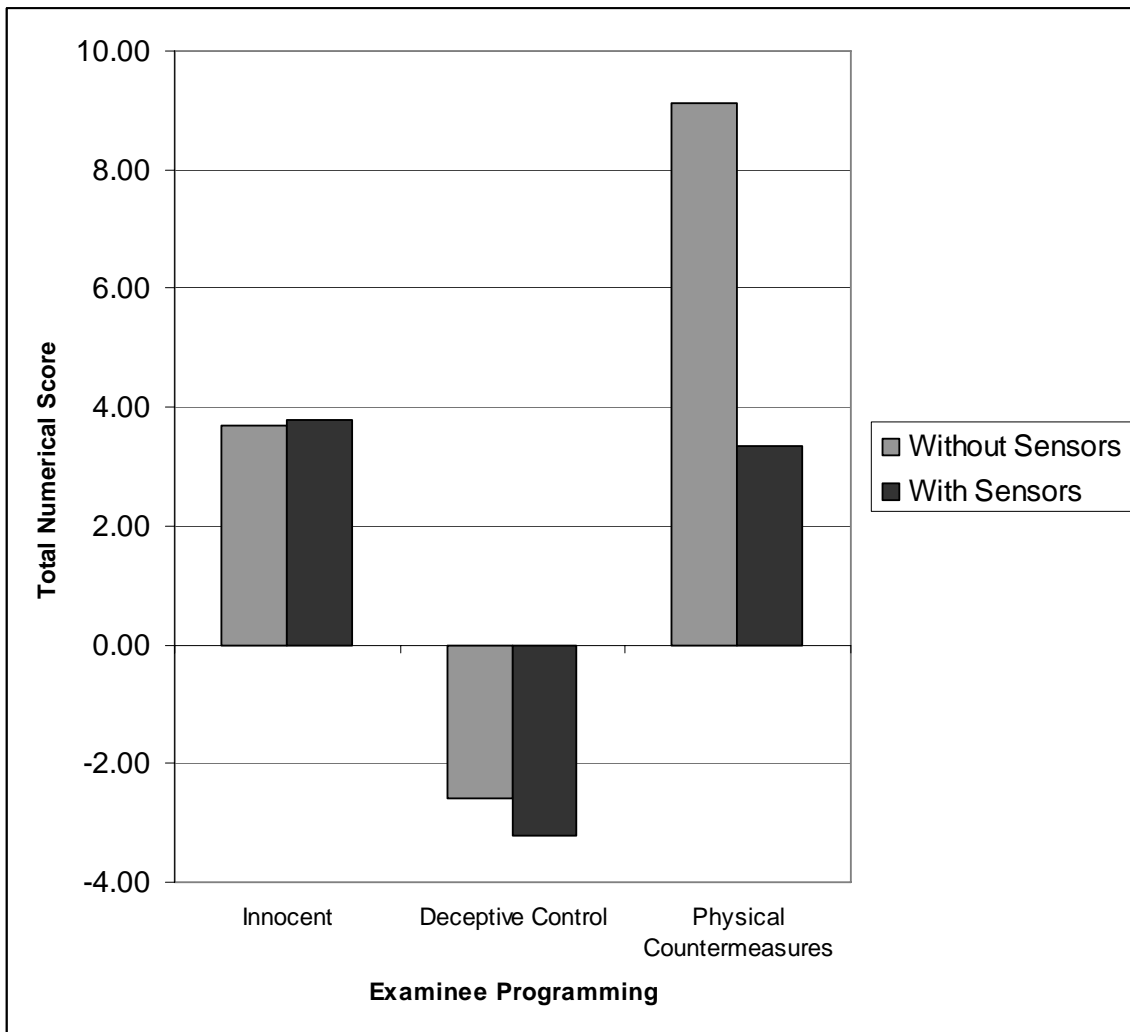
Figure 3. Average countermeasure scores when sensor data are either present or absent.



The effect on numerical scores was similar. The inclusion of the sensor data did not influence the average total numerical scores for innocent subjects ($t[14] = 0.05$, ns;) nor deceptive control subjects ($t[14] = 0.76$, ns). However, the addition of the sensor data on the polygraph charts did lower the numerical scores of the deceptive countermeasuring

subjects significantly ($t[37] = 3.79$, $p < 0.017$). Said differently, the sensor data only affected the numerical scores of deceptive countermeasuring subjects: countermeasuring subjects received significantly lower numerical scores when the sensor data were displayed on the charts. See Figure 4.

Figure 4. Average numerical scores when sensor data are either present or absent.



Discussion

The present analysis points to potential benefits for adding a data channel on field polygraphs. Blind scorers were better able to detect physical countermeasures when the sensor data were viewed along with the

traditional polygraph channels than when the sensor data were not displayed. Moreover, the sensor data mitigated the positive numerical scores that had been given to the countermeasure cases in the no-sensor condition though the scorers had been given no instructions in that regard. Moving the numerical scores of deceptive counter-

measurers in the negative direction can reduce false negative and increase true positive results. Finally, the sensor data did not affect either the numerical scores or the countermeasure scores of examinees who were not programmed to use countermeasures. Sensor data affected the countermeasure group only.

Limitations

1. Analog studies can be criticized for lacking the level of motivation and jeopardy that might take place in real world settings. The generalizeability of the present findings has

not been established. Our data may over- or under-estimate the ability of polygraph examiners to detect physical countermeasures in the field. Whether high motivation and jeopardy improves or diminishes countermeasure detection is unknown.

2. The countermeasures sensors used here were more extensive than the conventional seat cushion currently found in wide distribution in the field. Whether physical countermeasures of the type used in this study can be detected with the seat cushions alone was not addressed.

Appendix A. Examiner Score Sheet

File # _____ Scorer _____ Date _____

Chart 1

Question #	3	5	8			
Pneumograph						
Electrodermal						
Cardiograph						

Chart 2

Question #	3	5	8			
Pneumograph						
Electrodermal						
Cardiograph						

Chart 3

Question #	3	5	8			
Pneumograph						
Electrodermal						
Cardiograph						

Totals → _____

Decision _____ DI _____ NDI _____ Inc

Countermeasures

Check one → _____
 Certainly Probably Unsure Probably not Certainly not

Comments

References

- Abrams, S., and Davidson, M. (1988). Counter-countermeasures in polygraph testing. *Polygraph*, 17(1), 16-20.
- Benussi, V. (1914). Die atmungssymptome der lüge [The respiratory symptoms of lying]. *Archiv fuer die Gesamte Psychologie*, 31, 244-273. In 1975 an English translation appeared in *Polygraph*, 4 (1), 52-76.
- Blakemore, F.T. (1953, August). Can criminals beat the lie detector? *Science and Mechanics*, 80-84.
- Honts, C.R. (1984). Countermeasures and the physiological detection of deception. *Psychophysiology*, 23, 566-567 (Abstract).
- Honts, C.R. (1987). Interpreting research on polygraph countermeasures. *Journal of Police Science and Administration*, 15(3), 204-209.
- Honts, C.R., and Alloway, W.R. (2007). Information does not affect the validity of a comparison question test. *Legal and Criminological Psychology*, 12(2), 311-320.
- Honts, C.R., Amato, S.L. and Gordon, A.K. (2001). Effects of spontaneous countermeasures used against the comparison question test. *Polygraph*, 30(1), 1-9.
- Honts, C.R., and Hodes, R.L. (1983). The detection of physical countermeasures. *Polygraph*, 12(1), 7-17.
- Honts, C.R., Hodes, R.L., and Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, 70(1), 177-187.
- Honts, C.R., Raskin, D.C., and Kircher, J.C. (1983). Detection of deception: Effectiveness of physical countermeasures under high motivation conditions. *Psychophysiology*, 20, 446-447.
- Honts, C.R., Raskin, D.C., and Kircher, J.C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*, 1(3), 241-247.
- Honts, C.R., Raskin, D.C., Kircher, J.C. and Hodes, R.L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. *Journal of Police Science and Administration*, 16(2), 91-94.
- Krapohl, D.J. (1996). A taxonomy of polygraph countermeasures. *Polygraph*, 25(1), 35-56.
- Maschke, G.W., and Scalabrini, G.J. (2002). *The Lie Behind the Lie Detector*. Available at <https://antipolygraph.org/pubs.shtml>.
- Masserman, J.H., and Jacques, M.G. (1952, April 19). Do lie detectors lie? *The Nation*, 368-369.
- Miller, R. G. Jr. (1991) *Simultaneous Statistical Inference*. New York: Springer-Verlag.
- Murray, K.E. (1989). Movement recording chair: A necessity? *Polygraph*, 18(1), 15-23.
- Ohnishi, K., Tanaka, Y., and Matsuno, K. (1968). Study of a new index for lie detector: Re-examination of body movement. *Polygraph Reports*, 53, 12-16.

- Reid, J.E. (1945). Simulated blood pressure responses in lie detector tests and a method for their detection. *American Journal of Police Science*, 36(1), 201-214.
- Reid, J.E., and Inbau, F.E. (1977), *Truth and deception: The polygraph ("lie detector") technique* (2nd ed.). Williams & Wilkins: Baltimore, MD.
- Stephenson, M., and Barry, G. (1986). Use of a motion chair in the detection of physical countermeasures. *Canadian Association of Police Polygraphists Newsletter*. Reprinted in *Polygraph*, 17(1), 21-27.
- Stewart, W.S. (1941). How to beat the lie detector. *Esquire*, 16(5), 35, 158, 160.

Exploration of a Two-Stage Approach

Stuart M. Senter and Andrew B. Dollins

Abstract

Modifications to the psychophysiological detection of deception Zone Comparison Test decision rule process were explored in an attempt to increase accuracy. Two two-stage models for producing decisions following conventional physiological data scoring were proposed. Total score cutoff values were used to produce decisions during the 3T (Total) stage, with totals of -6 or less producing decisions of deception indicated, totals of $+6$ or greater producing decisions of no deception indicated, and totals between these cutoffs producing a no opinion (NO) decision. During the 3S (Spot) stage, scores assigned to individual question pairs were evaluated to produce decisions, in addition to total cutoffs. The assigned scores were totaled for each of three relevant-comparison question pairs, to produce a spot score for each question pair. If the total score was -6 or less or if any spot score was -3 or lower, then a decision of deception indicated was produced. If the total score was $+6$ or higher and if all three spot scores were $+1$ or greater, a decision of no deception indicated was generated. If neither of these criteria were met, a decision of NO was rendered. The two-stage 3T3S model consisted of the 3T (i.e., total cutoff) stage, followed by the 3S (i.e., spot score) stage to resolve NO decisions. The stages were implemented in reverse order in the 3S3T model. Data from three laboratory and two field studies were evaluated, (encompassing 1420 chart assessments) to compare accuracy of the two-stage models with the accuracy of models using only one of the two stages (i.e., 3T or 3S). The 3T3S and 3S3T models resulted in 8.0% more correct decisions, 10.4% fewer NO decisions, and 2.4% more incorrect decisions than the 3T and 3S models.

Introduction

Payne, Bettman, and Johnson (1988) empirically showed that the use of different decision heuristics produced different levels of effectiveness, in terms of accuracy and effort expenditure, depending upon the context of the choice situation. For example, in cases of extreme time pressure, decision models that involved only a small number of steps outperformed those models that were more intensive computationally. Based on work such as this, Payne, Bettman, and Johnson (1990) developed the notion of the adaptive decision maker, indicating that individuals have the ability to adapt and shift the strategies they use in order to produce effective decisions given the environment in which decisions are made. To further examine

the adaptive decision maker hypothesis, a set of polygraph decision models were evaluated to determine their overall effectiveness, and their effectiveness in evaluating deceptive and truthful cases.

In previous work we explored the effectiveness of decision rules used for the psychophysiological detection of deception (PDD) Zone Comparison Test (ZCT). Senter, Dollins, and Krapohl (2004) and Senter and Dollins (2004, 2008) have all shown, using data collected in the laboratory, that decision accuracy for the ZCT is heavily influenced by the decision processes that follow the assignment of scores to relevant-comparison question pairs. It was found that the proportion of correct decisions could be increased by using a three-or-five question

Acknowledgments

This research was funded by the Defense Academy for Credibility Assessment, Fort Jackson, South Carolina, as project DoDPI01-P-0007. The views expressed in this report are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

series rule, relative to the standard three question series rule. According to the three-or-five question series rule, three question series are collected and evaluated. If there is insufficient data to produce a decision of deception indicated (DI) or no deception indicated (NDI), then two additional question series are collected and combined with the initial three question series to produce a decision. Senter et al. (2004), Senter and Dollins (2004), and Senter and Dollins (2005) found 12%, 8%, and 12% increases in the proportion of correct decisions, respectively, with the three-or-five question series rule relative to the three question series rule. These increases were mostly due to the resolution of no opinion (NO) decisions into correct decisions.

An obstacle to the use of the three-or-five question series rule in the Federal Government is the lack of field-based research to demonstrate that the decision accuracy increase found in laboratory-based studies would generalize to the field. This is a problem that can be rectified through the collection of five question series in field settings. It would not be prudent to suggest policy changes without supporting empirical data. It may, however, be possible to improve decision accuracy without altering data collection procedures or policy.

For those not familiar with the ZCT polygraph procedure, two types of questions are used to evaluate examinee veracity [Department of Defense Polygraph Institute (DoDPI), 2001]. Relevant questions directly address the issue under consideration (e.g., Did you shoot that man?). Comparison questions contain content similar to the relevant question, but are separated from the issue under consideration, usually with a caveat or time bar (e.g., Prior to this year, did you ever harm anyone?). A ZCT question series usually contains three relevant-comparison question pairs. In theory, deceptive examinees show greater responses to the relevant questions, and truthful examinees show greater responses to the comparison questions. Data evaluation involves the examination of respiratory, electrodermal, and cardiovascular recordings. Responses are quantified by identifying specific reactions or features in each data channel. For a review of these features, see

Bell, Raskin, Honts, and Kircher (1999), and Swinford (1999).

PDD question series are evaluated by comparing the response to a relevant question to that of an adjacent comparison question for each data channel. If the response to the relevant question is larger than the response to the comparison question, a negative value is assigned to the question pair for that channel. If the response to the comparison question is larger, a positive value is assigned. Values can be assigned using a three-position scale (-1 to +1) or a seven-position scale (-3 to +3). The three-position scale is used to indicate a differential response, without evaluating the magnitude of the difference. The seven-position scale allows for the assignment of integers between -3 and +3 to indicate the magnitude of response differences. The assigned scores are then summed to produce a decision that the examinee is being truthful or deceptive.

Two prominent decision rules are currently used with the ZCT. One approach, published by investigators at the University of Utah and used by some polygraph training institutions, is the total cutoff rule whereby all assigned scores are totaled and a decision depends on whether the total value exceeds cutoff criteria (Bell et al., 1999). If the total is +6 or greater, a NDI decision is made. If the total is -6 or less, a DI decision is made. If neither criterion is met, a NO decision is rendered. A second approach, called the spot score rule, is taught by other training institutions and used with specific issue PDD examinations throughout the Federal Government. This decision rule also uses the -6 and +6 cutoffs, but has different criteria for producing decisions. The spot score rule requires the values for each relevant-comparison question pair (also known as a 'spot') be summed over question series repetitions to produce three 'spot scores'. A DI decision is made if the -6 criterion is met, or if the value of at least one question spot score is -3 or less. A NDI decision is made if the +6 criterion is met, and if the value of all three question spots is +1 or greater. If the criteria for producing a DI or NDI decision are not met, then a NO decision is made.

Previous work by Senter and Dollins (2008) showed the differential impact of these

two decision rules as a function of examinee veracity. Decision models that included only the total cutoff rule generated more correct decisions with truthful participants than with deceptive participants ($M = 78.4\%$ vs. 70.5% , respectively). Decision models that included the spot score rule resulted in more correct decisions with deceptive participants than with truthful participants ($M = 83.0\%$ vs. 61.4% , respectively). Senter and Dollins also showed that these rules could be used in combination to produce high levels of accuracy for both deceptive and truthful participants. Decision models that included both the total cutoff rule and the spot score rule produce high levels of accuracy for both deceptive and truthful participants ($M = 84.0\%$ and 76.9% , respectively). Thus, initial indications suggest that in a high deceptive base rate population, the spot score rule should be more effective, and in a low deceptive base rate population, the total cutoff rule should be more effective. However, the decision models that included both decision rules were only examined in conjunction with the three-or-five question series rule, a data collection procedure that has not been field tested or used by the Federal Government.

In this study we examine the accuracy of decision models that include combinations of the spot score rule and the total cutoff rule,

using the standard three question series. Four decision models were examined. The 3T model used only the total cutoff rule to produce decisions. The 3S model used only the spot score rule to produce decisions. The 3T3S model used the total cutoff rule as a first stage, and if a NO decision was produced, the spot score rule was used. The 3S3T model used the spot score rule as a first stage, and if a NO decision was generated, the total cutoff rule was used. These models are described in Table 1. A compilation of assigned scores from laboratory and field ZCT polygraph examinations were used to determine the performance of these four models.

Based on the work of Senter and Dollins (2008), we predicted the 3S model to produce the highest accuracy with deceptive participants, and the 3T model to produce the highest accuracy with truthful participants. We predicted that the 3T3S and 3S3T models would produce the highest accuracy rates for two reasons. Senter and Dollins found that models which included both the 3T and 3S components produced the highest accuracy. Also, Payne et al. showed that decision models that were more computationally intensive (i.e., included additional steps or stages to arrive at a decision) generated higher accuracy rates than simpler decision models when time pressure was not a factor.

Table 1

Decision Models

Model	Description
3T	total cutoff rule only
3S	spot score rule only
3T3S	total cutoff rule used first, if no opinion, spot score rule used
3S3T	spot score rule used first, if no opinion, total cutoff rule used

Method: Laboratory Data

Blind evaluators assigned values to physiological response data from three laboratory studies. The data are described in Senter et al. (2004; Study 1), and DoDPI Research Division Staff (2001; Study 2 and 3). Data set details are shown in Table 2. Physiological response data from the three laboratory studies were quantified by blind

evaluators and decisions were produced by applying the four decision models described in Table 1 with computer algorithms using the assigned scores. Data are first presented descriptively as a function of decision, and second inferentially with assigned values averaged across scorer to produce a single decision per case, so that statistical power will not be artificially inflated.

Table 2

Frequency of Observations Used in Laboratory Decision Model Evaluation

Study	N Participants		N Scorers	N Decisions
	Deceptive	Truthful		
1	50	50	5	500
2	16	16	3	96
3	16	16	7	224
Total	82	82	15	820

Results

The percentages of correct, incorrect, and NO decisions delivered by the four decision models are shown in Figure 1. The 3T model was more accurate for truthful examinees, the 3S model was more accurate for deceptive examinees, and the two-stage models were about equally accurate for truthful and deceptive examinees. Averaging across studies, the 3T model resulted in 71.5% correct, 2.7% incorrect, and 25.9% NO decisions for truthful examinees ($N_{\text{decisions}} = 820$). However, for deceptive examinees, the 3T model generated 60.0% correct, 8.8% incorrect, and 31.2% NO decisions ($N_{\text{decisions}} =$

410). In contrast, the 3S model produced 76.6% correct, 5.1% incorrect, and 18.3% NO decisions for deceptive examinees, and 54.9% correct, 13.9% incorrect, and 31.2% NO decisions for truthful examinees ($N_{\text{decisions}} = 410$).

The 3T3S and 3S3T models performed similarly for both deceptive and truthful examinees. The 3T3S model resulted in 75.6% correct, 8.8% incorrect, and 15.4% NO decisions for deceptive examinees, and 71.5% correct, 12.0% incorrect, and 16.6% NO decisions for truthful examinees. The 3S3T model generated 76.6% correct, 8.1% incorrect, and 15.4% NO decisions for

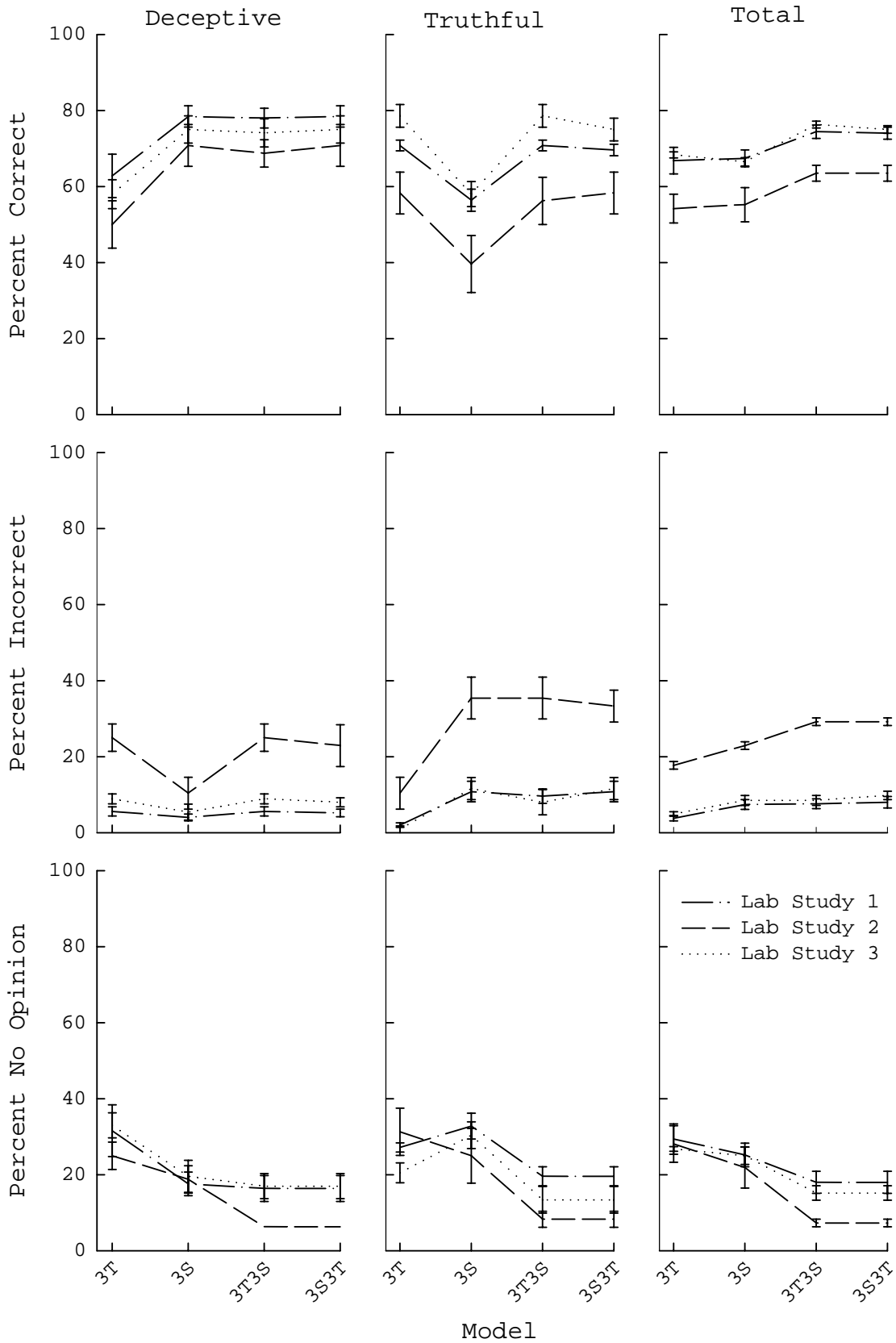


Figure 1. Mean (\pm SEM error bars) percentage of correct, incorrect, and no opinion decisions as a function of decision model and laboratory study.

deceptive examinees, and 69.5% correct, 13.9% incorrect, and 16.6% NO decisions for truthful examinees.

Overall, the 3T3S and 3S3T models delivered more correct decisions (73.7% and 73.1%, respectively) than the 3T and 3S models (65.7% and 65.7%, respectively). In addition, the 3T3S and 3S3T models generated fewer NO decisions (16.0% and 16.0%, respectively) than the 3T and 3S models (28.5% and 24.8%, respectively). However, the 3T3S and 3S3T models also resulted in more incorrect decisions (10.4%

and 11.0%, respectively), than the 3T and 3S models (5.7% and 9.5%, respectively).

Table 3 shows the percentage of agreement (cases where evaluators produced identical decisions) across the three laboratory studies. The percentage of agreement for the four decision models was very similar across all three studies, and there were no statistically significant differences in the proportion of agreement for the four models ($p > .05$).

Table 3
Average Percent of Scorer Agreement as a Function of Four Decision Models (Laboratory Data)

Model*	Study 1	Study 2	Study 3	Weighted Mean
3T	75	74	68	71
3S	78	72	68	72
3T3S	77	82	72	75
3S3T	77	80	72	74

Note. *See Table 1 for Model descriptions

While descriptive statistics are informative, we found it useful to provide some metric to discriminate between the effectiveness of the four models. A decision efficiency criterion (DEC) was calculated with an approach used by Kircher, Horowitz, and Raskin (1988) where, in essence, good decisions increase the index, poor decisions decrease the index, and moderate decisions produce no change. Participant veracity was coded as +1 for truthful participants and -1 for deceptive participants, and corresponding decisions were coded as -1 for DI, 0 for NO, and +1 for NDI deceptive. To reduce the

inflation of power produced by analyzing the data as function of decision, assigned scores for each participant were averaged across scorer for each study. Thus, for a given participant with three scorers, if total scores generated by the three scorers were +4, +9, and +5, respectively, these scores averaged to +6.0, which was the value used to produce decisions by the various models. Spot scores were also averaged in this way. In doing this, each case produced only a single decision.

Decisions were then used as the predictor variable and veracity as the criterion

variable in a regression model. Each decision model was calculated separately to determine the amount of variance accounted for by each model, again, as a function of individual study, and across studies. The resulting R^2 values were then compared across models. These results are displayed in Table 4. With the exception of the 3S model for study 2, the R^2 values were all statistically significant, indicating that a significant proportion of variance was captured by each model. However, none of the values were significantly different from each other, both in the context of individual studies, and when the values were combined across the three studies, indicating that the DEC did not differ among the studies.

Due to the relatively low power to detect differences afforded by the multiple regression approach, a second set of inferential analyses was conducted using a conversion of the raw DEC values. The -1, 0, +1 DEC system was used for coding decisions, and -1 and +1 for coding ground truth.

Correct decisions were coded as +1, incorrect decisions were coded as -1, and NO decisions were coded as 0. The values generated for all decisions were then added and divided by the total number of decisions. This system was applied to reflect both the strength of correct decisions and the weakness of incorrect decisions delivered by each model. This analysis was conducted for each laboratory study, and also collapsing across all studies. These values are displayed in Table 5.

A repeated measures analysis of variance (ANOVA) was calculated using the converted DEC raw values produced by each model as a dependent measure. ANOVAs were calculated for each study, and collapsing across studies. For the individual studies, no results reached statistical significance. However, collapsing across the three studies, the results were significant, $F(3,489) = 2.64, p < .05$. The contrast between the 3T, 3T3S, and 3S3T models and the 3S model was significant, $F(1,163) = 5.79, p < .05$, and accounted for 94.8% of the omnibus model effect variance (Keppel, 1991, p. 135).

Table 4

Decision Efficiency Criterion Values for Decisions and Participant Veracity

Study	Model			
	3T	3S	3T3S	3S3T
1	.613	.514	.561	.564
2	.230	.089	.113	.113
3	.554	.373	.486	.377
Weighted Total	.527	.404	.459	.440

Discussion

The increase in the proportion of correct decisions by the 3T3S and 3S3T models, relative to the 3T and 3S models is due primarily to the resolution of NO decisions. The resolution of NO decisions also resulted in an increased proportion of incorrect decisions for the 3T3S and 3S3T models relative to the 3T and 3S models. However, when the percentage of correct decisions is considered excluding NO decisions, the level of accuracy for all four

models is quite similar (3T = 92.0%, 3S = 87.4%, 3T3S = 87.7%, 3S3T = 86.9%). More cases were resolved using the two-stage models than the one-stage models (3T3S = 84.0%, 3S3T = 84.0%, 3T = 71.5%, 3S = 75.2%). It could, thus, be argued that the four models are equivalent in total accuracy, but differ in utility. It should, however, be noted that the 3T model is more accurate for truthful participants, and the 3S, 3T3S, and 3S3T models are more accurate for deceptive participants.

Table 5

Converted Decision Efficiency Criterion Raw Values as a Function of Decision Model and Study

Study	Model			
	3T	3S	3T3S	3S3T
1	.670	.620	.680	.680
2	.406	.281	.344	.344
3	.625	.531	.625	.563
Weighted Total	.610	.537	.604	.591

As Figure 1 shows, this pattern of results was consistent for all studies examined, which suggests a reliable impact for all four decision models. Keppel (1991, p. 76) argued that in lieu of statistical significance, replication may be considered the strongest indicator of a reliable effect. Figure 1 shows a clear replication of results generated by the decision models across the three studies.

The ANOVA planned comparison results indicate that while the 3T model produced fewer correct decisions relative to the 3T3S and 3S3T models, its low production of errors resulted in a high accuracy percentage when decisions were made. All three models showed superior performance relative to the 3S model. The DEC values reflected these trends, but did not show any statistically significant differences.

The application of the four decision models to the three sets of laboratory data clearly show that the 3T3S and 3S3T models reduce NO decisions and increase the total percentage of correct decisions relative to the 3S and 3T models. We felt it important to determine the extent to which these models would generalize to field data, and thus applied the same four decision models to confirmed field data.

Method: Field Data

As per the laboratory data, the four decision models were applied to assigned scores from four sets of confirmed field data.

Data set details are provided in Table 6. The assigned scores were also generated by blind scorers, and not by the original examiners. Data sets 1 and 2 were obtained from Blackwell (1999) and Krapohl, Dutton, and Ryan (2001), respectively. Unlike the laboratory data, there was not an equal number of deceptive and truthful cases across the two studies. On average, 57.5% of the field cases were from deceptive examinees. It is the experience of the authors that more deceptive than truthful examinees are tested in the field. The frequency of deceptive examinees in these field samples may, thus, be more representative of actual data than the laboratory data previously described.

Table 6

Frequency of Observations Used in Field Decision Model Evaluation

Study	N Participants		N Scorers	N Decisions
	Deceptive	Truthful		
1	65	35	3	300
2	50	50	3	300
Total	115	85	6	600

Results

Figure 2 shows the percentage of correct, incorrect, and NO decisions as a function of examinee veracity and study. The performance of the 3S model was highly dependent on whether it applied to truthful or deceptive examinees. This model produced 86.1% correct, 4.9% incorrect, and 9.0% NO decisions for deceptive examinees ($N_{\text{decisions}} = 345$), and 42.3% correct, 22.8% incorrect, and

34.9% NO decisions for truthful examinees ($N_{\text{decisions}} = 255$). The 3T model was less effective than the 3S model at detecting deceptive examinees, but was more balanced in performance across examinee veracity. The 3T model produced 67.1% correct, 5.5% incorrect, and 27.5% NO decisions for truthful participants, and 71.0% correct, 7.0% incorrect, and 22.0% NO decisions for deceptive participants.

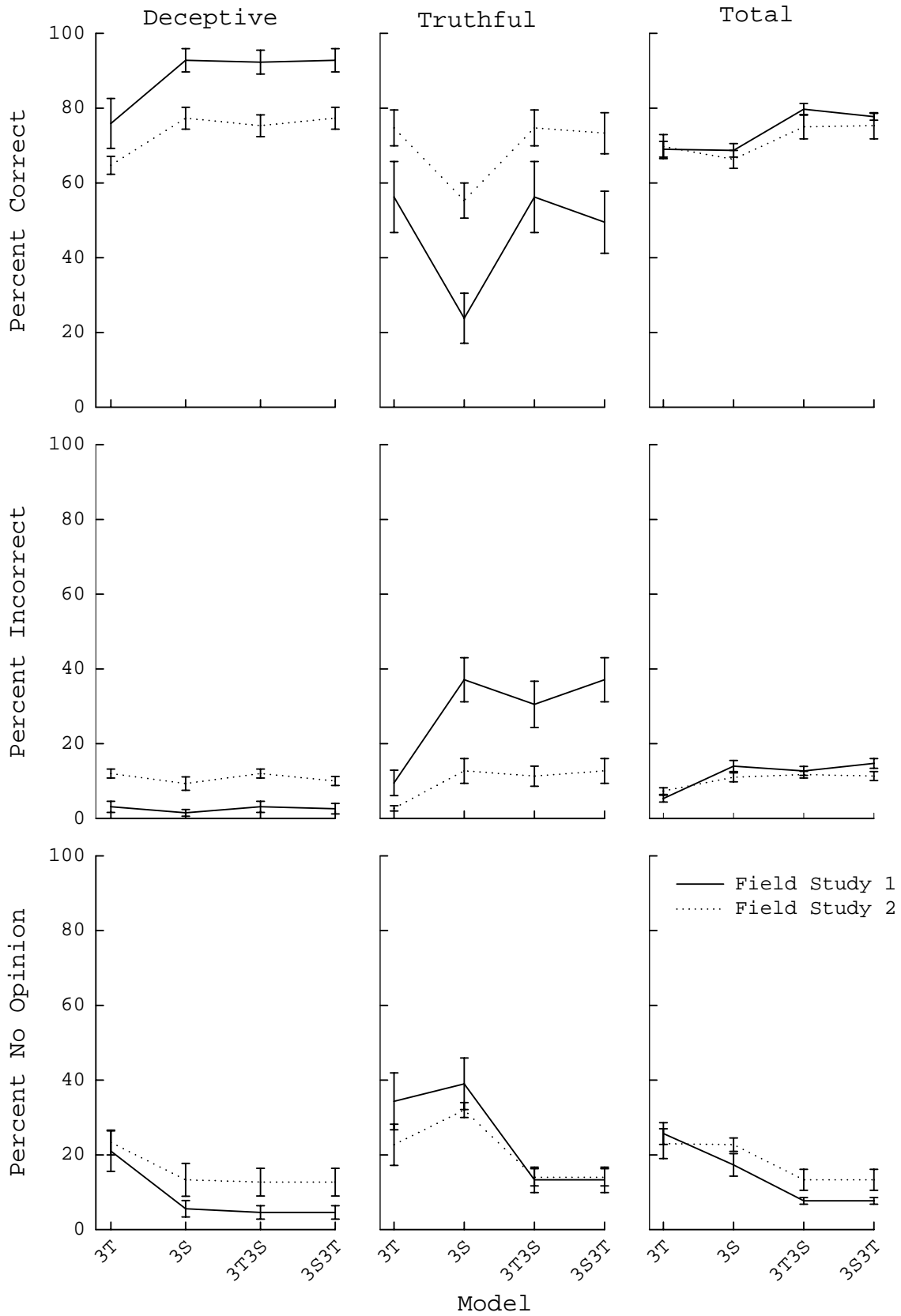


Figure 2. Mean (\pm SEM error bars) percentage of correct, incorrect, and no opinion decisions as a function of decision model and field study.

The 3T3S and 3S3T models were effective for both deceptive and truthful examinees, but similar to the 3S model, more accurately predicted veracity for deceptive than for truthful examinees. The 3T3S model resulted in 84.9% correct, 7.0% incorrect, and 8.1% NO decisions for deceptive examinees, and 67.1% correct, 19.2% incorrect, and 13.7% NO decisions for truthful examinees. The 3S3T model delivered very similar numbers, with 86.1% correct, 5.8% incorrect, and 8.1% NO decisions for deceptive examinees, and 63.5% correct, 22.8% incorrect, and 13.7% NO decisions for truthful examinees.

Collapsing across examinee veracity ($N_{\text{decisions}} = 600$), the 3T3S and 3S3T models generated more correct decisions (77.4% and 76.5%, respectively) than the 3T and 3S models (69.4% and 67.5%, respectively). In addition, the 3T3S and 3S3T models produced fewer NO decisions (10.5% and 10.5% respectively) than the 3T and 3S models (24.3% and 20.0%, respectively). Finally, the 3T3S and 3S3T models produced more incorrect decisions (12.2% and 13.0%,

respectively) than the 3T and 3S rules (6.3% and 12.5%, respectively).

Table 7 shows the average percentage of agreement among evaluators for the field data. No significant differences were found in the percentage of agreement as a function of decision model in either Study 1 or 2 or among the totals across studies. Table 8 shows the DEC values calculated using the field data. The DEC values were all statistically significant (indicating that all accounted for a significant amount of variance with ground truth), but none of the values were significantly different from each other, both in the context of individual studies, and when the values were combined across the two studies.

Table 9 shows the converted DEC values (i.e., correct decisions coded as +1, incorrect decisions coded as -1, and NO decisions coded as 0) across the two field studies. Repeated measures ANOVAs were again calculated using these values generated by each model as a dependent measure.

Table 7

Average Percent of Scorer Agreement as a Function of Four Decision Models (Field Data)

Model*	Study 1	Study 2	Weighted Mean
3T	66	71	69
3S	80	65	73
3T3S	80	74	77
3S3T	81	73	77

Note. *See Table 1 for Model descriptions

Table 8

Decision Efficiency Criterion Values for Decisions and Participant Veracity

Study	Model			
	3T	3S	3T3S	3S3T
1	.614	.397	.520	.474
2	.558	.520	.531	.585
Weighted Total	.586	.459	.526	.530

Table 9

Converted Decision Efficiency Criterion Raw Values as a Function of Decision Model and Study

Study	Model			
	3T	3S	3T3S	3S3T
1	.720	.580	.710	.680
2	.670	.650	.690	.720
Weighted Total	.695	.615	.700	.700

The values were significantly different for Study 1, $F(3,297) = 5.78$, $p < .05$. The contrast between the 3T and 3T3S models and the 3S model was significant, $F(1,99) = 10.06$, $p < .05$, and accounted for 98.9% of the omnibus model effect variance. The model effect was not significant for Study 2, but was significant across all studies, $F(3,597) = 5.55$, $p < .05$. The contrast between the 3T, 3T3S, and 3S3T models and the 3S model was significant, $F(1,199) = 10.65$, $p < .05$, and accounted for 99.7% of the omnibus model effect variance.

Discussion

As can be seen by examining Figure 2, the pattern of results delivered by the four decision models was very consistent across the two field studies. Across both studies, the 3T3S and 3S3T models generated among the most correct decisions for both deceptive and truthful examinees, and they produced the most correct decisions collapsing across examinee veracity. The increased percentage of correct decisions is due to the reduced percentage of NO decisions delivered by the 3T3S and 3S3T models (10.5% and 10.5%, respectively) relative to the 3T and 3S models (24.3% and 20.0%, respectively). However, as Figure 2 shows, there was also an increase in errors, indicating that some NO decisions were erroneously resolved. The average percentage of errors produced by the four models were: 3T = 6.3%, 3S = 12.5%, 3T3S = 12.2%, and 3S3T = 13.0%. Thus, while the 3T3S and 3S3T models resulted in more correct decisions than the 3T and 3S models, they also produced more incorrect decisions. However, excluding NO decisions, the percentage of correct decisions remained stable (3T = 91.6%, 3S = 84.4%, 3T3S = 86.4%, & 3S3T = 85.5%), indicating that accuracy for resolved decisions remained relatively consistent over the four models. The four models did, however, differ in utility. The 3S3T and 3T3S models both resolved 89.5% of the 600 decisions while the 3T and 3S models resolved 75.7% and 80.0%, respectively.

The DEC analysis, similar to the laboratory data, showed higher values for the 3T, 3T3S, and 3S3T models, and lower values for the 3S. However, these differences did not reach statistical significance. The converted value DEC analysis again showed that while

the 3T model generated a lower proportion of correct decisions, its low proportion of incorrect decisions resulted in values that rivaled or surpassed that of the 3T3S and 3S3T models. Replicating the laboratory data, the 3S model delivered converted DEC values that were significantly lower than any of the other three models.

General Discussion

Results of this study suggest several conclusions. First, as examination of Figures 1 and 2 indicates, the pattern of results generated by the four decision models was consistent across different studies and scorers for laboratory and field data, suggesting that this is a real and replicable phenomenon. Second, the overall accuracy of the four models, excluding NO opinions, is comparable (3T = 90.1%, 3S = 86.1%, 3T3S = 87.1%, and 3S3T% = 86.3%) suggesting that the one- and two-stage models are equally accurate. Third, the four models differ in utility with the two-stage models resolving, on average, 86.3% of the cases while the one-stage models resolving an average of 75.9% of the cases. These results indicate that the adaptive decision maker approach, that is, shifting strategies to fit the demands of the situation, is an effective technique for improving model utility. Fourth, the two-stage model increase in utility, due to an 10.4% increase in NO decision resolution, is accompanied by a corresponding 2.4% average increase in errors, relative to the one-stage models. Policy makers should consider the resource savings of increased utility versus the increase in decision errors when deciding which model to use. Finally, comparison of these results with those of our previous laboratory studies (Senter & Dollins, 2005; Senter, Dollins, & Krapohl, 2004), suggest that it may be possible to increase decision accuracy further by collecting additional data during the first PDD examination.

There are both advantages and disadvantages to the two-stage 3T3S and 3S3T models relative to the one-stage 3T and 3S models. Cases that had a -3 or lower in a spot but did not meet the -6 or +6 criteria resulted in DI decisions for the 3T3S, 3S3T, and 3S models, but not the 3T model. Thus, relative to the 3T model, more correct decisions and fewer NO decisions were

produced for deceptive examinees. However, more incorrect decisions were produced for truthful participants by the 3T3S and 3S3T models relative than the 3T model. Cases that exceeded the +6 total cutoff, but did not have positive values in all three question spots resulted in NDI decisions for the 3T3S, 3S3T, and 3T models, but not the 3S model. So, relative to the 3S model, more correct decisions and fewer NO decisions were delivered for truthful participants, but more incorrect decisions were generated for deceptive participants.

The net result of using the 3T3S and 3S3T models, for both laboratory and field data sets, is an overall increase in utility, in terms of producing DI or NDI decisions, relative to the 3T and 3S models (85.6% vs. 74.1%, respectively), and the maintenance of a similar level of correct decisions for resolved decisions (two stage = 86.5, one stage = 88.7). The 3T3S and 3S3T models produce higher accuracy than the 3T model for deceptive participants, and produce higher accuracy than the 3S model for truthful participants by simply combining the advantages of the two one-stage models.

Though both models produce nearly identical patterns of results, the question of whether the 3T3S or 3S3T model is more effective depends upon the priorities of the examiner. If it is more important to avoid false positive errors, the 3T3S model is slightly more effective, because a -3 in a question spot could still be classified as truthful if the sum of the three question spots is +6 or greater. If it is more important to avoid false negative errors, the 3S3T is slightly more effective because a total score of +6 or greater could still be classified as deceptive if there is a -3 or lower in any question spot.

In terms of the inferential statistics calculated for both the laboratory and field data, the 3T model is an effective decision model, not due to a high proportion of correct decisions, but rather to an extremely low proportion of incorrect decisions. This model performed well, producing relatively high converted DEC values for both laboratory and field data, in addition to high R^2 values. This model is attractive because it produces a high proportion of correct decisions, in cases where decisions are generated, hence the high

converted DEC values, but also produces a high proportion of NO decisions. The 3S model was more accurate for deceptive cases than for truthful cases, and consequently received consistently low raw and converted DEC values.

Regarding the adaptive decision maker hypothesis, the results clearly show contexts where each model would be the optimal approach to employ. In the context of high deceptive base rates, the 3S model should be the most effective. In the context of low deceptive base rates, the 3T model should be the most effective. Finally, in cases where a definitive decision was necessary and only a single data collection session was possible (as opposed to cases where multiple data collection sessions were possible), the 3T3S and 3S3T models should be the most effective. Thus, the results of this study have indicated what decision models should be most effective given predicted population base rates for deception. This corresponds to the findings of Payne et al. (1988), where it was shown that the most effective decision model varies across different factors and contexts. Further, Payne et al. showed that for multi-attribute, multi-alternative decisions, more labor-intensive decision approaches were more effective, such as the weighted-additive approach, provided that time pressure was not an issue. The increased proportion of correct decisions resulting from the use of the two-stage rules relative to the one-stage rules reflects the superiority of more effortful (i.e., involving more steps) decision models with PDD decisions as well.

While the current study showed robust effects across different laboratory and field data sets, it was not without its limitations. First, while the number of decisions produced in the application of the four models was quite large (i.e., 820 laboratory decisions, 600 field decisions), the number of individual cases was not as large (i.e., $N = 164$ and $N = 200$, respectively), and thus, the results may be somewhat misleading. A statistical power analysis (Bavery, 1996) indicates that 88 observations are required to accurately detect a difference (with a probability of .90) between a theoretical proportion of .50 correct responses and an observed proportion of .70 correct responses, using a two-tailed significance level of .01. A sample of 142

observations are needed, under the same conditions, to obtain a statistical power of .99. This analysis indicates that it is very likely that the results reported here (based on 164 laboratory and 200 field observations) will generalize to the real world—so long as the laboratory and field data are representative of the real world. While confidence would be increased with more observations, this study does represent a substantial sample. Other factors, unrelated to the number of observations, however, could influence how well these results would generalize to the real world. First, the data used in this study was not randomly chosen. It was taken from available laboratory studies, and a database of confirmed field cases maintained by the Defense Academy for Credibility Assessment. As discussed in the recent National Academy of Sciences study (2002) confirmed field cases may be a biased sample of all PDD examinations completed. We recently reviewed a sample of 242 specific issue PDD examinations completed by a government agency and found that only 97 (40.1%) of those examinations could be confirmed using objective criteria.

Second, it should be noted that the number of scorers contributing to this study was small. Fifteen different scorers were represented in the laboratory studies, and six were represented in the field studies. Although the pattern of results generated by the four models was very consistent across all studies, conclusions drawn from these results would be strengthened substantially by further cross-validation, using more scorers and data from additional laboratory and field studies. However, it should also be noted that the number of scorers reported here is quite high relative to other similar studies. Third, it is unknown whether producing definitive (e.g., DI or NDI) decisions within a single series is more accurate than the resolution of NO decisions using a second session of question series. While the 3T3S and 3S3T models produced more correct decisions than the 3T and 3S models, it is possible that resolution using a second examination may increase correct decisions without the increase in incorrect decisions. More research is necessary to answer this question.

While the two-stage models produced more correct decisions than the one-stage models in terms of the production of correct decisions, Senter and Dollins (2008) showed evidence that the combination of the two stage rules with a rule that integrated five question series in the second stage (referred to as the 3T5S and 3S5T models) could produce even larger increases in accuracy. Using scoring data from laboratory Study 1 and Study 3 (the only data for which assigned scores for five question series was collected), the 3T5S and 3S5T models resulted in more correct decisions (80.7% and 80.4%, respectively) than the 3T3S and 3S3T models (74.7% and 74.0%, respectively), a similar percentage of incorrect decisions (8.3% and 9.6%, respectively, vs 7.5% and 8.1%, respectively), and fewer NO decisions (11.0% and 10.0%, respectively, vs 17.8% and 17.8%, respectively). Thus, the three-or-five question series rule further increases correct decisions (beyond the 3T3S and 3S3T rules) by accurately resolving NO decisions.

In conclusion, the application of the one-stage 3T and 3S models in the form of two-stage 3T3S or 3S3T models to ZCT examinations produces more correct decisions and fewer NO decisions, with a small increase in the number of incorrect decisions, relative to the two one-stage models. This approach represents a simple and effective way to increase accuracy and resolve more cases using a single session examination. Though the 3T3S and 3S3T models likely represent a step in the right direction, future work in this area should include continued empirical explorations for optimal models that produce high accuracy for both deceptive and truthful examinees. These investigations may include the exploration of new cutoffs and the examination of diagnostic trends across examination, as a function of question series and as a function of data channel. Finally, the initial success of these innovative models on ZCT data suggests that the application of new decision models to other formats, such as the modified general question test, may prove fruitful as well.

References

- Bavery, J. L. (1996). Stat-Power (Version 2.2) [Computer software]. Portland, OR: Qeu Systems.
- Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. *Polygraph*, 28, 1-9.
- Blackwell, N. J. (1999). Polyscore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examinations from actual criminal investigations. *Polygraph*, 28, 149-175.
- Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Department of Defense Polygraph Institute. (2001). Zone comparison test. Fort McClellan, AL: Author.
- Department of Defense Polygraph Institute Research Division Staff. (2001). Test of a mock theft scenario for use in the psychophysiological detection of deception: IV (DoDPI00-R-0002). Fort Jackson, SC: Department of Defense Polygraph Institute.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.
- 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.
- Krapohl, D. J., Dutton, D. W., & Ryan, A. H. (2001). The rank order scoring system: Replication and extension with field data. *Polygraph*, 30, 172-181.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). *The adaptive decision maker: Effort and accuracy in choice*. In Hogarth, R. M. (Ed.) *Insights in Decision-Making: A Tribute to Hillel J. Einhorn*. Chicago, IL: University of Chicago Press.
- Senter, S. M., & Dollins, A. B. (2004). Comparison of three versus three or five question series contingency rules: A replication. *Polygraph*, 33(4), 223-233.
- Senter, S. M., & Dollins, A. B. (2008). Optimal decision rules for evaluating psychophysiological detection of deception data: An exploration. *Polygraph*, 37(2), 112-124.
- Senter, S. M., Dollins, A. B., & Krapohl, D. J. (2004). A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute. *Polygraph*, 33(4), 214-222.
- Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph*, 28, 10-27.

An Introduction to the APA's Panel on International Developments in Polygraphy

Frank Horvath

In 2005, at the APA seminar in San Antonio, the first-ever APA "International" panel was organized. The purpose of the Panel was to serve as a forum for discussion of contemporary events in Polygraphy and Credibility Assessment in countries outside of the U.S. The advent of the internet, changes in social, political and legal areas, the menace of terrorism, and the growing problem of transnational crime have raised the need for an awareness of developments in the field of Polygraphy. The panel was the APA's initial step in that direction.

The International Panel presentations, hopefully, will be a continuing feature of the annual seminar. Generally the organization will be as follows: Each Panel will consist of three or four presenters, each from a different country. Panelists will make a 20-30 minute presentation, after opening remarks from the moderator. After the featured "country" presentations, the moderator will summarize and integrate the important points. That will be followed by a question and answer session, with questions posed by audience members to the panelists. This organization is intended to promote greater interest in international issues and a better understanding of how practices and policies in other countries are related to those in the U.S.

In addition to a presentation at the seminar, each panelist also agrees to prepare a more detailed paper, in a relatively consistent way, that will be submitted to the APA's Editor for publication consideration. Examples of items that are to be covered in each of the papers include: Who is credited

with the initial development of polygraph testing in the country? When? Who uses polygraph testing? How many examiners are there and how are they selected and trained? What kind of instrumentation is used? What are the dominant procedures ("techniques") in use? What are the legal issues of most concern? What is the public perception of Polygraphy?

In issue 36(3) of *Polygraph* a paper from the second International Panel was published. In this issue of we are pleased to publish a paper from the third International Panel in 2007. In this paper the author has described Polygraphy in Lithuania.

About the Author

The reader will see that the author of this article is listed as "Anonymous." This is unfortunate but in this case necessary. After I made certain revisions in the manuscript I returned it to the author for approval. The author then sought permission to submit it for publication. That permission was not granted. However, it was agreed that because publication submission was a part of the original agreement when the Panel was formed, submission could proceed provided that neither the author nor the author's employer are identified. Citing the author as "Anonymous" and proceeding with publication was, under the circumstances, the only way to make this information available to the polygraph examiner community. I do wish to thank the author for the time, effort and sincere interest in providing this paper.

The Use of the Polygraph in Lithuania

Anonymous

History

Lithuania restored its independence on 11 March 1990. Almost from that time, polygraph testing has been used in my country, over 15 years now. In Lithuania only particular state agencies have the right to conduct polygraph examinations. All of them are indicated in The Law on the Use of Polygraph adopted in 2000. These agencies are: the Ministry of Interior, the Second Investigation Department under the Ministry of Defense, the State Security Department, the Ministry of Finance and the Special Investigations Service.

The first institution to start using polygraph was the Supreme Council's Security Division (today referred to as VIP Security Department under the Ministry of Interior); that was in 1992. The mission of the VIP Security Department is to ensure the protection of the state's leaders and their families and the official guests of the Republic of Lithuania from threats to their lives and (or) health and threats to protected objects in the country. Polygraph testing in the VIP Security Department is primarily used internally as a check for loyalty.

In Lithuania's defense sector, polygraph testing was first employed in 1999 by the Second Investigation Department under the Ministry of Defense.

Around 2001-2002, the State Security Department identified the requirement for polygraph testing. Up to now, however, the State Security Department does not have the necessary equipment or personnel. If a requirement for a polygraph examination arises, the examination is conducted by other authorized state institutions. However, the State Security Department should shortly acquire a polygraph and select personnel to act as polygraph examiners.

In 2003, the Police Department under the Ministry of Interior started the use of polygraph testing.

The Special Investigation Service, an institution fighting corruption, organized crime and crimes against state service, does not have a polygraph examiner. However, it does have plans either for acquiring a polygraph and training polygraph examiners or asking for polygraph examination support from other state institutions.

The Customs Department under the Ministry of Finance does not employ a polygraph examiner and does not have a requirement for such examinations. Thus, the Customs Department does not plan on acquiring a polygraph examiner or conducting examinations.

To summarize, currently only the Ministry of Defense and the Ministry of Interior conduct polygraph examinations in Lithuania. Other agencies that could conduct polygraph examinations do not now have the necessary equipment. Some though have plans for acquiring the equipment and training personnel. There are also some institutions that believe that polygraph testing has no value in their activities, or is either too expensive or too unreliable for checking on and obtaining information.

Training

The training of our first polygraph examiners in the VIP Security Department was provided in Lithuania by a Canadian of Lithuanian descent, a former commander of the Canadian Special Operations Unit.

At first, polygraph instruments were acquired from Canada. The Lithuanian community in Canada provided the necessary financial resources and in 1992 the first polygraph equipment (two Lafayette instruments) arrived in Lithuania. Later on, when the Supreme Council's Security Division was transformed into the VIP Security Department, the polygraph instruments remained with the latter agency.

Polygraph examiners from the Second Investigation Department went through a two-week polygraph training course in 1998. This was conducted by Polish instructors. In 1998, one polygraph examiner from the Department, together with an examiner from the Ministry of Interior attended a two-month Basic Polygraph Examiner Course in the United States at the Maryland Institute of Criminal Justice. They returned to Lithuania and shared their knowledge with other examiners. Those two examiners are members of the American Polygraph Association.

In 2007 one more polygraph examiner from the Second Investigation Department received training in America at the International School of Polygraph in Georgia.

The polygraph examiners from the Ministry of Interior have undergone a two-week specialized polygraph course in Lithuania. One more examiner attended a two-month course at the Maryland Institute of Criminal Justice and shared his expertise with others. In addition, one polygraph examiner from the Ministry of Interior attended a one-week conference dedicated to polygraph testing that was held in Russia.

So, the level of professional training of Lithuanian polygraph examiners varies: three officers went through training in the U.S. in polygraph schools accredited by the APA; others received training from visiting instructors from Canada and Poland, supplementing their expertise with the knowledge of the examiners who had been trained in the U.S.

In Lithuania, no institution provides training and education for polygraph examiners. The only training option is to travel to seminars abroad or read available professional literature.

Polygraph examiners from the Defense sector liaise with the American Polygraph Association. They first attended annual APA seminars in 2006. The polygraph examiners from the Ministry of Interior have closer ties with the Russian Polygraph Examiners School and attend conferences in Russia.

Examiners

As we know, currently there are seven polygraph examiners in Lithuania. The background education of Lithuanian polygraph examiners varies, but all of them are university graduates – psychologists, teachers, lawyers, and engineers. Except for the Canadian examiner of Lithuanian descent, whom I mentioned earlier, we do not know any other Lithuanian polygraph examiners who reside abroad.

It is difficult to point to the most influential polygraph examiners in Lithuania. The examiners from the VIP Security Department gain most of the mass media attention and are better known by the public. As the VIP Security Department was the first to use polygraph testing in Lithuania, the journalists are used to asking them questions concerning polygraph testing. In addition, the Police Department used to rely on the VIP Security Department to support their polygraph testing in criminal investigations.

As the number of polygraph examiners is very limited in Lithuania, there is no effort to establish a Lithuanian Polygraph Examiners Association. However, the idea of establishing one in the future is now circulating. As far as we know, none of the Lithuanian examiners belong to any polygraph examiners' association that may exist in Europe.

Examinations

It is difficult to establish an exact number of polygraph examinations conducted annually in Lithuania. The types of examinations and the institutions conducting them do not provide such data. However, the approximate number of polygraph exams carried out every year is probably around 100.

As far as we know, most polygraph examinations are conducted by examiners in the Second Investigation Department. Since 1999 that Department conducted more than 800 polygraph examinations, most of which were related to personnel clearances for access to classified information. The number of exams conducted at the request of other state institutions is very limited. We believe that in the future the Defense Sector could

expand the employment of polygraph testing, for example, in the investigation of instances of malfeasance.

The State Security Department has issued a few requests for the polygraph examinations. These were carried out by other state institutions.

The Police Department conducts several polygraph examinations annually. We do not have data on the number of police requests for polygraph exams to the VIP Security Department, but we believe it is limited to about 10 requests annually.

The first polygraph examinations carried out in Lithuania were for personnel vetting purposes. The VIP Security Department used polygraph testing in the selection of personnel to state institutions, for screening regarding loyalty, during malfeasance investigations and upon the request of other state institutions. For example, the Ministry of Interior used the polygraph services of the VIP Security Department in criminal investigations. The police employ polygraph examinations on somewhat of an experimental basis such as while conducting felony or related crime investigations. The police usually issue a request for a polygraph examination if they have an investigation without enough evidence or if the investigation has otherwise reached an impasse.

In the Defense sector, the first polygraph examinations were conducted in order to establish personnel reliability. Since 1999, all managerial personnel have been required to undergo polygraph examinations. Currently, polygraph examinations in the Defense sector are employed for personnel clearance to classified information, for selection of personnel to the Second Investigation Department and for the Department's periodic personnel investigations.

The State Security Department has requested polygraph examinations as part of its ongoing investigations. This is done in order to gain additional information on the persons who presumably breached the law or are in some way linked to those persons.

The Police Department employs polygraph examinations during the personnel selection process and also during malfeasance investigations. The subordinate commands may employ polygraph testing during pre-trial investigations. We do not have information on the use of polygraph testing in clandestine operations in the department. As the subordinate commands do not have polygraph equipment, they rely on the VIP Security Protection Department's polygraph support in pre-trial investigations. That department is the most experienced in employing polygraph testing in the pre-trial phase. Usually, polygraph testing is employed in cases of murder, rape or the leak of sensitive information.

At the beginning, the Police Department employed polygraph testing for the selection of personnel. Other subordinate commands initially conducted only experimental examinations, for example, when all the legal proceedings were exhausted and when an investigator reached an impasse. Moreover, the first polygraph examinations were used in those pre-trial cases that gained major public attention. Currently, the Police Department gives priority to the use of polygraph testing in pretrial cases and clandestine operations, usually dealing with organized crime and corruption. However, polygraph testing is sometimes hampered by the lack of information or, on other occasions, by the leak of information to the mass media.

Testing Issues

Data on countermeasures to polygraph techniques and the number of attempts to do so during the polygraph examinations are very limited. The most often employed countermeasures are physical movements, even after the polygraph examiner has warned the examinee to stay still and to concentrate on the testing. In some cases, examinees have arrived for their polygraph exam having consumed large amounts of alcohol. In these cases the polygraph examinations were canceled. In several cases, examinees have employed psychological manipulation techniques; for example, they tell the examiner about the difficult life they have led, they try to move the examiner emotionally, they request a postponement of the examination or they complain about the

setting of the examination day saying the date is unacceptable to them.

In the Second Investigation Department, there have been no cases when an examined person in the course of a polygraph examination refused to participate in the polygraph examination and then asked to leave. Moreover, no cases have been recorded in which the examinee filed an official complaint against the polygraph examiner or the examination in general.

The State Security Department has not recorded any attempts by examinees to use countermeasures. However, in several cases persons have refused to undergo a polygraph examination. The VIP Security Department has also faced several refusals. Moreover, the VIP Security Department has experienced situations in which not all the information concerning the examination is given to the polygraph examiner. This makes polygraph testing less effective and less efficient.

The Police Department has not recorded any indisputable attempts by examinees to use countermeasures. However, examiners in that department have recorded that some examinees may have employed movements and attempts to psychologically change polygraph data.

In Lithuania, several polygraph examination methods are employed. The Second Investigation Department uses the Relevant/Irrelevant procedure (R&I), the Zone Comparison (ZCT) and Modified General Question Techniques (MGQT). The VIP Security Department uses a modified D.T. Lykken – B.A. Varlamov testing method; this is a variation of the Guilty Knowledge Test (GKT) as modified by Russian polygraph examiners referred to as “cognizance of circumstances.”

The Police institutions usually rely on R&I, Peak of Tension (POT) and Comparison Question (CQT) polygraph examination methods. If a police institution requests a polygraph examination from the VIP Security Department, the latter usually employs a mixed methodology and, when possible, the modified GKT. According to the VIP Security Department, the Law on the Use of Polygraph,

as it is today, limits the use of new polygraph examination methods.

There are no websites in the Lithuanian language about polygraph testing that would provide information on the use of polygraph countermeasures. I believe that this is partly due to the fact that the use of polygraph testing is very limited in Lithuania and for that reason this type of information is not seen as very important to the public.

Legal Issues

Until 2000 Lithuania had polygraph regulations only in the laws which regulated usage in the defense sphere. On August 29, 2000, the Law on the Use of Polygraph was passed. The draft was prepared by the Parliamentary National Security and Defense Committee. On 12 July 2002 the Government approved the Rules for the Use of Polygraph in Lithuania.

The Law on the Use of Polygraph raised some debate and doubt about whether it was in conformity with the Lithuanian Constitution. There were also questions about whether the results of an examination were reliable and what the consequences of a negative conclusion (e.g., “deception indicated”) would be. The mass media captured the comments made by the Chairman of the National Security and Defense Committee stating that some political powers were strongly opposed to attempts to legitimize polygraph testing as a preventive measure against fraudulent civil servants.

The law identifies the institutions that are authorized to conduct polygraph examinations. They are the Ministry of Interior, the Second Investigation Department, the State Security Department, the Ministry of Finance and the Special Investigations Service. Moreover, the law provides against the use of polygraph testing by other institutions. Thus, private, commercial companies or private persons cannot employ polygraph testing in Lithuania.

According to the law a polygraph exam can be performed only in these cases:

- * When permission to work with classified information is being issued and if there

are grounds to believe that the individual withheld or provided false biographical data;

* In cases of the investigation of malfeasance, internal or operational investigations;

* In cases of crimes or other misdemeanors or irregularities in working with classified information;

* In cases when there are grounds to believe that a person who deals with classified information is under unlawful pressure;

* In cases of a person's request to conduct repeated polygraph examinations on him/her.

The Law on the Use of Polygraph and the Polygraph Examination Rules are the only legal documents regulating the activities of polygraph examiners in Lithuania. However, each institution, which is granted the right to use polygraph testing, must also have internal documents regulating the conduct of polygraph examinations. Also there are several laws in Lithuania partially dealing with polygraph testing. These include: the Law on State Secrets (1999), the Law on Operational Activity (2002) and the Criminal Proceeding Code (2003).

The scope and content of the Law on the Use of Polygraph was largely predetermined by the novelty of the idea of using polygraph testing. There were a very limited number of polygraph examiners and the public knew almost nothing about the topic. The requirement for polygraph examinations was not generally known either, so the general opinion was that polygraph testing would be used in screening personnel for security clearances. Today the law needs to be changed, as it regulates only polygraph examinations on civil servants seeking to obtain security clearances.

According to the Police Department, 2005 marked a turning point in polygraph examinations, because for the first time a court accepted a polygraph examination report as evidence in a criminal investigation (This was a Vilnius District Court ruling in

three murder cases). Other Police subordinate commands, which requested the VIP Security Department to conduct polygraph examinations in pre-trial investigations also presented reports from polygraph examinations to the Court in 2005 and 2006 and they were accepted as subjective evidence of a criminal act. In other words, the polygraph results proved that the defendant knew the circumstances of the criminal act and was guilty for committing the criminal act. In these polygraph examinations the examiners used "cognizance of circumstances" examination methods.

The polygraph examiner from the VIP Security Department testified in three separate criminal trials as a specialist. Moreover, the polygraph examination reports led to the dismissal of 10 false versions in different investigations.

The Second Investigation Department has no experience in conducting polygraph examinations presented in Court. None of that department's examiners has participated in court hearings.

According to the Lithuanian Criminal Code, evidence is the data received following legal proceedings. That is why a polygraph examination report can be accepted by the Court as evidence in criminal cases. If a polygraph examination is conducted as part of operational activities or while checking information received using operational methods, the results of such an examination could only be treated as basic information about the crime and those involved in it. Such reports do not constitute evidence because the polygraph examination was employed based on the Law on Operational Activities and not on the Criminal Code. However, the results from polygraph examinations could be very important to future operational activities and they could even determine the success of operations or of pre-trial investigations.

Lithuanian police officers believe that in most criminal investigations, reports from polygraph examinations are not sufficient to bring charges or to start a trial. It is of utmost importance to collect all other crime evidence. The polygraph examination report could only be treated as important

information in bringing criminal charges against somebody or in acquitting him/her of a crime. However, there are certain crimes where a polygraph examination report could be the only evidence determining the fate of a suspect. There was one exceptional case in Lithuanian police history when regional police headquarters requested the support of polygraph examiners in investigating an accident in which one of three men taking a boat trip fell from the boat and drowned. The relatives of the deceased man accused the two survivors of an attempt to murder him. Based on the polygraph examination report the case was dismissed.

Political and Social Issues

The introduction of polygraph examinations in the Defense sector has met our expectations and has brought very good results in revealing negative biographical and other information that a person is trying to conceal, in investigating the circumstances that are unfavorable to the person under examination, and also in revising the personal data provided by the person in question. Persons seeking employment in the Defense Sector and those already employed do not want to risk their reputation and professional career; thus they reveal biographical facts that would otherwise remain undisclosed. Such methods as interviewing, data collection and self-answered questionnaire items do not provide results as good as those provided by a polygraph examination.

Polygraph examinations also serve as a preventive and disciplinary measure, thus increasing the reliability of personnel and preventing civil servants and officers from engaging in illegal activities. These persons know that they will undergo polygraph examinations on a regular basis and they stay away from engagements that would be a cause for employment difficulties. However, in the Defense sector the polygraph examination report cannot serve as the only basis for terminating an employee or refusing to employ a candidate. Information from other sources has to confirm the polygraph test results.

It was not easy to introduce polygraph testing into the Defense system. The Defense civil servants and officers expressed discontent and animosity in regard to both

polygraph examiners and the examination process itself. However, such a negative attitude diminished as time passed.

The specialized and regional police headquarters know very little about polygraph examinations and their reliability. That is why most of them are skeptical of polygraph testing.

The general conclusion could be drawn that polygraph testing is not popular in Lithuania. Even though the use of polygraph is strictly governed, both the society and civil servants are mistrustful of polygraph examiners and the reliability of such examinations. There are a lot of discussions going on concerning human rights issues and polygraph examinations are seen by some to intrude into people's private lives. However, from time to time, especially with the eruption of a new political scandal, the mass media starts to speculate that every politician should undergo a polygraph examination, especially those in a higher echelon. The mass media also covered a case when several senior officials were alleged to have engaged in unauthorized disclosure of classified information. They were offered to undergo a polygraph examination but they refused to do so. However, the mass media doesn't inform the public of the results of polygraph examinations even in major legal cases.

From time to time one can see articles about polygraph testing in the press. Television stations also have broadcasted several discussion shows dealing with polygraph testing. However, polygraph testing does not attract much mass media attention and Lithuanian television stations do not have shows focusing on the elements of polygraph examinations. Nonetheless, the general mass media attitude towards polygraph testing is rather more positive than negative.

The VIP Security Department told me that the mass media quite often contact them with requests of information about polygraph testing. However, their interest is usually very superficial. Nonetheless, the VIP Security Department tries to use every opportunity to inform the public and to popularize polygraph testing. The VIP Security Department together with the Police Department has

compiled several publications about polygraph testing for distribution to the press.

The Police Department believes that mass media interest in the polygraph testing is very low. It has not been contacted by the mass media on this issue.

Polygraph examiners in Lithuania do not face any specific social or political problems, as their work environment is restrictive in nature and the general public does not know much about their work. So the conclusion is that Lithuanian public interest in polygraph testing and the challenges to it is low. Polygraph examination reports very rarely reach the courts as evidence, even though examinees never object to the results of a polygraph examination.

In 2006, an opinion poll concerning polygraph testing was conducted via the Internet. The audience of the poll was contractors (66.2 percent), managers (20.55 percent) and others (13.24 percent). The respondents fell into three age categories: 26-40 (56.36 percent), 18-25 (41.82 percent) and over 40 (1.82 percent). The objective of the poll was to gather information on the relevance of polygraph examinations.

We do not know who initiated the opinion poll. Also, it is difficult to evaluate the accuracy of the personal data presented. So the results of the poll do not necessarily represent the real perceptions and knowledge of polygraph testing in Lithuania.

The poll came out with the following results: As reported, 69.8 percent of the respondents know how polygraph testing functions. About 74 percent of the respondents believed that voluntary consent to undergo polygraph examinations would not violate their right and freedoms. Fifty-four percent of the respondents agreed to undergo a polygraph examination if it was required for a well-paid job. About 72.9 percent of respondents agreed to undergo a polygraph examination if it was required to prove their innocence in a wrongful accusation. About 59 percent of the respondents believed that they would not be able to cheat a polygraph examination. However, when asked if they would use polygraph examinations before employing personnel in important positions,

54.8 percent of the respondents replied negatively. About 47.3 percent of the respondents believed that the reliability of polygraph testing is 50-50, and only 34.2 percent believed that the testing is 90 percent reliable. About 18.7 percent of the respondents believed that polygraph testing was unreliable.

The results of the poll revealed that most of the respondents knew about the operational aspects of a polygraph and that they would consent to being examined in order to be promoted to higher ranks or to dismiss the criminal charges. However, most of the respondents, if acting as organization leaders, would not use polygraph examinations before employing people in important positions. Also, the respondents questioned the reliability of polygraph testing. In summary, however, the respondents had a more positive than negative attitude toward polygraph testing.

Research

In Lithuania polygraph testing has been used for 15 years. Despite this, not a single scientific research study has been done on polygraph testing in the country. There are students in several universities who have written their graduation papers on polygraph testing. Vilnius Technical University has conducted experiments with a voice stress analyzer prototype and has a large experimental database. However, there is no consistent research or experiments with polygraph testing. Scientific research on polygraph testing is not now an agenda item in the scientific research institutes.

However, every Lithuanian institution that conducts polygraph examinations strongly supports the requirement for specialized education and cooperation between subject matter experts. Also, they believe that scientific research on polygraph testing is important and they would participate in and cooperate with persons engaged in this type of research.

Instrumentation

Lithuanian institutions mostly use computerized polygraphs produced by the Lafayette Instrument Company in the U.S.

Currently, one private computer distribution company based in Vilnius would like to provide and sell Russian-made polygraphs in Lithuania. These instruments do not have a certificate yet which would provide permission to sell this instrument in the European Union, in which Lithuania has been a member since 2004.

We have not heard of any on-going developments of polygraph technologies or sensors in Lithuania, except for the voice analysis system that I mentioned earlier, which was developed and tested in one of our Universities.

Problems / Issues

The major issues that we face while working with polygraph technologies are the lack of knowledge on polygraph testing, the lack of periodical training for polygraph examiners, the difficulties in obtaining refresher training, the lack of co-operation among the institutions conducting polygraph examinations, and the difficulties in obtaining the legal basis for polygraph examinations in other countries. The most pressing requirement is for regular advanced training for polygraph examiners, something which currently is not available.

The Police Department believes that their polygraph examinations are hampered by the fact that the examination itself constitutes only a method of criminal investigation, but, nevertheless, it is governed by the law. Other criminal investigation methods and the application of these methods are not governed by any law.

Another major issue is the outdated Law on the Use of Polygraph, which does not meet today's requirements. It needs to be changed in order to grant other institutions/companies and private persons a right to use

polygraph testing. Moreover, the target audience for polygraph examinations should not be oriented only to personnel clearances for access to classified information.

Lithuanian polygraph specialists do not face any specific problems. However, one challenge that could be worth mentioning is the mentality of some examinees. For a number of years, the people in Eastern Europe, including Lithuania, had to live in a very different political system from what we have today. People in the previous system had to accept what the official leaders said and they had to hide their true political and religious beliefs. In other words, people used to say one thing but think another. This kind of suppression was caused by the fear of being prosecuted or deported based on beliefs that were different from the official position. For several decades people were double-faced and lying was a justifiable measure of survival and adjustment to society. A couple of generations of children were raised in these decades. That mentality hampers polygraph testing of the older generation who experienced the "old" governmental system. Because they tended to suppress their biographical data or to varnish it, they learned not to trust the government and, similarly, government-administered polygraph testing. For this reason, it is sometimes difficult for a polygraph examiner to develop satisfactory comparison questions. Many of these examinees, while under examination and telling of their past life, tend to stress that in the "old" days, certain activities or actions were normal; they were not crimes because they had to be done to survive and everyone did them. Today the same activities are considered to be illegal. So the problem is not evident in the younger generation in Lithuania; these younger people grew up in a democratic state and are more like those in the U.S. and elsewhere with respect to how they perceive comparison questions.