Polygraph

VOLUME 27 1998 NUMBER 2

Contents

Published Quarterly

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

Managing Sex Offenders in the Community With the Assistance of Polygraph Testing

George H. Baranowski

Abstract

To quote a January 1997 research brief by the National Institute of Justice, "A 'cure' for sex offending is no more available than is a cure for epilepsy or high blood pressure. But use of a variety of interventions can help manage these disorders." An important aspect in this assemblage is the use of Polygraph Testing. A realistic objective of treatment is to provide sex offenders with the tools to manage their inappropriate behavior. A therapist can, in many cases, teach offenders self-management skills for avoiding high-risk situations through identification of decisions and events that precede them and through correction of their thought distortions.

Treatment focuses on recognizing and managing deviant sexual behavior and offenders' thoughts and attitudes that promote it. However, in pursuing safe and effective treatment of sex offenders in the community, therapists must obtain full disclosure of offenders' sexual histories. Sex offenders must carefully assess their lives and identify relationships, emotional states, attitudes, and behaviors that they may consider "normal" which are not acceptable to the larger community. Use of the polygraph helps ensure that offenders fully reveal their sexual histories--information that is essential to the development of effective treatment programs.

Polygraph testing is useful for periodic monitoring of the offender in treatment and focuses on his/her activities in the community setting. The goal of polygraph examination is to obtain information necessary for risk management and treatment, and to reduce the sex offender's denial mechanisms. The examiner evaluates the offender's answers to test questions and renders an opinion; truthful, deceptive, or inconclusive. Deceptive results flag areas that the treatment provider and supervising officer need to investigate further. Every effort is made to assist the offender in obtaining a positive evaluation so that treatment can be informed and relevant. To this end, polygraph data should be used in conjunction with other information when making decisions about case management of sex offenders.

Keywords: Polygraph, sex offender testing

George H. Baranowski is Director of Mindsight Consultants in Michigan City, Indiana and is a member of the APA. Requests for reprints should be addressed to the author at Mindsight Consultants, 128 West 10th Street, Suite 102, Michigan City, Indiana 46360-3514.

Introduction

There is growing utilization of polygraph testing in the treatment and control of child sexual abusers. The issues of validity, reliability and admissibility of the polygraph are vital questions to many. As a quick answer, current studies show that field polygraph testing has an overall accuracy rate of 98%, which compares favorably with other scientific measurement disciplines that are routinely admitted as evidence, including ballistics, handwriting analysis, or hair and fiber evidence. Many professionals who work with child sexual abuse cases see considerable utility in the use of polygraph tests both as a means of probation or parole surveillance of convicted child sexual abuse offenders and as an aid to the therapists working with these offenders. The National Conference on Sentencing Advocacy (Practicing Law Institute, 1991), recommended an expansion of the use of polygraph monitoring of probated offenders, especially sex offenders.

Several jurisdictions have begun sex offender testing programs. As of this writing, Alaska, Arizona, Colorado, Florida, Idaho, Indiana, Texas, and Washington have such programs. A bill in New Jersey has just been introduced to provide polygraph monitoring tests every six months to sex offenders who committed their crime against a victim under the age of 18. An option to sentencing would give the New Jersey Court the ability to extend probation to "lifetime community supervision," requiring polygraph monitoring. Arizona has had such enactment the past 5 years, with outstanding results. For example, in Maricopa County, over 1200 probationers have been involved in the program which shows only 1.5% have re-offended (19 individuals)(1997). This figure, as opposed to a figure of 94% re-offense in communities who are without treatment and polygraph testing, stands in strong testament to this program's benefit. As of this writing, the Florida legislature has passed an amendment to the Florida Statute to mandate the use of polygraph testing as part of the treatment program of sex offenders. A recent article in the *Chicago Tribune* (May 16, 1997) pointed out that Texas Officials report there are nearly 50,000 sex offenders on probation who are subject to polygraph testing, electronic monitoring and neighborhood notices. The Hunt County Texas probation department advised they would not supervise a sex offender without polygraph testing, the article reported. A natural transition may occur which will see polygraph testing in probation monitoring of offenders in non-sexual cases. As reported in the *Los Angeles Times* (June 10, 1997), U.S. District Judge Richard Haik postponed the sentencing of a convicted drug dealer who advised he had only dealt drugs this single incident. The possible sentence may be 5 years in prison and a $350,000 fine. Judge Haik has ordered the defendant to undergo a polygraph test to verify the defendant's assertion.

Well established scientific principles provide the basis for the modern polygraph test. The parent science is psychophysiology, a recognized specialty within the field of psychology. The polygraph was not invented for, nor is it limited to, use in detecting deception. It is a scientific instrument long used in the psychophysiology field to measure and record simultaneously reactions that relate to psychological states. There has been a movement among some polygraph professionals to implement a change in traditional polygraph terminology toward what they consider to be a more descriptive model. For example, they suggest the discipline itself be termed "forensic psychophysiology". The description which traditionally has been called "polygraph examiner" would be termed "forensic psychophysiologist". The overall complex of activities conducted by the forensic psychophysiologist would be called a "psychophysiological detection of deception" test or "PDD". (It's noted that the term polygraph also works as

a "polygraph detection of deception" test). To paraphrase Dr. William Yankee's discussion on this matter (1994), he clarifies that the polygraph is actually the instrument and not the process. The term polygraph means many writings. Dr. Yankee noted that this instrument is utilized in the science of forensic psychophysiology and the practitioner is the forensic psychophysiologist. To illustrate his position Dr. Yankee noted that, a surgeon may use a scalpel in his work, it would be inappropriate to call him a "scalpelist". However, other professionals in the field respond that the traditional terms of "polygraph, polygraph examiner, polygraphy, and polygraphist", are well ingrained within the public and professionals alike, and that these terms are both accurate and adequate. Whether or not these suggested changes will prevail is unknown. The matter will obviously come before the polygraph committees of the ASTM, (American Standards and Testing Materials) who are working on the establishment of standards for the field. This paper does not intend to take a position in this matter, and for purposes of simplicity, traditional terminology was utilized.

It is without question that no single modality, including polygraph testing, is a panacea for the various problems associated with the prevention, detection, and treatment of child sexual abusers and their victims. However, it is a tool that is being used successfully in several important ways to assist professionals in other areas. Polygraph Examiners work with the offenders, those accused of offenses, probation and parole officers, judges and offender therapists. PDD testing, as it relates to child sexual abuse, finds three types of polygraph examination activities relevant: (1) specific testing in the investigation of accusations, (2) probation monitoring (or surveillance), (3) disclosure testing.

Specific Testing

Upon occasion, both public and private polygraph examiners conduct polygraph testing on individuals accused of sexually molesting children. Specific issue testing, as it is called, serves to assist the investigators in determining whether charges should be brought. A polygraph test in such cases is no different from testing an individual suspected of another offense, such as burglary, murder, or shoplifting. Like other offenders, child sexual abusers have denial, irrational thinking behaviors, and other protective mechanisms for their self-preservation. Training in polygraph question formulation emphasizes morally neutral questions that define events. For example, the examinee would be asked, "Did you put your hand on Child X's vagina?" If indications are the examinee did touch the child in that manner, the social and legal meaning of that incident will be deemed independently of any rationalization, feeling of guilt, or lack of feeling of guilt, on the part of the examinee. In a polygraph test, the examiner does not present questions that would require the examinee to be subjective. For example, it would not be useful to ask, "Did you touch child X in a sexual way?"

In some cases, child sexual abuse is ongoing at the time of testing, or the abuser has committed multiple offenses that can range into the hundreds or thousands. Having a high number of offenses does not make child abusers unique and untestable. Thieves, burglars, drug dealers often have scores of offenses in their past. When caught, the central or legal focus is on that particular offense that drew public attention to the offender, and of course, just one offense defines the legal status of the offender. With burglars who have committed multiple crimes for example, there may be little benefit in discovering exactly which burglaries can be attributed to that individual offender. The situation is different

however, with child sexual abusers. Identifying victims of child sexual abusers is important because some who did not report the abuse may need assistance from therapists or other helping professionals. Full disclosure polygraph testing is a technique that can be useful in this process and will be discussed later in this brief. However, specific testing is an important investigation tool in the process of identifying and convicting, or in some fashion gaining legal control over the child sexual abuser. If the offender does not acknowledge at least the focus incident, there is little hope of discovering the extent of that person's offense history.

Often specific polygraph testing is requested by an accused who demands an examination to clear himself/herself of allegations. A fairly common example is a hotly contested divorce and child custody case (McGraw & Smith, 1992). One parent may bring false allegations of child sexual abuse to gain advantage in the custody hearing. With only the word of one parent against another in such cases, polygraph testing can be a useful investigative tool. The results of many such tests clear an innocent person before charges are brought. Polygraph tests have also been conducted on similar issues brought before a child welfare agency such as the Department of Family and Children or some other type of child protection investigation body. The primary goal of such agencies is the protection of the child when suspicions of child abuse come forward. Too often such allegations involve a victim too young to articulate consistent allegations and where there is little or no physical evidence. Such suspected victims may be put into foster care during the time of investigation. A secondary goal of protective agencies is family reunification when appropriate. Polygraph tests have been an important aid for the courts and these agencies to assist in a proper decision for everyone's benefit and protection. Through its use, offenders have found their way into treatment programs and non offenders have been reunited with their families.

There are other applications of specific testing where the offender claims he pled guilty to the crime but alleges he/she did not commit the offense charged. This is a subject who has been adjudicated and sentenced to treatment as part of his/her plea agreement and order of probation, but then maintains in group or individual treatment sessions that he/she did not commit the alleged offense, and only pled guilty at the direction of his/her counsel to achieve a perceived lighter sentence then would have been received if found guilty at trial. Continuing to deny his/her offense is of little benefit to the therapist or the offender. A tool in breaking this type of denial to the original offense is also the use of specific testing, where the focus issues deal with the charged offense. Assuming this subject's position is false, and realizing that both the therapist and treatment group now have proof of his/her commission of the offense, the denying offender must revisit his/her position, abandon denial as a strategy, and allow treatment to progress.

Probation or Parole Monitoring (or Surveillance)

Increasingly, polygraph examiners are being asked to play a role in protecting the community and the victim from convicted child sex abusers by performing periodic polygraph examinations designed to provide early warning of potential problems. As noted in a publication describing the evolution of an alternative sentencing program in Alabama (Williams, Morrison & Terrell, 1993), the polygraph examiner is only one additional safeguard among several that include the traditional probation or parole officer, therapists, volunteer monitors and alternative sentence caseworkers. Failing a periodic surveillance polygraph examination should be cause for further investigation on the part of

therapists, probation officers, and others involved in the case. A common condition of probation or parole for a child sexual abuser is to stay away from the victim. As often is the case, the victim may be the daughter of a woman who is still in a relationship with the offender. The court may not try to keep the victim's mother from continuing the relationship with the offender, but typically would forbid the offender from visiting the residence where the victim lives with the mother. Discovering in the course of a polygraph examination that the offender has violated the mandate to stay away from the victim's home can result in proper remedial actions before the offender again sexually abuse the child again.

Consider the probation/parole supervision process as it occurs without the services of the polygraph examiner. The probation/parole officer periodically interviews the offender. The officer asks basically the same questions as the polygraph examiner. For example, "since our last discussion, have you had sexual contact with anyone under the age of 16?" If the offender answers "no," the officer must look for additional clues to validate the answer. Body language, perhaps, is the method the officer uses to make a decision concerning the offender's veracity. The present writer calls this judgment process "demeanorology". It is far less effective in judging the truthfulness of a verbal statement than the use of an instrument that records uncontrollable physiological reactions to these crucial questions.

Disclosure Examinations

A separate and significant area of polygraph testing associated with child sexual abuse involves disclosure testing. Disclosure tests explore all of the probated offender's past sexual behaviors in addition to those known to the court and/or therapist. There are several good reasons for using disclosure tests: (a) to provide the therapist with a more complete picture of the types of sexual behaviors of the offender in order to better plan treatment strategies, (b) to help the offender overcome denial as a precursor of psychological healing, (c) to identify any additional victims who may not have come forward since they may be in need of therapy, (d) to make sure that the offender has not committed new crimes while negotiating for an alternative sentence to known crimes. A main purpose of the disclosure test is to provide assistance to the therapist working with offenders. After the therapist has worked with an offender for several weeks and is beginning to feel that the offender has made some disclosures but has not told everything, the therapist finds it advantageous to send the offender for polygraph testing by a specialist in child sexual abuser testing. The Indiana Polygraph Association has, as many other state jurisdictions, guidelines for special training or certification programs for these specialists. Such examiners require advanced specialized training.

The specialist may use a number of methods of exploration during this phase. One technique may direct the offender to complete a comprehensive sexual history questionnaire and then tests the offender on the issues that are crucial to the offender's treatment and probation program. Such disclosure tests frequently help the offender to move further along toward a program of honesty. It also can give the therapist better tools for treatment planning if the offender has sexual proclivities not previously known to the therapist.

The examiner's testing protocol and interview expertise goes beyond having the examinee fill out a questionnaire and merely testing broadly whether they had lied somewhere in the questionnaire. Sexual issues must be explored with an in-depth interview and using special testing skills if there is to be an effective clinical polygraph examination. An erroneous perception had existed that such exploration into

the examinee's sexual history is the exclusive job of the treatment provider specialist. Therapists would be quick to point out that such exploration is often an ongoing process lasting throughout the period of treatment, and, more importantly, it is the responsibility of everyone on the treatment team. Therapists advise it is not uncommon for the polygraph examiner, through the process of interview and testing, to obtain more sexual disclosure information than obtained during traditional treatment processes. In fact, the mere mention that a polygraph examination is going to be conducted, will frequently produce disclosures among offenders in treatment who had staunchly denied any such acts previously. This point is dramatically illustrated in Dr. Abrams' book, *Polygraph Testing of the Pedophile*, (1993) describing a therapist's example of the degree of denial held by sex offenders and the effectiveness of the polygraph disclosure process (pages 17-18).

> There was a group of eight men in a particular treatment group that had seemed to be doing quite well. It was unusually stable, with the same members remaining in the group for several months. I think they'd had one new member in recent months. And what had happened is that the group had coalesced together to work on what they had disclosed without really pushing each other too hard to open any new doors. And I was getting increasingly comfortable, and so in January, I made the decision to begin polygraphing each of the men in this group.
>
> Of the eight, the number of charges that they were accused of was 16. The number of crimes which they were ultimately sentenced for, was only 12. Through the use of the interview, sexual history questionnaires and group process, up until January of this year (1990) that group of eight men had disclosed a total of 2,085 deviant behaviors in their history, deviant or inappropriate sexual behaviors. By May 1, after having completed a series of polygraphs and working with these men, that number had grown to 13,680 deviant or inappropriate sexual behaviors that had been disclosed.

That number was somewhat skewed by one individual who had adamantly maintained he had sexually abused only one daughter on three occasions throughout the entire time the individual was in treatment. Since the polygraph disclosure process came into use, the therapist related that this individual increased this disclosure to 7,500 occasions that he had at least 1,000 different victims. The point stressed in this example by therapist Dr. Humbart was that this man had been in treatment for several months and had previously completed all the commonly utilized screening devices available to that therapist. Moreover, throughout that period continued to maintain he only had one victim.

Treatment Team Approach

Successful sex offender treatment programs are those designed as a combined effort between the courts, probation, polygraph and treatment. The courts provide the authority to implement by judicial order, probation, treatment and polygraph monitoring. The prosecuting attorney's office is important to this team because of their concern for both the law and the protection of the communities to which sex offenders are returned. The probation officer maintains intensive contact with the

offender providing authority and guidance. (In some counties in Indiana, it is not uncommon to have probation officers attending and participating in the group sessions along with the therapists and those on probation.) The polygraph examiner is that member of the team who provides information, a symbol of deterrence, and has a foremost role in protecting society during the treatment and probation phase. The treatment provider team member has the major workload in this effort since therapy is the ultimate purpose in the process. He/she is the probationer's education provider, the confronter, the guide, and the instructor for alternative actions and thoughts. This is all too great for one entity alone. A team effort of involvement and cooperation is crucial.

Disclosure of new offenses during testing requires an agreement among probation, prosecution and judicial authorities. In some jurisdictions new offense disclosures can result in additional charges for that disclosed offense. These types of legal issues raise the question of granting limited immunity from prosecution to offenders who disclose new crimes. Jurisdictions vary regarding immunity policies. Some jurisdictions, like Colorado and some counties in Indiana, do not automatically offer limited immunity, but prosecutors make thoughtful decisions about further prosecution on a case-by-case basis. Decision makers in one jurisdiction have concluded that to prosecute all reported offenses would infringe on Fifth Amendment rights, and others point out that if the probationer is advised of the possibility of new charges, he/she would not be prone to make any new disclosures, despite its benefit to treatment. Yet another grants limited immunity for similar past offenses if the offender meets several containment conditions, including actively participating in an approved treatment program, pleading guilty, and gaining employment that meets the approval of the probation/parole officer. The positive side of such disclosures is that victims can be sought out to see if they are in need of therapy or other assistance, and the treatment provider is afforded the benefit of valuable information to assist in the offender's therapy.

Does Polygraph Detection Of Deception Work?

Social reality, being largely a matter of perspective, is different for each person. Accordingly, there is no one "truth" that can serve as the base for investigators to pinpoint. However, the polygraph examiner is not conducting tests to find the ultimate truth. The task in polygraph testing is much more basic. Polygraph testing allows the polygraph examiner to form an opinion on whether the examinee is being truthful or deceptive concerning his/her perception of a particular event. As a hypothetical example, the examinee is accused of fondling the vagina of a sleeping girl. The child legitimately has no memory of the act. The accused party being examined is asked a question during the course of a polygraph test, "Did you put your hand on X's vagina?" Assuming this scenario is true, the examinee's reality is that he knows that he did put his hand on the child's vagina. Whether the placing of the hand on the vagina was deviant, immoral, harmful, and/or illegal is a subjective matter determined by social mores, laws and even politics. The "ground truth" is a complex mixture of all of these processes and the polygraph test can play only a small (but important) part in the examination of the issue.

It is true that some child sexual abusers are able to rationalize and use denial so that they have no guilty feelings for a particular inappropriate sexual event. Indeed, some have gone beyond feeling that they caused no harm to the belief that the victim was the instigator and even enjoyed the episode. However, the success of polygraph testing does not depend upon the examinee feeling

guilty about placing his/her hand on the child's vagina or penis. The polygraph instrument does not measure guilt. Rather, it records psychophysiological responses concomitant to deception. The examiner and the examinee may interpret the sexual behavior in different ways (the examiner may place moral and legal implication on the act, while the examinee may not.) However, the question determines whether the examinee had a hand on the genitals, not how he/she feels about having done so.

Validity and Reliability Findings

There is a considerable body of research demonstrating the reliability and validity of polygraph testing. Research on the reliability and validity of polygraph testing should and does continue. However, as in other disciplines, there are pragmatists who continue to work and to do the basic research in the field while the methodologists wrestle with the methodology in the lab. Methodologists play a valuable role in any discipline by being the "professional unbelievers." Using the analogy of "pragmatists and unbelievers" is helpful in preparing for a short review of the rather vast literature on the validity and reliability of polygraph testing. Both types of research are important to the understanding and acceptance of polygraph testing.

Some methodologists would argue that the subject matter of detection of deception is simply too complex to study from a scientific viewpoint. The study of the validity and reliability of polygraph testing is no more or less complex than the study of any interaction among humans. The polygraph examiner conducts tests to determine if examinees are being truthful (as they perceive the truth) in response to questions designed to clarify an issue in dispute. The polygraph test yields recorded data on the physiological responses of the examinee. The physiological data are recorded as permanent tracings. The data are available to researchers along with the questions asked, the answers given by the examinee to each question, and the examiner's opinion or conclusion formed soon after the examination. One can study the results of polygraph testing done under field conditions (for example, law enforcement agency examinations of criminal suspects done as an aid to general criminal investigation), or polygraph tests can be conducted in the laboratory by way of experimentation. While experimental laboratory polygraph testing has the advantage that one can determine with some precision whether the person administering the test was right or wrong, some researchers note that laboratory findings cannot easily be generalized to field conditions (Barland & Raskin, 1976).

In settings where a large volume of polygraph testing is accomplished (a large commercial enterprise or government agency), there is an adequate supply of completed work from which researchers can draw samples. In each polygraph examination, the examiner has recorded a conclusion after study of the physiological data. The accepted outcomes are: (a) Non-Deception Indicated (NDI), (b) Deception Indicated (DI), (c) Inconclusive (INC) – the examiner is unable to reach a decision of DI or NDI from the available data. The methodologist researching the question of how well polygraph testing works has some interest in the "inconclusive" cases and intense interest in the examinations resulting in NDI or DI opinions by the examiner. It is those cases that allow the researcher to categorize opinions as right or wrong when independent verification can be obtained. (For example the examinee confesses to the crime, or someone else confesses to the crime). Stan Abrams, Ph.D., estimates that 60% of these tests can ultimately be verified independently of the polygraph testing (1973).

To present some perspective on the number of polygraph tests

available for study, one can look to the U.S. Army Criminal Investigation Division (CID). The Director of the U.S. Army Crime Records reported that the CID conducts about 1,600 to 1,700 polygraph tests each year (Hardy, 1994). If we use Dr. Abrams' estimation that 60% of these tests will ultimately have the polygraph examiner's decision verified (or discredited), this agency by itself is adding some 1000 cases per year to this stockpile of data available to researchers. The CID, as a matter of standard operating procedure, uses a quality control program where the data from each examination are independently reviewed by a senior polygraph examiner. (Most federal agencies use this form of quality control as well). This review of polygraph test charts by a person who does not know the facts of the case is called "blind" chart analysis. Blind chart analysis is not regarded so much as a test of validity, but more a reasonable measure of reliability. Quality control does give the agency a good estimate of the reliability of polygraph testing at any given point in time. For an assessment of validity, agencies such as the Army CID could provide historical data for which independent verification could be sought.

There are many well-known validity and reliability studies of polygraph testing (Barland & Raskin, 1976; Edwards, 1981; Elaad & Schahar, 1985; Kirby, 1981; Kleinmuntz & Szucko, 1984; Rafky & Sussman, 1985; Weaver, 1980; Wicklander & Hunter, 1975; Yamamura & Miyake, 1980; Yankee, Powell & Newland, 1985). A meaningful approach to looking at validity and reliability of polygraph testing is to evaluate only those studies published in recent years. One summary of studies done since 1980 was compiled by Ansley (1990), with the following results:

> Examiner decisions in these studies were compared to other results such as confessions, evidence, and judicial disposition. The ten studies reviewed considered the outcome of 2,042 cases, and the results, assuming every disagreement was a polygraph error, indicate a validity of 98%. For deceptive cases, the validity was also 98%, and for non-deceptive cases, 97%. The studies were from police and private cases, using a variety of polygraph techniques, conducted in the United States, Canada, Israel, Japan and Poland. (P. 169).

One type of research on polygraph testing focuses on interrater reliability, or the degree of agreement among different scorers of the same polygraph data (Ansley 1990). Today's model of chart analysis requires numerical scoring of physiological data recorded during the polygraph test. Uniformity in test formats allows other polygraph examiners to score the data on a particular case without knowledge of extra-polygraphic details. Any subjectivity introduced into the test evaluation, such as the examinee's difficult, surly, charming or cooperative disposition, is effectively removed. When confirmation of the outcome is added to the blind analysis of charts, one then has measures of both reliability and validity.

It is curious that studies appear to show that blind evaluators tend to perform slightly less well than the original examiners. Although blind evaluators have good scoring percentages, original examiners usually make more correct decisions, and fewer inconclusive calls. In the Ansley (1990) summary of ten recent studies, where blind evaluations were done on confirmed cases, the blind evaluators were correct in 89% of 218 polygraph tests, while the decision of the polygraph examiners who had originally done the same 218 tests

were correct 95% of the time. This compilation of research studies excludes the inconclusive outcomes so that only tests with confirmed NDI or DI outcomes are considered. The studies were completed by a variety of capable researchers (Arellano, 1990; Elaad, 1985; Franz, 1989; Hontz & Driscoll, 1988; Honts & Raskin, 1988; Jayne, 1990; Matte & Reuss, 1989; Patrick & Iacono, 1987; Raskin, Kircher, Honts & Horowitz, 1988).

Validity Studies

Tables 1 and 2 below outline summaries of the nine out of the ten studies where NDI and DI decisions were separated. A Polish study (Widacki, 1982) is omitted. It showed an overall accuracy of 92%, but did not separate the NDI and DI decision results.

Table 1
Summary of Testing Accuracy on Confirmed "No Deception Indicated" (NDI) Tests

Authors and Date	Number of Confirmed NDI Tests Studied	Correct NDI Decisions	Percentage Decisions Correct
Arellano, 1990	18	18	100%
Edwards, 1981	363	356	98%
Elaad & Schahar, 1985	100	95	95%
Matte & Reuss, 1989	54	54	100%
Murray, 1989	21	18	86%
Patrick & Iacono, 1987	30	27	90%
Putnam, 1983	65	62	95%
Raskin *et al.*, 1988	28	27	96%
Yamamura & Miyake, 1980	65	61	96%
Totals:	744	718	98%

Table 2
Summary of Testing Accuracy on Confirmed "Deception Indicated" (DI) Tests

Authors and Dates	Number of Confirmed DI Test	Correct DI Decisions	Percentage Decisions Correct
Arellano, 1990	22	22	100%
Edwards, 1981	596	587	98%
Elaad & Schahar, 1985	74	73	99%
Matte & Reuss, 1989	60	60	100%
Murray, 1989	150	150	100%
Patrick & Iacono, 1987	51	51	100%
Putnam, 1983	220	219	99%
Raskin *et al.*, 1988	57	54	95%
Yamamura & Miyake, 1980	30	24	98%
Totals:	1,250	1,240	98%

Tables adapted from Ansley (1990, p. 171).

A Word About Admissibility

Admissibility of not only polygraph evidence, but scientific evidence in general, has been governed in the federal courts for over half a century by *Frye v. U.S.* (1923). *Frye* excluded expert testimony that was based on use of a simple and crude precursor to the modern polygraph testing process on the ground that expert opinion based on a scientific technique would be inadmissible until the technique is "generally accepted" as reliable in the relevant scientific community.

With few exceptions, such as *U.S. v. Piccinonna* (1989), *Frye* kept polygraph evidence out of the federal trials for about fifty years, despite an expanding understanding of psycho-physiological principles underlying the modern polygraph and the extensive development of validated techniques. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, (1993), the U.S. Supreme Court finally put the *Frye* ruling to rest. A Ninth Circuit Court opinion had upheld the exclusion of expert scientific testimony linking birth defects to an anti-nausea drug. The Ninth Circuit based its decision on *Frye* and the fact that the testimony of the expert was not based on principles sufficiently established to have general acceptance in the relevant scientific community.

The Court stated that although the *Frye* "general acceptance" test had historically been the dominant standard for determining the admissibility of novel scientific evidence at trial, the Federal Rules of Evidence 702 superseded *Frye* and now governs admission of expert testimony. The rigid "general acceptance" requirement would be at odds with the "liberal thrust" of the Federal Rules, including the "general approach of relaxing the traditional barriers to 'opinion' testimony". The Court added that its decision does not mean that there are no limits on the admissibility of purportedly scientific evidence. Indeed, the Court referenced numerous criteria which may assist the trial court in the exercise of its discretion.

The most thorough published exposition of the applications of the *Daubert* factors to polygraph evidence in particular is found in a case from the U.S. District Court for the District of New Mexico. In *U.S. v. Galbreth*, (1995), the court conducted an exhaustive review of the *Daubert* factors in the context of a very complete evidentiary record and ruled that the defense polygraph evidence should be admitted at trial. Other courts have also ruled that polygraph evidence may meet the *Daubert* standards. Recently, the Ninth Circuit Court reached the same conclusion in *U.S. v. Cordoba* (1997), however, this decision was reversed by the United States District Court, Central District of California.

Summary

Any professional in any discipline can make an error on occasion, including polygraph examiners. However, polygraph testing is able to produce high quality accurate information quickly for other professionals working with sexual offenses. In doing so, it can deter child sexual abuse to the benefit of everyone.

References

Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Science*, 18, 313-326.

Abrams, S., & Abrams, J.B. (1993). *Polygraph testing of the pedophile.* Portland, OR: Ryan Gwinner Press.

Ansley, N. (1990). The validity and reliability of polygraph decisions in real cases. *Polygraph*, 19, 169-181.

Arellano, L.R. (1990). The polygraph examination of Spanish speaking subjects. *Polygraph*, 19, 155-156.

Baranowski, G.H. (1995). *Polygraph monitoring of individuals on probation for child sexual abuse.* Paper presented at the 30th Annual Seminar and Workshop of the American Polygraph Association, Las Vegas, NV.

Barland, G.H. & Raskin, D.C. (1976). *Validity and reliability of polygraph examinations of criminal suspects.* (Report No. 75-NI-99-001). Washington, D.C.: National Institute of Law Enforcement and Criminal Justice.

Brown, R.H. (1977). The emergence of existential thought: Philosophical perspectives on positivist and humanist forms of social theory. In J.D. Douglas & J.M. Johnson (Eds.) *Existential Sociology* (pp. 77-100). New York: Cambridge University Press.

Collins, R. (1981). On the micro-foundations of macro sociology. *American Journal of Sociology*, 86, 984-1014.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993), 509 U.S. 579.

Edwards, R.H. (1981). A survey: Reliability of polygraph examinations conducted by Virginia polygraph examiners. *Polygraph*, 10, 29-272.

Elaad, E. (1985). *Validity of the control question test in criminal cases* (Technical report). Jerusalem, Israel: Criminal Investigation Division, Israel National Police Headquarters.

Elaad, E., & Schahar, E. (1985). Polygraph field validity. *Polygraph*, 14, 217-223.

Franz, M.I. (1989). *Technical report: Relative contributions of physiological recordings to detect deception* (DOD contract MDA904-88-M-6612). Atlanta, GA: Argenbright Polygraph, Inc.

Frye v. U.S. (1923) 293 Fed. 1013 D.C. Cir.

Hardy, L.H. (1994, July). *Polygraph testing of victims: Focus on sex crime victims.* Paper presented at the 29th Annual Seminar and Workshop of the American Polygraph Association. Nashville, TN.

Honts, C.R. & Driscoll, L.N. (1988). A field study of the Rank Order Scoring System (ROSS) in multiple issue control question test. *Polygraph*, 17, 1-15.

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.

Jayne, B.C. (1990). The relative contributions of physiological detection of deception. *Polygraph*, 19, 105-117.

Kirby, S.L. (1981). The comparison of two stimulation tests and their effect on the polygraph technique. *Polygraph*, 10, 63-76.

Kleinmuntz, B. & Szucko, J.J. (1984). A field study of the fallibility of polygraph lie detectors. *Nature, 308, 449-450.*

Matte, J.A. & Reuss, R.M. (1989). A field validation study of the quadri-zone comparison technique. *Polygraph*, 18, 187-202.

McGraw, J.M. & Smith, H.A. (1992). Child sexual abuse allegations amidst divorce and custody proceedings: Refining the validation process. *Journal of Child Sexual Abuse*, 1, 49-62.

Murray, J.K. (1980). The polygraph technique: Past and present. *FBI Law Enforcement Bulletin.*

Murray, K.E. (1989). Movement recording chairs: A necessity? *Polygraph*, 18, 15-23.

Patrick, C.J. & Iacono, W.G. (1987). Validity and reliability of the control question polygraph test: A scientific investigation. *Psychophysiology*, 24, 604-605.

Practicing Law Institute. (1991, April). *National conference on sentencing advocacy: 1991.* Washington, D.C.: Author.

Putnum, R.L. (1983). *Field accuracy of polygraph in the law enforcement environment.* Unpublished manuscript, American Polygraph Association archives.

Rafky, D.M. & Sussman, R.C. (1985). Polygraph reliability and validity: Individual components and stress of issue in criminal cases. *Journal of Police Science and Administration*, 13, 282-292.

Raskin, D.C., Kircher, J.C., Honts, C.R. & Horowitz, S.W. (1988). Validity of control question polygraph tests in criminal investigations. *Psychophysiology*, 25, 474.

Rock v. Arkansas (1987), 483 U.S. 44.

U.S. v. Cordoba (1997) 104 F.3d 225, 9th Cir.

U.S. v. Piccinonna (1989) 885 F.2d 1529, 11th Cir.

U.S. v. Posado (1995) 57 F.3d 428, 5th Cir.

U.S. v. Galbreth (1995), F.Supp. 877 D.N.M.

Weaver, R.S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, 9, 94-108.

Wicklander, D.E. & Hunter, F.L. (1975). The influence of auxiliary sources of information in polygraph diagnosis. *Journal of Police Science and Administration*, 3, 405-409.

Widacki, J. (1982). *Analiza przestanek diagnozowania w. badanich poligraficznuych.* Univwersytetu Slaskiego, Katowice.

Williams, V.L., Morrison, J. & Terrell, J. (1993). Advocation alternative sentencing: A polygraph examiner role in the political economy of alternative sentencing. *Polygraph*, 22, 150-163.

Williams, V.L. (1995). Legal issues: Response to Cross and Saxe's "A Critique of the Validity of Polygraph Testing in Child Sexual Abuse Cases". *Journal of Child Sexual Abuse*, 4, 55-71.

Yamamura, T. & Miyake, Y. (1980). Psychological evaluation of detection of deception in a riot case involving arson and murder. *Polygraph*, 9, 170-181.

Yankee, W.J. (1994). A case for forensic psychophysiology and other changes in terminology. *Polygraph*, 23, 188-194.

Yankee, W.J., Powell III, J.M. & Newland, R. (1985). An investigation of the accuracy and consistency of polygraph chart interpretation by inexperienced and experienced examiners. *Polygraph*, 14, 108-117.

* * * * * *

A Guide to Department of Defense Polygraph Institute Research Interests

Andrew B. Dollins

Editor's note:

The present paper outlines the polygraph research plans for the Department of Defense Polygraph Institute, and is provided for information purposes to polygraph practitioners, as well as to researchers. For further information, contact Dr. Dollins at the Department of Defense Polygraph Institute, Building 3195 at 13th Street, Ft. McClellan, AL 36205-5114, Tel: 254-848-3803, or visit the DoDPI World Wide Website at www.dodpoly.org.

Keywords: Acquaintance test, automated testing, cardiovascular, directed lie, drugs, PDD, polygraph, probable lie, question formulation, research, semantics

Introduction

The Department of Defense Polygraph Institute (DoDPI) anticipates that the DoDPI research budget will be increased during FY99. In anticipation of this increase, DoDPI encourages investigators to submit research proposals addressing the topic areas below. Formal solicitation of this research requirement was originally published in the *Commerce Business Daily*, by the Office of Naval Research, under the Broad Agency Announcement entitled Defense Personnel Security Research Center, on 24 July 1997. Proposals submitted should be made in response and in accordance with this BAA.

The Institute funds thesis, dissertation, and institutional research awards. The maximum award amount for a masters degree thesis is $5,000 per student. The maximum award for a dissertation thesis is $15,000 per student. The maximum amount for institutional awards is $150,000 per project. Institutions are eligible to receive simultaneous awards for different projects. Awards are for a one-year period. The Institute will support multi-year projects by funding the first year and making decisions on subsequent year funding dependent on the (a) Institute's evaluation of current year deliverables (usually a report) and (b) availability of funds. Proposals should be submitted to the Defense Personnel Security Research Center (PERSEREC). Telephone, mail, and electronic mail inquiries about projects are welcome. Points of contact are provided at the end of this document. It is strongly suggested that applicants obtain a copy of the pamphlet *Personnel Security Thesis, Dissertation, and Institutional Research Awards* from Claire Rigg, Howard Timm, or Andrew Dollins. This pamphlet describes the application procedure and the information to be included in the proposal. Written requests for the pamphlet should include a self-addressed mailing label.

Overview

The Institute's research mission is to: (a) evaluate the validity of psychophysiological detection of deception (PDD) techniques used by the Department of Defense (DoD); (b) investigate countermeasures and anti-countermeasures; and (c) conduct developmental research on PDD techniques, instrumentation, and analytic methods. While the Institute will evaluate all research proposals within our mission objectives, those which address the topics identified below will receive priority. Proposals to use procedures similar to those currently

in use by Federal agencies that conduct PDD examinations will receive priority over proposals which use other, less applicable, procedures. While the Institute is interested in, and will support, basic research, the majority of our funding will be awarded to proposals describing applied research that is of immediate use to the PDD community. The research topics of interest have been divided into the categories of Applied Topics, PDD Data Analyses, Deterrence, and New Technology for the sake of convenience.

Applied Topics

Standard Scenario Development

The objective of this project is to develop sets of subject manipulation procedures which can be used, in the laboratory, to reliably and accurately manipulate and test subject veracity. The goal is to correctly determine the veracity of at least 80% of the subjects tested, including subjects for which no opinion could be rendered (which should be less than 10% of the total subjects tested). In addition, the standard scenario should be as portable as possible. That is, the results should be reproducible in other laboratories using the same procedures. Priority will be given to proposals which use psychophysiological detection of deception (PDD) techniques currently used by Federal PDD examiners (*e.g.*, Zone Comparison Test, You Phase, Modified General Question Test, Relevant/Irrelevant Test). The subject manipulation procedures will be used as a standard for testing other applied questions, such as those identified below.

Effects of the Cardiovascular Auscultatory Cuff

An auscultatory cuff (*i.e.*, blood pressure cuff) is commonly used to collect cardiovascular data during a PDD examination. The cuff is normally fastened around the upper arm. Before the introduction of electronic amplifiers the cuff was inflated to around 90mmHg (millimeters of mercury) while a series of questions were asked during the PDD examination. Due to the advent of electronic amplifiers, the average pressure has been reduced to 60mmHg. The discomfort caused by the cuff, however, remains a limiting factor in the number of questions which may be asked in a single series during a PDD examination. Some believe that the discomfort of the cuff enhances the deception detection potential of the examination. Still others believe that the cuff is entirely unnecessary. The objective of this project is to determine if the auscultatory cuff itself influences the accuracy of the examination. If it does not influence the accuracy of the examination procedure, it could be replaced by an alternative sensor. Without the cuff, the examination length could be increased and additional questions asked--possibly improving accuracy by supplying more data.

Usefulness of Acquaintance Test

A procedure purported to demonstrate that a PDD examination works, commonly referred to as a stimulation test, stim test, numbers test, or acquaintance test is used during most PDD examinations. A variety of specific techniques fall into this category, each of which is intended to stimulate the examinee and demonstrate the utility of the PDD instrument. Acquaintance tests may also be used as a practice test for the examinee--to obtain a normal physiological pattern, or to orient the examinee to the setting. The acquaintance test is usually completed before collecting data concerning the actual questions of interest, or between the first and second series of questions. Studies have not determined whether the acquaintance test is better before the actual examination or following the first question series. The objective of this project is to

determine what, if any, influence the acquaintance test has on the examination.

Compound Questions

A relevant compound question is one that addresses two distinctly different issues, such as, "Did you steal that money or that ring?" The DoDPI policy is that use of compound questions is inappropriate because such questions are confusing to examinees--and could interfere with physiologic reactions if only part of the question is meaningful to the examinee. Compound questions are, however, routinely used by some agencies during screening examinations. We have been unable to locate any reports of controlled systematic studies investigating the use of compound questions. The Institute is interested in supporting controlled systematic studies designed to determine if the use of compound questions influences examination accuracy.

Caveats

During a pretest interview examinees will frequently admit to transgressions which are not relevant to the purpose of the PDD examination. Admissions may be related to the primary topic of the examination (*i.e.*, an examinee accused of using heroin may have used marihuana, or an examinee accused of robbing a bank may have robbed a convenience store several years ago) or related to the comparison questions (*i.e.*, an examinee may admit to stealing from a relative when asked, "Have you ever stolen anything of value?") In both cases, the examiner must attempt to isolate the topic of interest from the admission. This is usually done by adding a qualifying phrase, or caveat, to the original question. For example, "Have you ever stolen anything of value?" could be changed to "Other than what you told me, have you ever stolen anything of value?" The most common caveats are "Other than what you told me ..." and "Before ... (some specific time or age), did you ever ..." We have located only two controlled systematic studies investigating the effectiveness of caveats used with PDD examination questions. The results of one study suggested that caveats are effective, while results of the other indicated caveats were not effective. The Institute is interested in supporting further research regarding the effectiveness of caveats.

Directed *versus* Probable Lie Comparison Questions

Most PDD examination procedures require that the examiner ask the examinee relevant and comparison questions. Relevant questions address the main focus of the examination (*e.g.*, "Did you steal that money?" "Did you break that lock?") Comparison questions are designed to address an issue that is similar in nature, but unrelated to the main focus of the examination (*e.g.*, "Have you ever cheated a friend out of anything?" "Before your last birthday, did you ever lie to anyone in authority?") Responses to relevant questions are evaluated by comparing them to responses to comparison questions. The most commonly used comparison questions are labeled probable lie questions. The examiner will compose the probable lie comparison questions using information obtained during the pretest interview. The examiner is not absolutely certain the examinee's response to these questions is a lie, hence the label probable lie. An alternative comparison question, labeled a directed lie question, has been proposed and is currently used during some screening examinations. Examinees are simply instructed to be deceptive to a question and directed to think about, and imagine, an uncomfortable personal incident related to the asked question. The examiner never knows what the lie was or what the examinee is thinking. Directed lie questions are easier to use because they do not require an extensive interview with the

examinee. It has not, however, been verified that a specific issue examination using directed lie comparison questions is as effective or accurate as one using probable lie comparison questions. The Institute is interested in supporting studies designed to compare the efficacy of probable and directed lie comparison questions.

Stimulation Between Tests

During the data collection phase of a typical PDD examination, the same series of questions is repeated two or three times with a brief rest between series. During the rest time, the pressure in the auscultatory cuff is released and the examinee is instructed to relax. Some believe that questioning the examinee between question series enhances the examinee's physiologic reactivity, others believe questioning an examinee between question series is unethical and manipulative. The Institute would like to support controlled systematic investigations regarding the question of stimulation between PDD tests.

Question Phrasing

Composing the questions to be asked during a PDD examination can be a major task in itself. Approximately 25 hours of the DoDPI introductory course are dedicated to teaching principles of question construction. It is believed by some, that changing a single relevant word in a question, such as changing the word steal to take (*i.e.*, "Did you steal that money?" to "Did you take that money?") can influence the results of the examination. Still others believe that the syntactical complexity and the number of words in a question can influence the results of an examination. The objective of this project is to determine if changes in question phrasing can influence the accuracy PDD examination results.

Question Order/Sequence

A variety of PDD examination question formats are currently in use. It is anecdotally suggested that changing the position or order of questions asked during the intest can influence the accuracy of the examination. We have found little evidence of systematic controlled investigations regarding the influence of changing the sequence of questions asked during a PDD examination. Proposals to investigate the influence of question presentation order on PDD examination accuracy are of interest to the Institute.

Drug Effects

There are very few reports in the literature regarding the effects of drugs on PDD examination accuracy. This obvious important research genre is an investigational priority for the DoDPI. As a first step toward determining the effects of drugs on PDD examinations, the Institute is soliciting proposals from experts in psychopharmacology, microbiology, and related fields to develop a well-documented prioritized list of substances to be investigated. The Institute is further interested in proposals to investigate the effects of those substances.

Arousal Level *versus* Accuracy

Current computer programs designed to evaluate PDD examination results weigh activity from the electrodermal channel quite heavily. Computerized polygraph developers report that activity from the electrodermal channel accounts for approximately 50% of subject veracity predictability. During manual evaluation, the pneumographic, cardiovascular, and electrodermal information is weighed equally, but anecdotal evidence suggests that examiners score the electrodermal channel more easily than the other two channels. Skin conductance (an electrodermal measure) research has

shown that response amplitude increases as tonic (resting) skin conductance increases. It may be possible to reduce the error and no opinion rate by determining which examinees are likely to have pronounced electrodermal responses. The objective of this project is to determine if there is a relationship among subjects tonic skin conductance level, the amplitude of their responses recorded during a PDD examination, and accuracy of veracity detection. Proposals to investigate the relationship between other measures of arousal and the detection of deception would also be of interest. It is anticipated that this research will be exploratory.

Recorded Voice

The examiner's question asking technique is considered to be a factor influencing the accuracy of a PDD examination. Changing the volume, tone, or pitch of the voice used to ask the examination questions could influence PDD examination results, as could changing the speed used to pronounce specific words. While each of these factors could and possibly should be investigated, another solution is to ensure that the questions are repeated in exactly the same manner throughout the examination by recording and replaying them. There is, however, a possibility that examinees react differently to questions asked via recording *versus* an actual human. The objective of this project is to determine if using recorded questions influences examination accuracy.

Pretest Content Analysis

The pretest portion of the PDD examination is believed to be very important because it influences examinee's understanding of, and reactions to, the examination procedures. If a pretest is done well, the examination process is explained, the examinee is put at ease, unnecessary anxiety is alleviated, the examination questions are thoroughly explained and their meaning is made clear to the examinee. In addition, the examiner gains information which can be used to more precisely formulate examination questions--and possibly further interview the examinee. The pretest should also provide a basis for the examiner to determine if the examinee is physically and mentally suitable for the examination. Anecdotal evidence suggests that the examination results are greatly influenced by the examiner's pretest expertise. There has been little formal investigation into the relationship between information reviewed during the pretest and PDD examination accuracy. The objective of this research is to determine if there is a relationship between the content of the pretest interview and the PDD examination results. The first step in this project will be to analyze the content of pretests which were recorded under controlled conditions, and for which the ground truth of subsequent PDD examinations is known, to determine if the pretest discussion influences examination accuracy in a predictable manner. The Institute can supply tape recordings of pretests for this project.

Test Format Comparison

A specific procedure and order for asking questions during a PDD examination is referred to as a question or test format. For example, a Zone Comparison Test format usually includes three comparison questions and three relevant questions which address the same issue. Variations on the Zone Comparison format include only two comparison questions, two relevant questions, or both. Other test formats can include one, two, three, or four relevant questions and two, three, or four comparison questions. In addition, the relevant questions may address the same or multiple issues. Additional variations on specific issue and screening test formats exist and there is little, if any, evidence supporting one method in favor of another. The objective

of this research is to determine if PDD test format variations influence the accuracy of decisions regarding examinee veracity.

Statement Content/Honesty Test Analysis

Information obtained during the pretest interview is used to formulate and refine PDD examination questions. Examiners currently base their pretest interview on information gathered during the investigative process that precedes the examination. One possible method of improving the pretest is to require examinees to provide a written response to specific questions well before the PDD examination, then use statement content analysis, honesty testing, or both techniques to identify information that the examinee may be reluctant to disclose. The information gathered during the analysis could then be used to assist, and possibly enhance, the question formulation process. The Institute is interested in proposals to systematically investigate the use of statement content analysis, honesty testing, and similar techniques to improve the accuracy of the PDD process.

PDD Data Analyses

The physiologic data recorded during a PDD examination has traditionally been evaluated visually by trained examiners. Examiners look for changes in the amplitude, duration, or frequency of respiratory, cardiovascular, and electrodermal (*i.e.*, skin resistance or conductance) activity following the presentation of a question or statement that was clearly defined and explained during the examination pretest. These changes are referred to as responses, features, or evaluation criteria. Usually the responses are identified, a numerical value is assigned by comparing responses to one question with responses to at least one other question, and the numerical values are combined to produce a score from which a decision regarding the subjects' veracity is made.

The evaluation rules and procedures may differ according to the type of PDD examination used, the policy of the agency sponsoring the examination, and the examiner's training, skill, and experience. Preliminary evidence further suggests that there may be discrepancies in how multiple PDD examiners score the same information (called blind scoring). These discrepancies can result in decreased accuracy in determining examinee veracity. The objective of the PDD Data Analysis category is to improve the consistency, reliability, and accuracy of PDD data analysis techniques.

PDD Feature Analysis

The objective of this multipurpose, multiphase, project is to: 1) determine what criteria examiners actually use to evaluate PDD data, 2) determine which of the used criteria are predictive of deception, and 3) investigate methods of combining the most predictive criteria to achieve the highest possible accuracy rate. The DoDPI currently has data from over 800 actual examinations for which examinee veracity is known. These data will be made available for this analysis. The amount of available data will increase with time because the data collection effort is ongoing. It is anticipated that the successful proposal will be multiphase with progress reports submitted at the end of each phase.

Does Artifact Removal Improve Accuracy?

The raw data collected during a PDD examination typically contains artifacts such as centering adjustments and off-scale tracings. If the data were collected and

recorded using a computerized polygraph instrument, these artifacts can be removed, and the data displayed without the artifacts. The objective of this project is to determine if the accuracy of visual data evaluation is influenced by removing artifacts from the data prior to scoring.

Deterrence

It has been suggested, and is believed by individuals within and outside of the federal government, that PDD testing has a deterrence effect. That is, requiring employees to undergo periodic and aperiodic PDD examinations will discourage or prevent them from engaging in undesirable and illegal activities, such as theft or disclosure of sensitive information. While this may be a valid supposition, there is little supporting evidence. The objective of this project is to determine if the knowledge that a PDD examination is imminent will deter subjects from engaging in undesirable activities. One possible approach to answering this question is to examine the frequency of transgressions that occurred in similar companies prior to, during, and after the Employee Polygraph Protection Act of 1988. Another possibility would be to devise a laboratory procedure to test the deterrence effect of an imminent polygraph examination.

New Technology

The physiological measures recorded during a PDD examination have changed little during the last 20 to 40 years. The Institute would like to promote the investigation of alternative measures, technology, and analyses techniques. Among the new measures which have been suggested are: Voice stress, thermal imaging, pulse transit time, visual activity (*i.e.*, pupil dilation and eye movement), electroencephalography, electromyography, electrocardiology, and vagal tone. The objective of this research is to identify and test new measures, technology, and analyses which could be used to detect deception. It is important that proposals in this area clearly identify and support a nexus between the proposed work and the detection of deception.

Points of Contact

Direct inquiries related to research issues and topics to:

Andrew Dollins
DoDPI
13th St., Building 3195
Ft. McClellan, AL 36205-5114
Tel: 256-848-3803
Fax: 256-848-5332
DOLLINSA@MCCLELLAN-DPI.ARMY.MIL

Direct inquiries related to proposal submission and project monitoring to:

Howard Timm
Tel: 408-656-5027
or
Claire Rygg
Tel: 408-656-5024
The Personnel Security Research Program, Defense Personnel Security Research Center
99 Pacific St., Bldg. 455-E
Monterey, CA 93940-2481
Fax: 408-656-2041
Timm: TIMMHW@OSD.PENTAGON.MIL
Rygg: RYGGCM@OSD.PENTAGON.MIL

Direct inquiries related to contractual issues to:

Brian Glance
Office of Naval Research
Code 252
Ballston Centre Tower 1
800 North Quincy St.
Arlington, VA 22217-5660
Tel: (703) 696-2596
FAX: (703) 696-0993
glanceb@onrhq.onr.navy.mil

Formant Structure of Vowels Spoken During Attempted Deception

Dean A. Pollina

Douglas A. Vakoch

Lee H. Wurm

Abstract

Changes in the formant structure of vowels were examined within the terminal words of sentences spoken during attempted deception. Subjects read three types of sentences related to a mock crime. Sentences were 1) believed to be true, 2) believed to be false, or 3) unlikely but possibly true. Subjects were instructed to do their best to deceive the lie detector. Formants extracted from sentences believed to be false were higher in amplitude (formants 1 and 3) and lower in frequency (formant 2) than formants extracted from the same vowels within identical sentences spoken when that sentence's truth value was uncertain. These results suggest that emotional arousal experienced during the act of deception can cause subtle changes in the quality of the speaker's voice.

Keywords: Voice stress

Dr. Pollina is with the State University of New York at Stony Brook. Dr. Vakoch is with Vanderbilt University. Dr. Wurm is at the State University of New York at Binghamton.

Address all correspondence and request for reprints to Dean Pollina, Department of Neurology, Health Sciences Center T12-020, State University of New York at Stony Brook, Stony Brook, NY 11794-8121. E-mail: dpollina@neuro.som.sunysb.edu.

Formant Structure of Vowels Spoken During Attempted Deception

Speech has been shown to convey information about a speaker's affective state (Hecker, Stevens, von Bismarck, & Williams, 1968; Lippold, 1970; Simonov & Frolov, 1977; Tolkmitt & Scherer, 1986; Vakoch, 1994; Williams & Stevens, 1969). Emotions that increase arousal reportedly cause increases in the mean fundamental frequency (F_0) of the speaker's voice. These findings suggest that the increases in arousal that accompany deceptive communication should cause elevation of the F_0 of the deceptive statements. Several studies have reported data supporting this hypothesis (Ekman & Friesen, 1974; Streeter, Krauss, Geller, Olson & Apple, 1977). However, other studies investigating frequency modulations in the voice during attempted deception have failed to report any reliable effect (Horvath, 1978; Nachshon, Elaad, & Amsel, 1985).

There are several possible explanations for these discrepant findings. One potential difficulty results from the relatively poor sensitivity of traditional measures such as mean F_0 when used for classifying an individual subject's statements as either deceptive or truthful (Tolkmitt & Scherer, 1986). Simple measures of central tendency may miss small but reliable differences in the shape of individual subjects' speech waveforms that could be useful in the classification process. A second problem results from the use of less than optimal classification rules. For example, early studies are difficult to interpret because they are based on subjective chart evaluations made by an experimenter (Nachshon, Elaad & Amsel, 1985; Williams & Stevens, 1969). A third contribution to the variability has to do with the length of the utterances and subsequent speech waveforms subjected to analysis. The speech waveform will show a great deal of variability when the speech sample is large.

In the following study we explored the effect of varying the strength of subjective belief in propositions on speech waveforms that were recorded as the subjects stated these propositions. We used formant structure analysis in place of more traditional waveform measures such as the mean F_0 (formants are concentrations of energy in certain frequency ranges in a speech signal). Formant structure analysis of the steady-state portion of vowels has previously been shown to be sensitive to induced cognitive and emotional stress (Tolkmitt & Scherer, 1986). In the current study, digitized speech samples from the first stressed vowel in each relevant word were analyzed. It was anticipated that these measures would reveal differences in the structure of the speech waveforms elicited to uncertain statements spoken during different arousal conditions. The hypothesis tested was that the act of lying would cause the structure of the resulting speech waveforms to change in a specific, quantifiable way relative to similar speech samples obtained when a subject was telling the truth.

The paradigm has been tested recently using event-related brain potentials (ERPs), which are electric polarity shifts recorded at the scalp and thought to reflect cognitive processes such as attention (Pollina & Squires, in press). These researchers found that the structure of several ERP components changed as subjects viewed propositions whose truth values differed (*i.e.*, propositions that were either likely to be true, likely to be false, or unlikely but possibly true). The present study attempts to extend these results using a measure (voice quality) that can be influenced by the autonomic nervous system.

Method

Subjects

Twenty-five undergraduates from the State University of New York at Stony Brook participated. None of the subjects had any knowledge of the specific hypotheses under study. Each received experimental credit in an introductory psychology course upon completion of the experiment. All subjects were native speakers of English.

Group 1 consisted of seven females and six males, whose ages ranged from 18 to 22 (mean = 19.0). Group 2 consisted of six females and six males, whose ages ranged from 18 to 25 (mean = 19.9). Further information about the two groups is provided below.

Stimuli

Information about a fictional murder mystery was presented on a computer screen using a program developed for this study. A 79-word scenario provided information about the victim and the crime, and was followed by a list of nine suspects. For each suspect, biographical information was generated. This information consisted of the suspect's home town, birth state, year of birth, height, occupation, type of car owned, and financial worth.

Subjects were divided into two groups. The specific biographical information presented to subjects was the same for each suspect and did not depend on the subject's suspect choices. Prime suspect information for subjects in group 1 corresponded to the second-choice suspect for subjects in group 2. Second-choice suspect information for subjects in group 1 corresponded to the eliminated (from consideration) suspect for subjects in group 2. Eliminated suspect information for subjects in group 1 corresponded to the prime suspect information for subjects in group 2 (see Table 1).

Procedure

Each subject was asked to imagine that he or she was a police inspector in charge of a murder case. The subject studied computer files that contained information about nine suspects. The subject read the computer file on each of the nine suspects. For eight of the suspects, only information consistent with their being the murderer was presented. The ninth suspect (Bobby) had an "air-tight alibi," and subjects were told to eliminate Bobby from the list of suspects. After reading the suspect files, subjects filled out a "case report" (a written questionnaire), in which they were asked which suspect they believed to be the murderer and why. Next, subjects were told that if they were not absolutely sure that their prime suspect committed the crime, they should indicate a second-choice suspect.

Subjects were then presented with seven sentences listing information connected with their prime suspect. When subjects indicated that they had studied the information thoroughly, seven additional sentences were presented. These sentences listed information connected with the subject's second-choice suspect. After studying these sentences, a final set of seven sentences were presented. These sentences listed biographical information connected with Bobby, who was no longer considered capable of committing the crime. Subjects were not told that the information paired with their suspect choices depended only on group assignment (and not on the specific suspect choices they made).

After memorizing this biographical information, subjects were presented with the following information:

> Hello Inspector. Some time has passed since the investigation. You are still the police

chief, but some important events have transpired. For one thing, despite your beliefs about which suspect was guilty of the murder, you arrested Bobby! Although you knew him to be innocent, he had been a thorn in your side. You had always been sure that he was a member of organized crime, and was responsible for gun-running, racketeering, and money laundering operations in your district for years, yet you were never able to make even a single charge stick. Taking what you believed to be your big chance to be rid of Bobby once and for all, you arrested him and charged him with Alice's murder. Unfortunately for you, Bobby's defense attorney has not taken kindly to you. You have been accused of arresting a man whom you knew to be innocent of the crime. Under pressure from the court, you have agreed to take a lie-detector test. Your job will be to convince the lie detector that you believe Bobby to be the murderer. Under NO circumstances are you to reveal your true beliefs during the lie detector test. If you manage to fool the lie detector, you will be given a commendation for a job well done.

Next, subjects were asked to recall the biographical information they had previously learned. The method was a cued-recall test in which the experimenter orally presented the subject with a sentence fragment (*e.g.*., "Bobby was born in ...") and the subject completed the sentence. After performing this task to a criterion of one correct pass through all 21 biographical sentences (seven for each of the three suspects), the subject was led into the testing area and informed that the test was about to begin. Subjects were told that the test was an attempt to discover whether "voice patterns" could be used to detect deception and that a second experimenter would attempt to determine the subject's "true beliefs." They were also told that if they were able to defeat the lie detector test, they would receive a $20 prize.

Each subject was placed in a sound attenuating chamber and given three sheets of paper, each with seven biographical statements. Seven identical sentence beginnings were presented, in the same order, three times. The sentence endings consisted of biographical information that corresponded to information connected with that subject's prime suspect (on one sheet of paper), second-choice suspect (on a second sheet of paper), and the suspect (Bobby) who was presumed innocent (on a third sheet of paper). The subject was instructed to read each sentence aloud clearly and distinctly. The subject's voice was recorded as the biographical information was read.

This procedure resulted in sentences being read that fell into one of three truth-value categories: PROBABLY FALSE sentences were produced by the sentences believed to be false (biographical information associated with Bobby); PROBABLY TRUE sentences were produced by the statements that the subject believed most strongly--the strength of the subject's belief having been determined by his or her prime suspect in the case report; and finally, POSSIBLY TRUE sentences were produced by the statements corresponding to the subject's second-choice suspect. During the testing session, the six possible

presentation orders of PROBABLY TRUE, POSSIBLY TRUE, and PROBABLY FALSE sentence-type information were counterbalanced.

Formant Analysis

With a Technics Model AT9400 microphone (Matsushita Electric Industrial Company, Osaka, Japan), the responses of the subjects were recorded in analog form (onto a Technics Model RS-B18 tape recorder). The microphone was placed at a fixed distance (approximately 122 cm) from the subject. Recorded responses were then digitized off-line at a sampling frequency of 10 kHz and a filter cut-off frequency of 4800 Hz, using the Computerized Speech Research Environment (CSRE) software package, version 4.2 (AVAAZ Innovations, Inc., London, Ontario). Of the seven biographical facts recorded for each sentence-type, only the sentences stating the murderer's home city (The murderer is from Boston; The murderer is from Atlanta; The murderer is from Oakland) was digitized.

Next, the speech waveforms were edited so that only the stressed vowel in each of the three target words remained. Formant extraction from the vowels was performed using CSRE. For each vowel, separate formant structure analyses were performed on two successive time windows within each waveform; 0-25.6 msec and 25.6-51.2 msec. This frame width ensured that even the lowest male F_0 would fit into a single frame. In all cases, the stationary part of the vowel was longer than the resulting 51.2 msec analysis epoch. The points within each window were processed with a Hamming tapering window.

Results

Separate between-group MANCOVAs were performed on formant amplitudes and frequencies extracted from the stressed vowels within each of the three sentences, with group membership serving as the between-groups factor and the first three formants extracted from each waveform used as dependent variables. In each analysis, sex served as to covariate, because male and female speakers tend to differ greatly on F_0 (men typically have much lower voices than women). Separate analyses were conducted on formants extracted within each of the two (0-25.6 msec and 25.6-51.2 msec) time windows. Post-hoc tests for group effects utilized the Tukey HSD test to correct for experiment-wise error.

The analysis conducted on formant frequencies comprising the stressed vowel (the second [æ] in Atlanta showed a significant group effect within the first (0-25.6 msec) time window ($\underline{f}(3, 19) = 7.87$, $\underline{p} < .002$). These frequency differences were due to higher formant frequencies in the POSSIBLY TRUE group than in the PROBABLY FALSE group (Fig. 1). This was true for all formants, but reached significance only for formant 2 ($\underline{p} < .002$). In the second (25.6-51.2) analysis epoch, there was no significant effect of true value on frequency.

Formant amplitudes for the stressed vowel in Atlanta were significantly higher in the PROBABLY FALSE group than in the POSSIBLY TRUE GROUP within both the first ($\underline{F}(3, 19) = 8.33$, $\underline{p} < .001$) and second ($\underline{F}(3, 19) = 6.06$, $\underline{p} < .005$) time windows (Figures 2 and 3). Post-hoc tests show that these effects were due to significant formant 3 differences in the first time window ($\underline{p} < .0004$), and significant formant 1 differences in the second time window ($\underline{p} < .001$).

None of the anaylses conducted on formant frequencies or amplitudes comprising the stressed vowels in either Boston (PROBABLY TRUE and POSSIBLY TRUE) or Oakland (PROBABLY FALSE and PROBABLY TRUE) reached statistical significance.

Discussion

The present study examined the formant structure of stressed vowels

within the terminal words of sentences that were considered likely to be true, possibly true, and probably false. The underlying assumption being tested in this study was that the emotional stress experienced during attempted deception would lead to a change in the spectrum of speech sounds, and consequently to the formants extracted from this spectrum.

Two explanations for the formant structure changes seem most plausible. The first explanation is that increased arousal produced by the presentation of the highly task-relevant (lied about) sentences yielded changes in the formant structure. Most subjects reported that they were motivated to beat the lie detector and receive $20. Sentences in the PROBABLY FALSE condition may have been most arousing to subjects because in this condition, subjects were stating information believed to be false with the intent to make it sound true. Thus, in the PROBABLY FALSE condition, the subjects were lying. On the other hand, the POSSIBLY TRUE statements were not as task-relevant, because subjects were less sure of the sentences' truth value.

Another possibility is that these differences were due to the strategy that the subjects used in attempting to pass the lie detector test. Presumably, the subjects knew that they could pass the test by making the sentences that they believed to be false sound true. However, if the formant changes were due to the subjects' attempts to make the false sentences sound true, then we should also expect differences when comparing the PROBABLY TRUE and PROBABLY FALSE conditions. The lack of a significant difference between these two conditions argues against this explanation. The first explanation is more likely: Because the deceivers were motivated to avoid detection, an increase in the autonomic arousal accompanying the deceptive statements was responsible for the observed differences.

If the first explanation is correct, however, one might also predict that differences attributable to belief condition would be observed in other group and condition comparisons. That is, it is logically possible that the differences we find could be due to group assignment instead of belief condition. While this speculation is not particularly compelling, it cannot be dismissed out of hand. The best way to test it directly would be to use a Latin square design, which would result in the same target word being used in all belief conditions.

Changes in the arrangement of formants during vowel production are due to changes in the character of the whole vocal tract working as one resonant system (Fry, 1979; Lieberman & Blumstein, 1988). The autonomic nervous system can influence this resonant system in several ways. For example, it is known that sympathetic projections from the ganglionic chain project to the salivary glands (see Kandel & Kupfermann, 1995). Changes in the activity of these glands resulting from increased autonomic activity could lead to changes in the resonance characteristics of the vocal tract. Respiration and muscle tension are also influenced by sympathetic arousal, and can be expected to have an effect on speech production (Scherer, 1981). Another largely untested possibility is that phasic (short-term) arousal can influence the production of cortical motor programs that code for phonation and articulation.

It should be pointed out that the lack of significant effects for PROBABLY TRUE/POSSIBLY TRUE or PROBABLY TRUE/PROBABLY FALSE contrasts might be due to the arousal pattern hypothesized above, the fact that different vowels were used in the three sets of analyzes, or some combination of such factors. Atlanta [æ] has a low, front unrounded vowel. Oakland [ow] and Boston [ɔ], as spoken in the northeastern U.S., have mid, back, rounded vowels (O'Grady, Dobrovolsky, & Aronoff, 1993). It

may also be relevant that the vowel in Atlanta is preceded by an [l], because the so-called "liquids" (that is, [l] and [r]) produce changes in the spectral shapes of vowels that follow them. One of the goals of this study was to determine if a specific "lie response" exists within the formant structure across different vowels. However, significant differences in the present study were specific to a single formant from a vowel sound within a specific word. These results suggest that arousal-mediated effects on formant structure are subtle and possibly specific to the segment of speech subjected to analysis.

The significant group effect is interesting because it suggests that the spectral characteristics of the same vowel spoken under different belief conditions change during a situation (a lie detector test) that has been shown to be emotionally arousing to experimental subjects (see Ben-Shakhar & Furedy, 1990). Future research is needed to elucidate the mechanisms responsible for these changes. This, in turn, should lead to specific hypotheses about the formant structure expected from a segment of speech obtained during manipulations affecting arousal. This would be useful in forensic psychophysiology for the detection of deception. Once the mechanisms responsible for these formant structure changes are better understood, it may also be possible to use vocal parameters to monitor specific types of stress in clinical populations.

References

Ben-Shakhar, G. & Furedy, J. (1990). *Theories and applications in the detection of deception.* New York: Springer-Verlag.

Ekman, P. & Friesen, W. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology*, 29, 288-298.

Fry, D. (1979). *The physics of speech.* Cambridge: Cambridge University Press.

Hecker, M., Stevens, K., von Bismarck, G., & Williams, C. (1968). Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America*, 44, 993-1001.

Horvath, F. (1978). An experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception. *Journal of Applied Psychology*, 63, 338-344.

Kandel, E., & Kupfermann, I. (1995). Emotional states. In Kandel, E., Schwartz, J., & Jessell, T. (Eds.) *Essentials of neural science and behavior* (pp. 595-612). Norwalk, CT: Appleton & Lange.

Lieberman, P., & Blumstein, S. (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge, U.K.: Cambridge University Press.

Lippold, O. (1970). Physiological tremor. *Scientific American*, 224, 65-73.

Nachshon, I., Elaad, E., & Amsel, T. (1985). Validity of the psychological stress evaluator: A field study. *Journal of Police Science and Administration*, 13, 275-282.

O'Grady, W., Dobrovolsky, M., & Aronoff, M. (1993). *Contemporary linguistics: An introduction.* New York: St. Martin's Press.

Pollina, D., & Squires, N. (in press). Many-valued logic and the N400. *Brain and Language.*

Scherer, K. (1981). Vocal indicators of stress. In J.K. Darby (Ed.). *Speech evaluation in psychiatry* (pp. 171-187). New York: Grune & Stratton.

Simonov, P., & Frolov, M. (1977). Analysis of the human voice as a method of controlling emotional state: Achievements and goals. *Aviation, Space, and Environmental Medicine,* 48, 23-25.

Streeter, L., Krauss, R., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology,* 35, 345-350.

Tolkmitt, F., & Scherer, K. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception and Performance,* 12, 302-313.

Vakoch, D. (1994). Perceived similarity of vocal expression of emotion. *Journal of the Acoustical Society of America,* 95, 2974. [Abstract].

Williams, C., & Stevens, K. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine,* 40, 1369-1372.

* * * * * *

Table 1

Stimuli Used to Complete the Sentences Spoken During the Lie Detector Test

Group	Biographical Information		
	Probably True	**Possibly True**	**Probably False**
1	6' 4"	5' 9"	5' 8"
	1953	1959	1958
	Boston	**Atlanta**	**Oakland**
	Chef	Manager	Salesman
	Chevrolet	Honda	Buick
	Ohio	Florida	Texas
	$280,000	$260,000	$300,000
2	5' 8"	6' 4"	5' 9"
	1958	1953	1959
	Oakland	**Boston**	**Atlanta**
	Salesman	Chef	Manager
	Buick	Chevrolet	Honda
	Texas	Ohio	Florida
	$300,000	$280,000	$260,000

Note: Stressed (underlined) syllables within the emboldened words were digitized and subjected to formant structure analysis.

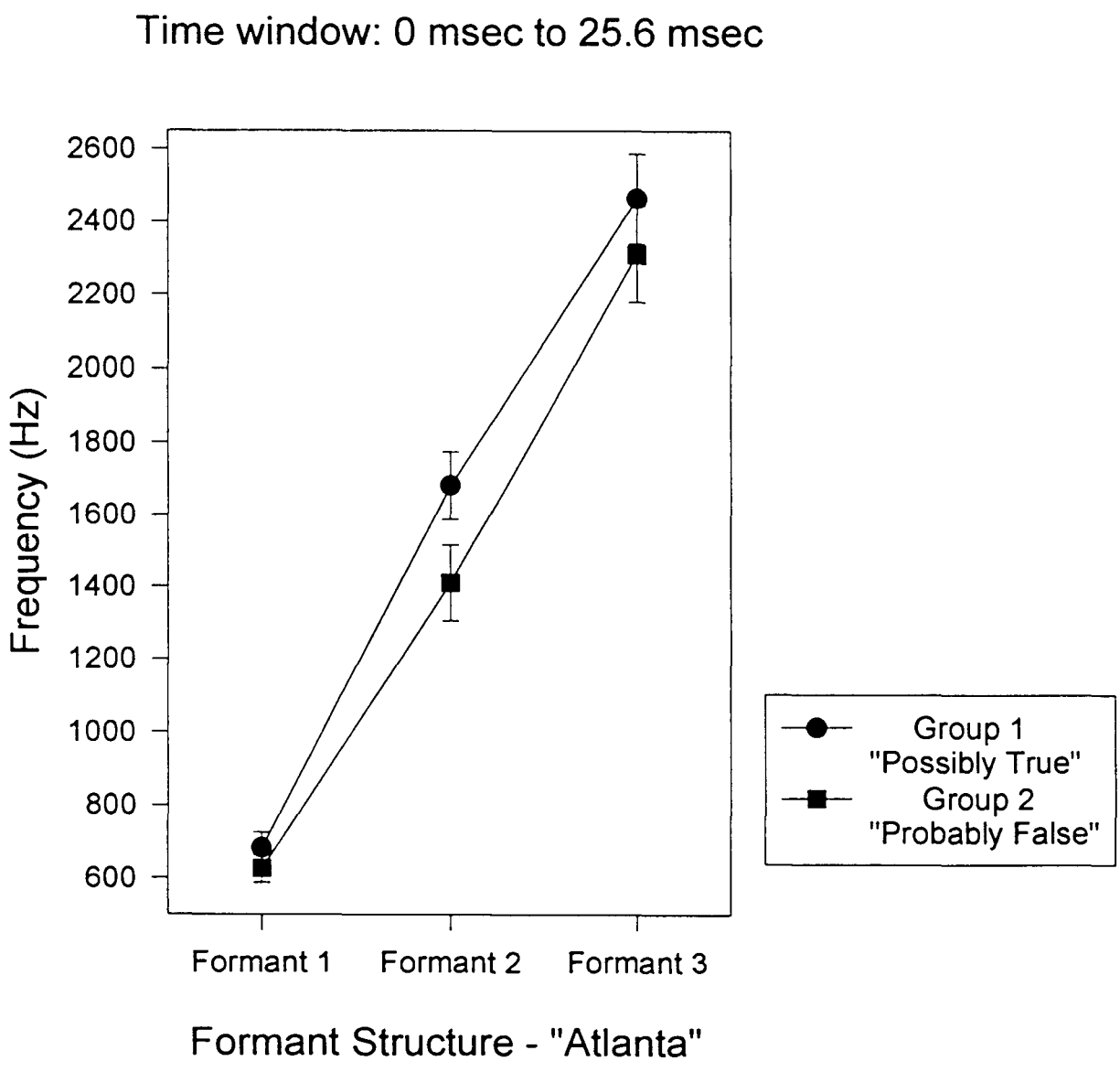

Figure 1

Mean formant frequencies comprising the stress vowel in the terminal word of sentences with different truth values. These frequencies were derived using sampled timepoints starting at the onset of the vowel, and ending at 25.6 msec. Error bars show +/- 1 SEM.

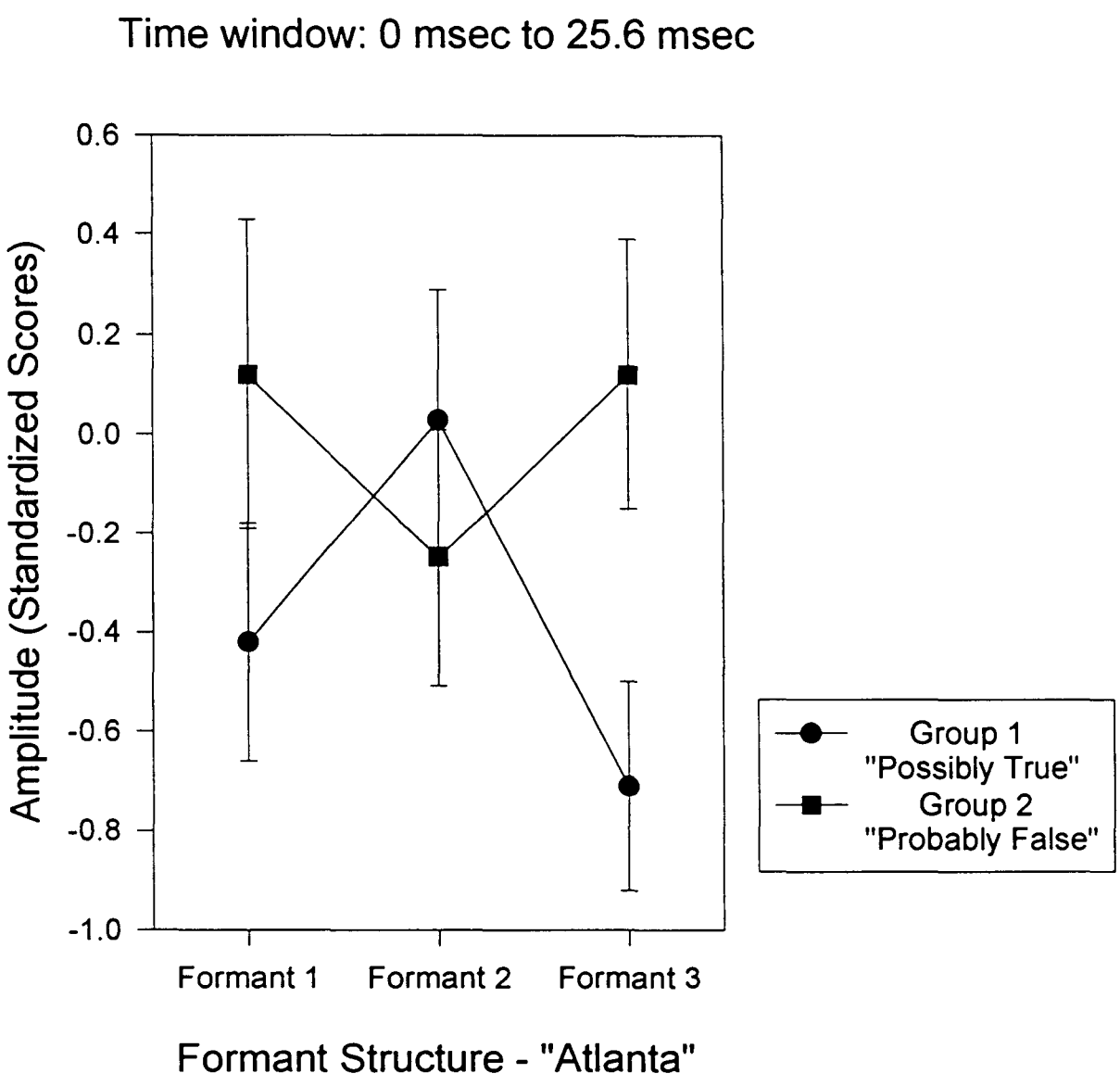

Figure 2

Mean formant amplitudes comprising the stressed vowel in the terminal word of sentences with different truth values. These amplitudes were derived using sampled timepoints starting at the onset of the vowel, and ending at 25.6 msec. Error bars show +/- 1 SEM.

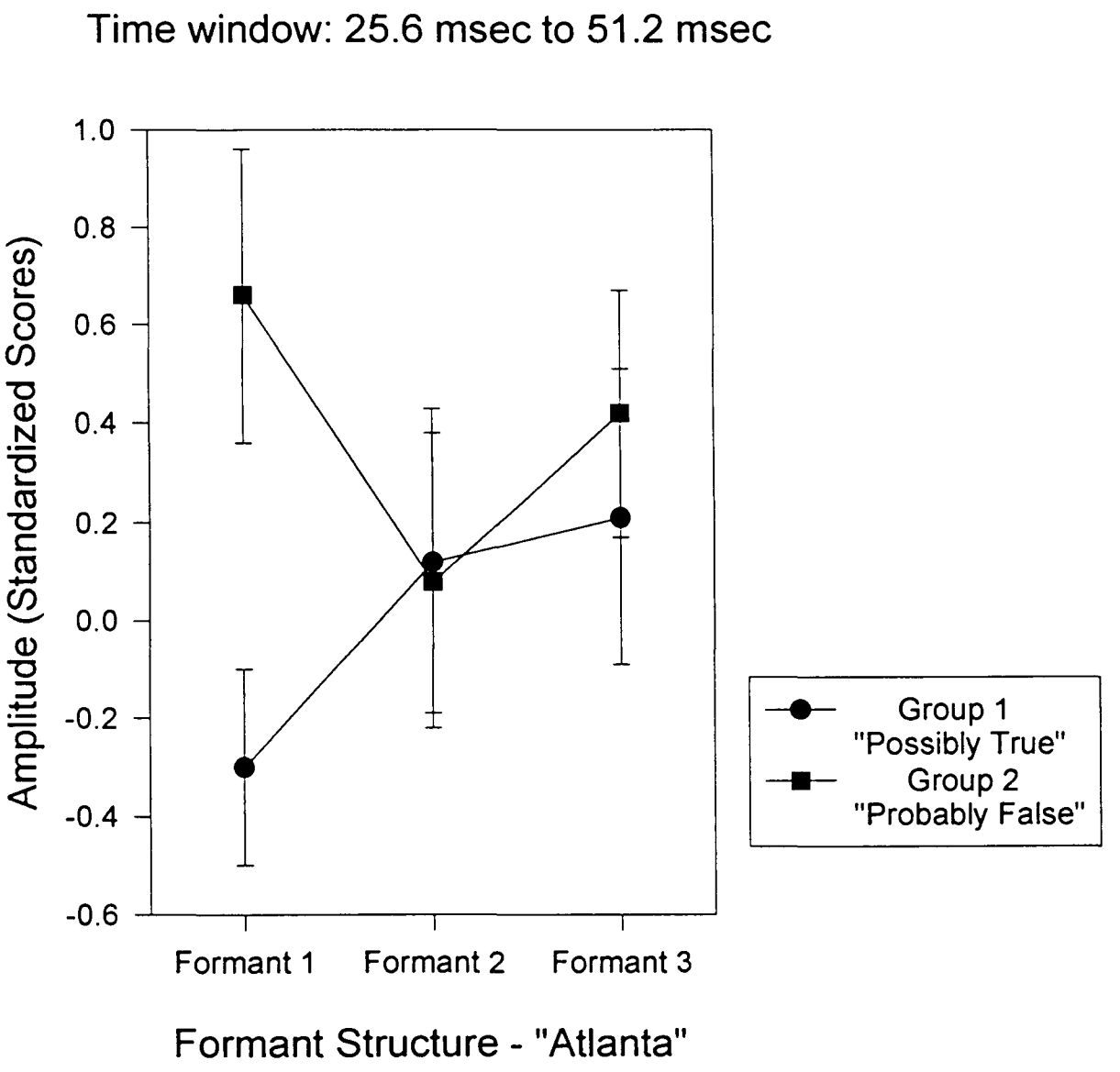

Figure 3

Mean formant amplitudes comprising the stress vowel in the terminal word of sentences with different truth values. These amplitudes were derived using sampled timepoints starting at 25.6 msec after the onset of the vowel and ending at 51.2 msec. Error bars show +/- 1 SEM.

The Zone Comparison Test

Norman Ansley

Abstract

The present work is a companion piece to the author's paper on the Modified General Question Technique, found in *Polygraph*, 27 (1) 35-44. It traces the history and development of zone formats from their obscure inception to current use. Included in this article is a review of the relevant scientific literature, and descriptions of the common zone technique variants.

History and Use

The concept of the zone comparison test, relevant and control questions paired for comparison purposes, first appeared in the scientific literature with the publication of a test format by Father Walter G. Summers, Ph.D., in 1939. He used the format in his own research, which included simulated and real criminal examinations (Summers 1939). He also taught the format to a number of law enforcement agencies, and as of 1952 it was the primary technique of the New York State Police (Kirwan 1952).

In 1961, Cleve Backster published a new test format, called the zone comparison. Like Summers' earlier test it featured pairs of control questions and relevant questions, plus irrelevant questions. All test formats employ one or more irrelevant questions as buffers, and interspersed in the test question sequence as needed to quell responses (Frisby 1979). Backster added some features, pre-test evaluation, a fixed format, and a scoring methodology. His format added two new types of questions, the symptomatic and the sacrifice relevant questions. The symptomatic question is designed to reduce inconclusive results owing to concern for outside issues. Field research has proven that it achieves that purpose (Capps, Knill & Evans 1993). The sacrifice relevant question was designed to serve as a buffer, employed prior to the first relevant question. Research has not substantiated the buffer role for the question, and while it is not scored, it appears to have predictive value (Capps 1991, Horvath 1994). The zone comparison test was quickly adopted by a number of polygraph schools, supplementing the commonly used Reid control question test, the Relevant-Irrelevant test, and the Peak of Tension test. The Reid format, with modifications has become the Modified General Question test, or MGQT, and the peak of tension group of tests includes the Guilty Knowledge Test, or GKT. When the Army polygraph course, which has become the Department of Defense Polygraph Institute (DoDPI), adopted the zone comparison, they modified it and changed the Backster scoring system.

The author is a Life Member of the American Polygraph Association and president of Forensic Research, Inc. Comments on this paper or requests for reprints should be sent to the author at 35 Cedar Road, Severna Park, Maryland 21146.

Since then minor modifications have been made by Backster, the United States Government (at DoDPI), the University of Utah, the Canadian Government Polygraph School, and by Dr. James A. Matte (Backster 1979, 1980; Kopang 1985; Matte 1980; United States Department of Defense Polygraph Institute 1991).

The zone comparison test is a format in widespread use in the United States and abroad. In a national survey by D.G. McCloud of the Virginia Beach Police Department in 1980, 64.9% used a zone comparison test format as their preference in specific issue examinations. In a survey of polygraph schools and courses accredited by the American Polygraph Association, Weaver (1992) reported that all of them taught a zone comparison technique.

Validity

There are three ways to assess the validity of zone comparison tests. One is to follow up on real cases that are subsequently confirmed by confession of the examinee, or someone else; conduct simulated examinations in which you control who is truthful and who is deceptive; and use a panel of judges to determine from evidence in the file (less the polygraph examination) who is guilty and who is innocent, and compare that judgement with the polygraph test results. Reliability is a component of validity and is the consistency of results obtained from a series of tests. No data on repeated tests is available from field research. We know something about the consistency of accuracy of pairs of relevant and control questions within field zone comparison tests, and one research project involved giving simulated zone comparison tests to persons on two consecutive days. Another research method which partially assesses reliability is to have examiners independently evaluate field or simulation charts and compare their results with those where there is independent proof of truth or deception. This paper will provide the results of studies of zone comparison tests from field research, laboratory simulations, independent analysis of field charts, and independent analysis of charts produced in laboratory simulations of zone tests. No panel comparisons have been performed since 1980. Moreover, there is good evidence that panels evaluating case file material against ground truth criterion do not perform well enough to establish any degree of confidence (Dohm & Iacono 1993).

In regard to terminology, many courts use the word "reliability" for the word "validity," without making the distinction noted above.

Field Validity

There have been several field studies relating to the validity of zone comparison polygraph examinations. A cutoff of 1980 has been used, as earlier research was more often based on polygraph instruments that employed pneumatic drives for respiratory and cardiovascular channels, and examiner training was not as good as it is in recent years; improved as basic polygraph courses came into compliance with the American Polygraph Association accreditation standards and advanced training became more common. We are now in the midst of another major change in instrumentation, as computer-based instruments are beginning to replace the standard solid-state amplifier driven instruments. Of the eight field studies reported, all but one (Capps, Knill & Evans 1993) involve the amplifier driven polygraph. The field validity research data appear in Table 1.

Notes on the Studies - Field Validity

Arellano (1984) is an interesting study of real cases because all of the 40 examinees were in the United States illegally, were being tested by their employers for theft,

were tested in the Spanish language, and the test format was Backster Zone. The ground truth was confession by perpetrator or someone else. The research involved both the examiner's scoring and the scoring by an independent reviewer. Both the examiner's scores and the independent scores matched ground truth in all cases.

The research by Capps, Knill and Evans (1993) is based on results of polygraph tests from a computer instrument. They were examining the effectiveness of the symptomatic question in reducing inconclusive decisions. In demonstrating that in real cases the symptomatic significantly reduced inconclusive decisions, they had among the examinations 36 cases confirmed by confession of the subject, and two cases of truthfulness confirmed by the confession of other persons. Because the examiner scored the charts before the confessions, the examiner scores could be compared to ground truth. There was one false positive, and no false negatives.

Edwards (1981) was a survey of police examiners in which he asked them to follow up on the outcome of as many cases as they could and report the results. Because, at that time, all police examiners in Virginia were trained by Cleve Backster in his basic course, and at annual refresher courses, it is probable that most of the examinations were conducted in Backster Zone Comparison, but is possible that a few tests involved other techniques. Because Edwards has the weakest research format, a survey, and accounts for 959 of the 1817 cases in Table 1, (53%), the data is presented in the table with and without the Edwards results.

Elaad and Schahar (1985) is not purely a study of Backster Zone Comparison as some Reid Control Question Test format examinations were included. The authors reviewed all the Israeli Police polygraph examinations for two years, 1973 and 1974 (n. 2,293) and were able to independently confirm the innocence of 100 subjects and the guilt of 74 subjects. They compared the results against the examiner's original scores. There was agreement in 95 of 100 innocent cases (95%) and in 73 of 74 guilty cases (99%).

Mason (1991) reviewed the exculpatory polygraph examinations of 87 male and female soldiers who were identified as illegal drug users through urinalysis testing. Under military regulations an exculpatory polygraph examination may be demanded by any military person who is to be disciplined or discharged for illegal drug use. As the reason for the examination was a positive urinalysis finding, the examiners knew beforehand what was probably the ground truth. A Department of Defense Zone Comparison Test was used in all cases, and the examiners scored the charts with the DoD method, a seven position scale with a +/- 5 inconclusive range. The charts were subsequently scored blind at CID headquarters. There was complete agreement between the examiner and reviewer results. In addition, in 21 cases defense counsel allowed post-test interrogation, and 18 of those confessed. Out of the 87, one was found not deceptive, and the Army CID opened an investigation. They found an invalid collection of the urine sample had taken place. A career was saved.

Matte and Reuss (1989) followed up on all cases conducted with the Quadri-Zone method for the years 1985, 1986, and 1987 by the Buffalo Police Department and all the cases of the Matte Polygraph Service in 1986 and up to April 1987, and compared the outcome of those cases with the examiner's original score. As with all these studies, inconclusive scores, which is no decision at all, were deleted from the data. They used confessions, investigative data, convictions, and combinations of those for ground truth. Of 258 cases available for study, they were able to confirm 115

cases. In all the cases the examiner's score coincided with the case outcome.

Patrick & Iacono (1987) compared the original examiner's decision with case outcome in 81 Canadian cases. The Canadian Police College control question test format is a zone comparison (Kopang 1985; Bradley, Cullen & Carle 1996). The examiner's decision was correct in 27 of 30 truthful subjects (90%) and in all 51 deceptive cases, for an overall accuracy of 78 of 81 (96%).

Putnam (1994) followed up on 552 polygraph examinations conducted by the Washoe County Sheriff's Office (Reno, Nevada) from 1 January 1979 to 1 September 1982. Using only the confession of the subject or another person for ground truth, he confirmed 285 cases, and about four-fifths were in Zone Comparison technique, and one-fifth in the Modified General Question Technique. The original examiner's score was correct in 62 of 65 truthful cases (95%). The score was correct in 219 of 220 deceptive cases (99%), with an overall accuracy of 281 of 285 (99%).

In Widacki (1982), the criminal cases conducted at the University of Katowice (Poland) were reviewed for those that could be confirmed by legal proceedings. Sixteen confirmed guilty and 22 confirmed innocent were compared with the examiner's numerical score. Tests were in the Backster Zone Comparison format. The results of guilty and innocent decisions were not reported separately, only the total of 35 correct of 38 cases (92%).

Notes on the Studies - Independent Analyses of Verified Field Charts

The independent analyses of field charts is an aspect of reliability, and a high degree of reliability is a necessary component of validity. By tradition, these studies are conducted with the examiners blind to all facts of the polygraph test, such as the issue, question formulation, and testing conditions. Research has demonstrated that examiners are more accurate at scoring charts when they have some facts about the testing (Holmes 1958; Wicklander & Hunter 1978). Table 2 lists the four studies in independent analyses of field charts where the reviewers saw complete sets of charts, used a scoring system compatible with the technique, and where the reviewers were examiners experienced with the technique.

Arellano (1984) was also reported in the field studies because both follow-up and independent analyses were performed. The conclusions of the original examiner and the independent reviewer were the same in each case.

Capps & Ansley (1992) were interested in what examiners attend to when they score charts. Eleven experienced examiners, trained in eight different polygraph schools and representing federal, police and private sectors of the profession, scored 40 sets of confirmed zone comparison specific issue criminal cases conducted in the DoDPI format. Using checklists of types of reactions for the three physiological channels, they indicated what kinds of reactions they used in making the decisions. Drawn randomly from a much larger pool of confirmed zone charts, the 40 sets had 17 confirmed truthful and 23 confirmed deceptive. Examiners were blind to all details of the cases and the distribution of truth and deception. They evaluated two sets at a time, mailed to them over a period of nearly a year. The original charts were from cases conducted by 17 examiners. All examinations were confirmed by confession of the subject or another person. In addition to the checklists examiners employed the DoDPI standard seven-point numerical method to each pairing of control and relevant questions (spots). A +/- 5 inconclusive range was used.

Excluding inconclusive scores, the examiners were correct in 361 (97%) of their 372 decisions. Of 143 truthful sets of charts, they were correct in 135 (94%) of their decisions. Of 229 deceptive sets, they were correct in 226 (99%) of their decisions.

Matte & Reuss (1989) had two examiners who did not conduct the original polygraph cases independently evaluate confirmed quadri-zone charts conducted by examiners at the Buffalo Police Department and the Matte Polygraph Service. The independent examiners had no case information and worked separately. Inconclusives excluded, both were correct in 106 nondeceptive decisions and 124 deceptive decisions, totaling 230 correct decisions.

Patrick & Iacono (1987) selected 89 sets of Royal Canadian Mounted Police polygraph examinations that were confession confirmed and had them scored independently by experienced examiners using a numerical scoring system. The test format is not identified, but probably Canadian Zone. The independent examiners, excluding inconclusives, were correct in 11 of 20 truthful cases (55%), and correct in 48 of 49 deceptive cases (98%). Overall, they were correct in 59 of 69 cases (86%). Why those examiners were so poor at evaluating the truthful cases is unknown.

Notes on the Studies - Simulated Zone Comparison Examinations

In simulated examinations, the tests are usually conducted in a laboratory setting, and college students or others are assigned roles. Commonly, a mock crime such as theft, is committed by half of the participants, and half the participants do nothing. The polygraph examiner is blind to the role of the person being tested. The participants in these studies lack the emotional arousal common to field examinations, and the lack of responsiveness is a serious drawback. The effectiveness of attempts to create arousal, or at least interest in the process, by cash rewards or appeals to pride, are problematic (Ansley 1995). Nonetheless, the research is valuable, as it is a means to try new formats, equipment, scoring methods and other variations without placing real cases at risk. Lacking the variations in field examinations, laboratory simulations may be conducted with precision, each case and person treated alike. Table 3.

Barland & Honts (1990) tested 40 soldiers with the DoDPI zone comparison format using a mock crime scenario. Examiners were six instructors at the Institute. They scored the tests with the DoDPI numerical method. In all of the laboratory simulations, inconclusive decisions are not included in the resulting data. In this research, examiners were correct in 14 of 18 (78%) truthful decisions, 14 of 19 (74%) deceptive decisions, and a total of 28 of 37 (76%) decisions.

Bradley & Ainsworth (1984) investigated the effect of alcohol on detection of deception. Although two techniques were applied, we report here only the results of Zone Comparison, inconclusive results deleted. Subjects were 40 university students, 32 guilty of a mock crime, eight innocent. There were 16 who committed the crime sober, and 16 while intoxicated. Eight of each of these groups were tested while sober, eight while drunk. The laboratory instrument recorded respiration, skin resistance, and heart rate. It did not record blood pressure (blood volume) which is recorded by a field instrument. The examiner decisions were correct in 6 of 7 (87%) innocent, 22 of 28 (79%) guilty, and 28 of 35 (80%) for all decisions. Alcohol before the test did not significantly alter the detection rates. Alcohol intoxication during the commission of the crime reduced

detectability with lower scores derived from the measurements of skin responses.

Bradley, Cullen & Carle (1996) tested 120 subjects about a real life embarrassing story or on a laboratory mock crime. Sixty were innocent, and sixty were assigned guilty roles. Police examiners tested forty subjects, half guilty, with the Canadian Zone format, employing standard scoring. The laboratory scientists tested 80 subjects, half guilty, employing a novel scoring method but the Canadian Zone format. Half of each police and laboratory group were tested on whether the embarrassing story applied to them, and half were tested about a mock crime. Police were correct in 16 of 16 (100%) truthful decisions, correct in 11 of 14 (79%) deceptive decisions, and correct in 27 of 30 (90%) decisions. Laboratory scientists using the novel scoring were correct in 19 of 24 (79%) truthful decisions, 19 of 20 (95%) deceptive decisions, and correct in 38 of 44 (86%) decisions.

Dawson (1980) conducted two types of tests, a standard zone comparison, and one in which he had the subjects delay their answer by eight seconds, a novel method. For our purposes, the novel method results were only slightly better, and the insignificant difference and difficulties of administration, have not led to any field application. Dawson's subjects were 24 student actors, aged 19 to 53, at the Strasberg Theatre Institute in Hollywood. Half committed a mock theft of a $20 bill. The guilty were to use techniques taught actors to appear innocent, which they considered a test of their acting ability. The examiner decisions were correct in 9 of 12 (75%) truthful decisions, and 12 of 12 (100%) deceptive decisions, and 21 of 24 (88%) of all decisions. Acting was not an effective countermeasure.

Forman & McCauley (1986) compared the validity of three polygraph test formats with 38 college women who had the choice of accepting two dollars and not opening an envelope, an innocent role, or opening it and taking a promissory note that was either for $2.00 or $10.00, a guilty role, but to keep the promissory note they had to appear truthful on the test in denying they opened the envelope. Sixteen chose the innocent role, 22 chose the guilty role. Each subject received an examination in which there were two charts in the positive control question test format (PCQT), two charts in a zone comparison test format (ZCT), and one chart in a guilty knowledge test format (GKT). With the zone comparison test, the examiner was correct in 7 of 15 (47%) of the truthful, 14 of 17 (82%) of deceptive, and overall, correct 21 of 32 (66%) decisions. The accuracy of the GKT was 69% and the PCQT 72%.

Ginton, Netzer, Elaad & Ben-Shakhar (1982) gave 21 Israeli police cadets the opportunity to change their answers in grading tests returned to them. What they did not know was that the paper had been chemically treated to detect changes, and seven cadets did change their answers to improve their score. After several days all cadets were informed they were suspected of cheating and offered polygraph examinations. To them it appeared their future depended on the polygraph outcome. Three guilty subjects confessed, one did not show up for the test, two took the test, and 13 innocent cadets took the test. The seven staff examiners who conducted the tests used a zone comparison format, and they did not know who was guilty and who was innocent. The examiners applied two types of chart evaluations, one the common 7-position scale with a +/- 5 inconclusive range, and the other, a global method. Using the numerical method, examiners were correct in 11 of 13 (85%) truthful, in 2 of 2 (100%) deceptive, and 13 of 15 (87%) decisions. Global analysis is seldom used with zone tests. In this research, examiners were correct in 7

of 10 (70%) truthful, 1 of 2 (50%) deceptive, and 8 of 12 (67%) decisions.

Yankee & Grimsley (1986) were interested in reliability of polygraph testing, specifically what happens when you test the same people two days in a row. Subjects were 72 college students who were tested on a mock crime with a zone comparison test format, by an experienced examiner, with a polygraph instrument that had one extra channel for a dual electrodermal recording, one for dry stainless steel electrodes and one for silver/silver-chloride electrodes. The groups of 36 men and 36 women were divided evenly into truthful and deceptive roles, and further subdivided into three groups of six subjects each for conditions of accurate feedback, inaccurate feedback, and no feedback. One group would be informed after the first examination that the test detected their role, one group would be told it failed to disclose their role, and one group would get no feedback from the first test. The tests, given in two consecutive days to each person, were identical in format. The DoDPI numerical scoring for zone comparison was used by the examiner, with a +/- 5 cutoff. On the first day, the examiner was correct in 30 of 30 (100%) of truthful, 26 of 30 (87%) of deceptive, and 56 of 60 (93%) of all decisions. On the second examination, the next day, the examiner was correct in 26 of 27 (96%) truthful, 16 of 22 (73%) deceptive, and 42 of 49 (86%) of all decisions. For both days, the examiner was correct in 98 of 109 (90%) decisions.

Notes on the Studies - Independent Analyses of Simulated Zone Test Charts

As with the independent analysis of field charts, the method provides a partial estimate of reliability. The advantage of this method is the certainty of truth or deception, and the uniformity with which the examinations are administered. The disadvantages is the usual lack of arousal, more difficult to interpret. Nonetheless, the studies have some utility. Since 1980, four studies of zone comparison have employed this method. Table 4.

The Forman *et al.* study discussed in the previous section included a blind scoring by a second examiner. The independent examiner had the same accuracy as the original examiner on zone charts (correct in 7 of 15 true decisions and 14 of 17 decisions of deception) and the PCQT charts. The original examiner outperformed the blind scorer in the interpretation of the GKT charts, with an accuracy of 69% against the blind evaluator's 56%.

Similarly, the Ginton *et al.* study found in the previous section also included a blind scoring of the 15 examinations produced from the testing of police cadets on test cheating. Eight examiners blind scored the 15 sets of charts two months after the study. These examiners averaged 82% accuracy scoring the charts of the innocent cadets, and 94% of the charts from deceptive cadets.

Honts & Carlton (1990) wanted to investigate the effect of motivation on soldiers taking a zone comparison polygraph examination to determine who had participated in a mock crime. Half of the 30 guilty and half of the 30 innocent were told they would get the afternoon off if the examiner decided they were truthful in denying the theft. For troops in basic training this was believed to be a significant incentive, having been determined by a survey of troops as the best the Army could offer. The examinations were conducted by experienced polygraph examiners on field instruments. Charts were scored later by independent evaluators, examiners who did not conduct the examinations and were blind to the details. They employed the DoDPI seven-position scoring system with a ±5 inconclusive range. The examiners

were correct in 20 of 23 (87%) truthful, 19 of 23 (83%) deceptive, and 39 of 46 (85%) decisions.

Kircher & Raskin (1988) wanted to compare numerical scoring of zone comparison charts with a computer analysis method. Half the 100 men recruited from the local community committed a mock crime, the other 50 were innocent. The zone question sequence was administered five times, rather than the customary three. Recordings were made on a laboratory instrument, a Beckman type R Dynograph including skin conductance, EKG providing heart rate through a cardiotachometer, finger blood volume from a photocell on the finger, relative blood pressure from a cuff, abdominal respiration, thoracic respiration, and all put on a magnetic tape. Independent numerical scoring was performed on these Utah Zone Comparison test format charts by an experienced examiner. He was correct in 43 of 46 (93%) truthful, 44 of the 47 (94%) deceptive, and 87 of 93 (94%) total decisions.

Conclusions

The field studies of zone comparison technique, published since 1980, involved following up on 1,817 real polygraph cases. In 1,784 of those (98%), the examiner's original score was correct. Research involving independent analyses of zone comparison polygraph charts from real cases involved the study of sets of charts from 711 cases. Examiners correctly scored 690 (97%) of those sets. Simulations of zone comparison tests in a laboratory setting involved 326 examinations, in which the examiners were correct in 274 (84%) of their decisions. Independent analyses of sets of charts from simulated zone comparison tests involved the study of 289 sets. Examiner decisions were correct in 246 (85%) of the sets. The research demonstrates that the zone comparison test format, used by a trained and experienced examiner, with a standard field instrument, is a reliable and valid method of detecting deception and verifying truth.

References Cited:

Ansley, N. (1995). Simulations and laboratory research. *Polygraph*, 24(4) 275-289.

Arellano, L.R. (1990). Research papers: The polygraph examination of Spanish speaking subjects. *Polygraph*, 19(2) 155-156.

Backster, C. (1961, Aug. 20). *Annual Report on Polygraph Technique Trends*. Academy for Scientific Interrogation.

Backster, C. (1990, Aug.) *Backster Zone Comparison Technique: Chart Analysis Rules*. San Diego: Backster School of Lie Detection. Paper presented at the 25th Annual Seminar of the American Polygraph Association, Louisville, Kentucky.

Backster, C. (1979). *Standardized Polygraph Notebook and Technique Guide: Backster Zone Comparison Technique*. San Diego: Backster School of Lie Detection.

Barland, G.H. & Honts, C.R. (1990). *A Laboratory Study of the Validity of the ZOC: An Executive Summary*. Department of Defense Polygraph Institute, Ft. McClellan, Alabama.

Bradley, M.T. & Ainsworth, D. (1984). Alcohol and psychophysiological detection of deception. *Psychophysiology*, 21(1) 63-71. Reprinted in *Polygraph*, 13(2) 177-192.

Bradley, M.T., Cullen, M.E. & Carle, S.B. (1996). Control question tests by police and laboratory questions in a mock crime and real events. *Polygraph*, 25(1) 1-14.

Capps, M. (1991). Probative value of the sacrifice relevant. *Polygraph*, 20(1) 1-6.

Capps, M. & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph*, 21(1) 39-43.

Capps, M. & Ansley, N. (1992). Numerical scoring of polygraph charts: What examiners really do. *Polygraph*, 21(4) 264-320.

Capps, M.H., Knill, B.L. & Evans, R.K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22(4) 285-298.

Dawson, M.E. (1980). Physiological detection of deception, measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology*, 17(1) 8-17.

Dohm, T.E. & Iacono, W.G. (1993, July). *Design and Pilot of a Polygraph Field Validity Study.* Contract Report No. DoDPI93-R-0006. Department of Defense Polygraph Institute, Fort McClellan, Alabama.

Edwards, R.H. (1981). A survey: Reliability of polygraph examinations conducted by Virginia polygraph examiners. *Polygraph*, 10(4) 229-272.

Elaad, E. & Schahar, E. (1985). Polygraph field validity. *Polygraph*, 14(3) 217-223.

Forman, R.F. & McCauley, C. (1986). Validity of the positive control polygraph test using the field practice model. *Journal of Applied Psychology*, 71(4) 691-698. Reprinted in *Polygraph*, 16(2) 145-160.

Frisby, B.R. (1979). Research into the semantics of the irrelevant question. *Polygraph*, 8(3) 205-233.

Ginton, A., Netzer, D., Elaad, E. & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67(2) 131-137.

Holmes, W.D. (1958). Degree of objectivity in chart interpretation. In V.A. Leonard (Ed.) *Academy Lectures on Lie Detection* (vol. 2, pp. 62-70). Springfield, IL: Charles C Thomas. Abstract in *Polygraph*, 19(2) 156-157.

Honts, C.R. & Carlton, B. (1990). The effects of incentives on detection of deception. *Psychophysiology Abstracts*, 27(4A) 539.

Horvath, F. (1994). The value and effectiveness of the sacrifice relevant question: An empirical assessment. *Polygraph*, 23(4) 261-279.

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.

Kirwan, W.E. (16 Oct. 1952). Letter to Norman Ansley.

Kopang, C. (1985, May 5). *The Canadian Police College Polygraph Examination Procedure.* Paper distributed at the Delta College Polygraph Workshop, Delta College, University Center, MI.

Mason, P.M. (1991). *Association Between Positive Urinalysis Drug Tests and Exculpatory Polygraph Examinations.* Fort Belvoir, Virginia: U.S. Army Crime Research Center.

Matte, J.A. (1980). *The Art and Science of the Polygraph Technique.* Springfield, IL: Charles C Thomas.

Matte, J.A. & Reuss, R.M. (1989). A field validation study of the quadri-zone comparison technique. *Polygraph*, 18(4) 187-202.

McCloud, D.G. (1989). *Polygraph Survey Results.* Unpublished manuscript. Virginia Beach Police Department.

Patrick, C.J. & Iacono, W.G. (1987). Validity and reliability of the control question test: A scientific investigation. *Psychophysiology*, 24(5) 605.

Putnam, R.L. (1994). Field accuracy of polygraph in the law enforcement environment. *Polygraph*, 23(3) 260.

Summers, W.G. (1939). Science can get the confessions. *Fordham Law Review*, 8, 334-354.

United States Department of Defense Polygraph Institute (1991). *Zone Comparison Lesson Plan.*

Weaver, R.S. (1992). *Polygraph Technique: Standardization or Versification? A 1992 Status Report.* Paper presented at the Seminar of the American Polygraph Association, Orlando, Florida.

Wicklander, D.E. & Hunter, F.L. (1975). The influence of auxiliary sources of information in polygraph diagnosis. *Journal of Police Science and Administration*, 3(4) 405-409.

Widacki, J. (1982). *The Analysis of Diagnostic Promises in Polygraph Examination.* Katowice, Poland: University of Slaskiego.

Yankee, W.J. & Grimsley, D.L. (1987). The effect of a prior polygraph chart in a subsequent polygraph test. *Psychophysiology*, 24(5) 621-622.

* * * * * *

Zone Comparison Test Question Sequence

Department of Defense Polygraph Institute 1991

1. Irrelevant. Are the lights on in this room? Yes.

2. Sacrifice Relevant. Regarding that stolen money, do you intend to answer truthfully each question about that? Yes.

3. Symptomatic. Are you completely convinced that I will not ask you a question on this test that has not already been reviewed? Yes.

4. Control. Prior to 1990, did you ever steal from someone who trusted you? No.

5. Strong relevant. Did you steal any of that money? No.

6. Control. Prior to coming to Alabama, did you ever steal anything? No.

7. Relevant. Did you steal any of that money from the footlocker? No.

8. Symptomatic. Is there something else you are afraid I will ask you a question about, even though I have told you I would not? No.

9. Control. Prior to this year, did you ever steal anything from an employer? No.

10. Weak Relevant. Do you know where any of that stolen money is now? No.

SKY - Optional

11. Suspect. Do you suspect anyone in particular of stealing any of that money? No.

12. Knowledge. Do you know for sure who stole any of that money? No.

13. You. Did you steal any of that money? No.

Table 1

Field Research

Authors	Nondeceptive			Deceptive			Total		
	Number	Correct	%	Number	Correct	%	Number	Correct	%
Arellano 1990	18	18	100	22	22	100	40	40	100
Capps, Knill & Evans, 1993	2	1	50	36	36	100	38	37	97
Edwards 1981	363	356	98	596	587	98	959	943	98
Elaad & Schahar, 1985	100	95	95	74	73	99	174	168	97
Mason 1991	1	1	100	86	86	100	87	87	100
Matte & Reuss, 1989	53	53	100	62	62	100	115	115	100
Patrick & Iacono, 1987	30	27	90	51	51	100	81	78	96
Putnam, 1994	65	62	95	220	219	99	285	281	99
Widacki* 1982							38	35	92
	634	613	97	1147	1136	99	1817	1784	98
Less Edwards	-363	-356		-596	-587		-959	-943	
	269	257	96	551	549	99	858	841	98

*Widacki did not provide a breakdown.

Table 2

Independent Analysis of Field Charts

Authors	Nondeceptive			Deceptive			Total		
	Number	Correct	%	Number	Correct	%	Number	Correct	%
Arellano 1990	18	18	100	22	22	100	40	40	100
Capps & Ansley, 1992	143	135	94	229	226	99	372	361	97
Matte & Reuss, 1989	106	106	100	124	124	100	230	230	100
Patrick & Iacono, 1987	20	11	55	49	48	98	98	69	86
	287	270	94	424	420	99	711	690	97

Table 3

Simulated Zone Comparison Examinations

Authors	Nondeceptive			Deceptive			Total		
	Number	Correct	%	Number	Correct	%	Number	Correct	%
Barland & Honts, 1990	18	14	78	19	14	74	37	28	76
Bradley & Ainsworth, 1984	7	6	86	28	22	79	35	28	80
Bradley, Cullen & Carle (Police) 1996	16	16	100	14	11	79	30	27	90
(Psychologists)	24	19	79	20	19	95	44	38	86
Dawson 1980	12	9	75	12	12	100	24	21	88
Forman & McCauley, 1986	15	7	47	17	14	82	32	21	66
Ginton *et al.*, 1982	13	11	85	2	2	100	15	13	87
Yankee & Grimsley, 1986 1st Series	30	30	100	30	26	87	60	56	93
2nd Series	27	26	96	22	16	73	49	42	86
	162	138	85	164	136	83	326	274	84

Table 4

Independent Analyses of Simulated Test Charts

Authors	Nondeceptive			Deceptive			Total		
	Number	Correct	%	Number	Correct	%	Number	Correct	%
Forman & McCauley, 1986	15	7	47	17	14	82	32	21	66
Ginton *et al.* 1982 (numerical)	102	84	82	16	15	94	118	99	84
Honts & Carlton 1990	23	20	87	23	19	83	46	39	85
Kircher & Raskin 1988	46	43	93	47	44	94	93	87	94
	186	154	83	103	92	89	289	246	85

Truth or Just Bias: The Treatment of the Psychophysiological Detection of Deception in Introductory Psychology Textbooks

Mary K. Devitt, Charles R. Honts, and
Lynelle Vondergeest

Abstract

This study examined the presentation of psychophysiological detection of deception (PDD; polygraph) testing in introductory psychology textbooks. We examined a sample of 37 introductory psychology textbooks published between 1987 and 1994 for content that discussed PDD testing. Excerpts concerning PDD were then checked for misdescriptions or inaccuracies and rated by two psychophysiologists and a social psychologist. The results showed that PDD received strongly negative treatment in the texts. Moreover, the treatments were often fraught with misdescriptions and inaccuracies. In addition there was an over-reliance on reviews as opposed to empirical studies. We discuss the significance of the problems of bias, reliance on secondary sources, and inaccuracies, and elaborated on the importance of balanced and error free presentations in the medium that serves as a first introduction to the science of psychology for so many people.

Keywords: CQT, literature review, polygraph, psychophysiological detection of deception.

Mary K. Devitt is from Oklahoma State University, Charles R. Honts is from Boise State University, and Lynelle Vondergeest is from the University of North Dakota. This article first appeared in the e-journal *The Journal of Credibility Assessment and Witness Psychology*, 1997, vol. 1, No. 1, 9-32, URL http://truth.idbsu.edu/jcaawp, published by the Department of Psychology of Boise State University. It is reprinted here with the kind permission of Dr. Honts (Editor) and Dr. Devitt (Principal Author). This article was edited by J. Peter Rosenfeld of Northwestern University. The authors wish to thank Eric Landrum and Marc Pratarelli for their assistance in reviewing and rating the textbook excerpts for this study. Thank you also to Peter Rosenfeld and the anonymous reviewers for their helpful comments and suggestions on earlier drafts of this article. Address correspondence and requests for reprints to Mary Devitt, Department of Psychology, 215 N. Murray, Stillwater, OK 74078.

Previous content analyses of Introductory Psychology textbooks have been conducted in areas such as the treatment of counseling *versus* clinical psychology (Leong & Poynter, 1991), transactional analysis (Douglass, 1990), humanistic psychology (Churchill, 1988), sensory deprivation research (Suedfeld & Coren, 1989), religion (Lehr & Spilka, 1989), parapsychology (Roig, Icochea & Cuzzucoli, 1991), the number of neurons in the brain (Soper & Rosenthal, 1988), the Little Albert legend (Paul & Blumenthal, 1989), the Yerkes-Dodson Law (Winton, 1987), the utility of idealized figures (Shepard, 1983), and racial diversity (Gay, 1988). Those studies have illustrated that misdescriptions, inaccuracies, theoretical biases, ambiguity, lack of objectivity, or lack of assimilation may be present in Introductory Psychology material. As a result, it appears that college students are not being well served when controversial material is inadequately and incompletely presented.

In the present study we address another controversial area that is frequently covered in introductory textbooks, that is, the psycho-physiological detection of deception (PDD). Psychophysiological detection of deception tests (also known as polygraph or lie detector tests) are psychological tests that are an important application of psychology in the real world (Honts, 1994a). In the United States and Canada, virtually all federal and local law enforcement agencies employ polygraph examiners who conduct investigative examinations with criminal suspects. The results of such tests often remove individuals from suspicion or result in confessions of wrongdoing following interrogations (Honts & Perry, 1992; Lykken, 1981; Raskin, 1986). Polygraph testing also finds application in the workplace (Honts, 1991). Although many screening uses of polygraph testing in the private sector were prohibited in 1988 (Employee Polygraph Protection Act), employers may still use polygraph to investigate specific losses, and several industries were exempted from the screening ban. Moreover, polygraph tests for pre-employment screening are widely used by federal, state, and local governments. Polygraph pre-employment screening of police officer applicants is particularly pervasive. Finally, polygraph testing plays a critical role in personnel selection and the security clearance process in the national security agencies (Department of Defense, 1991; Honts, 1991; 1994a). All employees of the National Security Agency and the Central Intelligence Agency must take and pass polygraph tests to obtain and retain their security clearances. There are proposals to expand greatly the numbers of individuals subject to such clearance testing (Department of Defense, 1991). Although the numbers of tests conducted in the national security system may be relatively small in absolute terms (*i.e.*, in the tens of thousands), in terms of the special trust and power placed in the hands of those who conduct PDD examinations, the importance of such tests can hardly be overstated (Honts, 1994a). It thus seems important that Introductory Psychology textbooks present a fair and unbiased picture of this important area of applied psychology.

Method

Materials. The data base for this study consisted of an exhaustive sample of the 37 Introductory Psychology textbooks offered to the psychology faculty of a medium sized Midwestern university during the academic year 1993/94. In the case that multiple editions of any textbook were made available, only the most current edition was used in this analysis.

Procedure. Each of the textbooks was searched for references to lie detection, polygraph, or detection of deception testing. If the textbooks

contained references to PDD testing, the words and number of pages devoted to the topic were counted. The textbook sections were then rated on a 7-point scale regarding their orientation toward PDD testing (1 = negative, 4 = neutral, 7 = positive). Three different individuals rated the orientation for all of the textbook excerpts. The first rater was a psychophysiologist (the second author of the present manuscript) who was highly familiar with the polygraph testing literature and who has testified as an expert on polygraph examinations in a number of courts of law in the United States and Canada. The second was an assistant professor of psychology (a colleague of the first author) who was trained as a psychophysiologist and who was not involved in polygraph research or in the polygraph controversy in any way. The third evaluator was an associate professor social psychologist (a colleague of the second author) who has not been involved in the polygraph controversy in any way, but does frequently teach large Introductory Psychology classes. The three evaluations were conducted independently of one another.

In addition, reference citations were recorded. The reference citations present in each textbook were counted, examined, and classified as to their orientation (either positive or negative) toward polygraph testing. The reference citations were also classified as either laboratory or field studies, or reviews. When research or reviews were cited, the descriptions of empirical research and reviews were examined for factual errors or misdescriptions. Also recorded were the types of polygraph usage (forensic testing, investigative testing, on-the-job screening, pre-employment screening, or national security screening) discussed in each textbook. The types of polygraph tests (Control Question Test, Concealed Knowledge Test, or Relevant-Irrelevant Test) mentioned were also recorded.

Results

General Statistics. The data collected in this study are presented and summarized in Table 1. The mean textbook length was 656.31 pages. For only those books that discussed polygraph testing, the mean textbook length was 655.9 pages. The mean number of pages devoted to a discussion of polygraph testing was 1.5 pages. Twenty-nine of the textbooks (78.4%) included some discussion of polygraph testing. Of the texts that discussed PDD, only 11 (29.7%) described empirical research.

Ratings. The mean ratings of the textbook excerpts were as follows: Polygraph-Expert/Psychophysiologist, M = 2.24, sd = 0.87, Independent Psychophysiologist, M = 2.55, sd = 1.01, Social Psychologist, M = 3.79, sd = 0.86. A repeated measures Analysis of Variance (ANOVA) was used to test for differences among the raters. This analysis revealed a significant difference between the means, F (2, 27) = 40.51, $p < .001$. This analysis was followed-up with single degree of freedom tests. The Bernoulli corrected p value was calculated by dividing alpha (.05) by the number post-hoc comparisons (3), for an alpha value of $p = .017$. The univariate tests indicated that the ratings by the two psychophysiologists were not significantly different, F (1, 28) = 1.95, $p > .1$, but that the ratings of the Polygraph-Expert Psychophsiologist and the Social Psychologist were significantly different, F (1, 28) = 71.95, $p < .001$, as were the ratings of the Independent Psychophysiologist and the Social Psychologist, F (1, 28) = 37.57, $p < .001$. Interestingly, neither of the psychophysiologists rated a single excerpt as positive. The average rating for the three evaluators is shown in Table 1.

Citations. Only four (16%) of the textbooks provided any positive citations, and those textbooks cited only review articles. The ratio of

negative citations to positive citations was over 15 to 1 (4.28/.28). To determine if differences in orientation, number, and type of citations existed between the textbooks that discussed both empirical studies and reviews (Mixed group) and those textbooks that discussed reviews only (Review group), *t*-tests for independent samples were conducted. There was a significant difference in orientation for the discussion type, $t(23) = 3.23$, $p = .003$, with the Mixed group providing a more negative discussion ($M = 1.64$) than the Review group ($M = 2.64$). Also noted was a significant difference in total number of citations provided by each group, $t(23) = 3.28$, $p = .003$. The Mixed group provided more citations ($M = 6.46$) than the Review group ($M = 3.0$). Finally, there was a significant difference in the number of negative citations $t(23) = 3.96$, $p = .001$, with the Mixed group presenting more negative citations ($M = 6.46$) than the Review group ($M = 2.57$).

Table 1
Analysis of the Presentation of Polygraph Testing in 37 Introductory Psychology Texts

Text (First Author Shown)
Segments with Both Empirical and Reviews Cited

Atkinson
Doyle
Dworetzky
Feldman
Huffman
Kalat
Lefton
Santrock
Wade
Wood
Worchel
Means

Only reviews Cited

Baron
Bernstein
Bootzin
Carlson
Gleitman
Laird
Meyers
Peterson
Pettijohn
Rubin
Smith
Weiten
Weiten (briefer version)
Wortman
Means

No Citations

Crooks
Darley
Riediger
Shaver
Means

PDD Not Discussed

Benjamin
Bourne
Gerow
Goldstein
Gray
McConnell
Ornstein
Zimbardo

Grand Means

Number of Words on PDD	Number of Negative Cites	Number of Positive Cites	Average Rating (1 = Negative)
1223	3	0	3.67
520	5	0	3.00
1899	12	0	1.33
625	10	0	2.33
587	8	0	2.33
875	7	0	3.00
502	4	0	2.67
722	7	0	3.00
764	3	0	2.33
959	6	0	1.33
1178	6	0	3.00
896	6.5	0	2.54
606	2	0	2.67
575	2	2	3.33
274	2	0	2.33
1135	2	0	3.67
301	2	2	3.33
584	3	0	1.67
1281	9	1	3.00
312	1	0	3.00
232	1	0	3.67
182	2	0	2.67
834	5	2	2.67
520	23	0	3.00
531	2	0	3.00
541	1	0	4.33
565	2.6	0.5	3.02
673			2.33
252			3.00
182			4.00
167			3.33
318			3.17
656	4.3	0.3	2.86

The frequency of various citations was also examined. The most frequently cited (14 times) review was the popular book by Lykken (1981). The most commonly mentioned (6 times) empirical field study was one by Kleinmuntz and Szucko (1984). Finally, the most commonly cited (2 times each) laboratory validity studies were the studies by Honts, Hodes, and Raskin (1985; concerning countermeasures) and by Szucko and Kleinmuntz (1981; concerning validity). Overall, 64 different citations were noted. Fifty (78.1%) of those citations were for reviews, nine (14.1%) were empirical laboratory studies, and five (7.8%) were empirical field studies. Fifteen reviews, two laboratory studies, and three field studies were cited more than one time each. Furthermore, over all of the textbook excerpts there were 113 citations (*i.e.*, some of the 64 separate citations were cited in more than one textbook). The most frequently cited author was David Lykken with a total of 29 citations for eight different publications. At least one of Lykken's works was discussed in 19 of the textbooks.

The types and uses of polygraph testing discussed in the excerpts were also assessed. Those results are presented in Table 2. Overall, 23 of the textbooks discussed some specific use or type of polygraph tests. In those texts that discussed types of polygraph testing (Control Question Test (CQT), Concealed Knowledge Test (CKT), and Relevant/Irrelevant (RI), 17 (74%) mentioned only one test type. The other six textbooks mentioned two types of polygraph tests. No textbook discussed more than two types of polygraph tests. Ten textbooks provided a discussion of the RI test. The CQT and the CKT were each discussed in nine textbooks. The uses of polygraph tests that were assessed included forensic testing, investigative testing, on-the-job screening, pre-employment screening, and national security screening. Overall, 23 (62.1%) of the textbooks included some mention of at least one of the uses of polygraph tests, although only the textbooks with citations (reviews and/or empirical research) discussed those uses. Pre-employment screening was discussed most often (17 times), followed by forensic and investigative testing (14 times each). On-the-job screening was mentioned 13 times, and national security screening was discussed in eight textbooks. Only one textbook discussed all of the polygraph uses. Thirteen textbooks discussed three or four of the uses, while nine textbooks discussed either one or two of the possible uses.

Table 2

Percent of Textbooks That Provided a Discussion of the Uses and Types of Polygraph Tests

Topic (Use/Type)	Research & Reviews	Reviews Only	No Citations
	(n = 11)	(n = 14)	(n = 4)
Forensic	34.6	64.3	00.0
Investigative	45.5	57.1	00.0
On-the-Job Screening	63.6	42.9	00.0
Pre-employment Screen	90.0	42.9	00.0
National Security Screen	45.5	21.4	00.0
Control Question Tests	36.4	35.7	00.0
Concealed Information	45.5	28.6	00.0
Relevant Irrelevant	27.3	35.7	50.0

Finally, the discussions of polygraph testing were examined for factual errors in the reported research. Overall, 25 textbooks provided discussions with research or review citations. Factual errors or misdescriptions were noted in 18 (72%) of those textbooks (*e.g.*, Feldman 1993), in describing a countermeasure study by Honts, Hodes and Raskin (1985), stated that subjects in that study had used a tack in the shoe as a countermeasure. No such manipulation was included in that study.) Details of the errors and misdescriptions in the excerpts are provided in Appendix A at the end of this article.

Discussion

Our analysis of the treatment of PDD in Introductory Psychology textbooks indicates that most textbooks present a negative view of the area. If the majority of research concerning PDD indicated poor validity, this view would clearly be justified. The question thus becomes what does the empirical literature have to say about the validity of PDD tests?

A Brief Review of the Empirical Literature on PDD

Despite their widespread application, polygraph tests have been, and continue to be, the source of great controversy in the scientific literature. Of the three techniques discussed in this paper, there seems to be general agreement in the scientific literature that the Relevant-Irrelevant Test lacks validity (Ben-Shakhar & Furedy, 1990; Honts, 1991; Iacono & Patrick, 1988; Kleinmuntz & Szucko, 1982; Lykken, 1981; Raskin, 1986; Saxe, Dougherty, & Cross, 1985). However, this may be a limited finding as the RI is used very infrequently in forensic settings and its applied uses seem to be limited to employment settings (Honts, 1991). If authors intend that their comments be directed to the use of the RI in employment settings they should state this clearly, as such incontrovertible agreement is noticeably lacking for the other two techniques.

The most commonly used test in the field is the Control Question Test. We will focus most of our analysis on validity studies of the CQT. The third technique, the Concealed Knowledge Test has been studied extensively in the laboratory, but has not achieved much application in the field. In the following section, we also review the empirical literature on the CKT.

The subsequent review also focuses on forensic applications of the polygraph. There is virtually no empirical scientific literature on the validity of PDD tests in employment settings, and thus there is nothing to review (Honts, 1991). Similarly, there is little empirical literature on the national security uses of the polygraph. However, what literature there is on the national security uses consistently produces near change estimates of validity (Barland, Honts & Barger, 1989; Honts, 1991; 1992; 1994a). We found no references to any of these sources in the Introductory Psychology textbooks.

Laboratory Studies Concerning Forensic Settings. A recent meta-analysis of 15 laboratory studies (Kircher, Horowitz, & Raskin, 1988) of the Control Question Test indicated a wide range of validity estimates. One study found near chance results, while six of the studies produce moderate validity estimates, and eight of the studies report validity coefficients of 0.7 or better. In four of the studies, the validity coefficients exceeded 0.8. The Kircher *et al.* meta-analysis noted that these laboratory studies differed widely in their ecological validity. Some studies used mock crimes and procedures that closely modeled field conditions while other studies were very artificial and used unrealistic procedures. Moreover, the Kircher *et al.*, meta-analysis indicated that

those laboratory studies that most closely modeled field conditions produced the highest accuracy rates. A similar state of affairs appears to exist in the Concealed Knowledge Test literature. A more recent review (Honts & Quick, 1995) of the most ecologically valid laboratory studies of both the CQT and the CKT produced overall estimates of accuracy of about 90% and approximately equal false positive and false negative error rates.

Regardless of their methodology, some (*e.g.*, Ben-Shakhar & Furedy, 1990; Lykken, 1981) have criticized all laboratory studies on the grounds that they lack ecological validity. These critics contend that it is not possible in the laboratory to mimic adequately the motivational and emotional context of being given a polygraph test when you are accused of a crime. Others have argued that if sufficient care is taken in creating a deceptive context in the laboratory, then laboratory studies can be useful in estimating the accuracy of the technique in the field (*e.g.*, Podlesny & Raskin, 1978; Kircher *et al.*, 1988).

The Kircher *et al.* (1988) review and meta-analysis should have been easily available to all of the authors of the Introductory Psychology textbooks considered in this analysis. It was published in a first tier psychology journal (*Law and Human Behavior*) that is published by APA Division 41, and is abstracted in all of the popular reference sources. We believe that it is telling that the laboratory study cited most frequently for estimates of validity is the Szucko and Kleinmuntz (1981; *American Psychologist*) study which produced the lowest estimate of accuracy (detection efficiency $\underline{r}$ = 0.21; the next lowest study, which produced an $\underline{r}$ of .51, accounting for six times the criterion variance, is the Kircher *et al.*, meta-analysis). Conspicuously absent from the textbook excerpts were references to equally available publications in first tier journals that produce high estimates for the validity of the Control Question Test (*e.g.*, Podlesny & Raskin, 1978, *Psychophysiology*; Ginton, Netzer, Elaad, & Ben-Shakhar, 1982, *Journal of Applied Psychology*; Kircher & Raskin, 1988, *Journal of Applied Psychology*; Dawson, 1981, *Psychophysiology*; Raskin & Hare, 1978, *Psychophysiology*). As a minimum, each of the studies cited above accounted for 10 times the criterion variance of Szucko and Kleinmuntz (the validity coefficient for Szucko and Kleinmuntz was .21 while the validity coefficients for the cited studies ranged from .65 for Ginton *et al.*, to .87 for Raskin and Hare). One is left with the inescapable conclusions that either the introductory psychology textbook authors gave only a cursory review to the laboratory data on the polygraph or they were biased in their choice of studies to cite.

Ben-Shakhar and Furedy (1990) provide a review of the laboratory studies of the Concealed Knowledge Test. At that time they found ten laboratory studies of the CKT that they felt were scientifically sound enough to include in their review (Balloun & Holmes, 1979; Bradley & Ainsworth, 1984; Bradley & Warfield, 1984; Davidson, 1968; Giesen & Rollison, 1980; Lykken, 1959; Podlesny & Raskin, 1978; Steller, Haenert, & Eiselt, 1987; Stern, Breen, Watanabe, & Perry, 1981; Waid, Orne, Cook, & Orne, 1978). However, no meta-analysis or quantitative analysis of the quality of these studies was reported. Over all ten studies, the accuracy with guilty subjects ranged from 61.1% (Balloun & Holmes, 1979) to 100% (Bradley & Ainsworth, 1984; and Bradley & Warfield, 1984). Accuracy with innocent subjects ranged from 80.6% (Waid *et al.*, 1978) to 100% in seven of the studies (Bradley & Ainsworth, 1984; Bradley & Warfield, 1984; Davidson, 1968; Giesen & Rollinson, 1980; Lykken, 1959; Podlesny & Raskin, 1978; Steller *et al.*, 1987). Only a single one of these studies received a single citation in one

textbook. That study was Bradley and Ainsworth (1984), one of two studies indicating 100% accuracy with both innocent and guilty subjects.

Field Studies Concerning Forensic Settings. In any event, laboratory studies cannot tell the complete story. Data from real world settings are necessary to compliment and extend the results from the laboratory. Unfortunately, validity estimates based on field studies are also mixed and highly debated. Much of the debate regarding field studies concerns the issue of what constitutes adequate methodology. There seems to be an emerging consensus among both proponents (*e.g.*, Honts & Perry, 1992) and critics (*e.g.*, Patrick & Iacono, 1991) that the following are the necessary minimum requirements for field studies of PDD: First, the subjects must represent the population for generalization. If one is interested in studying criminal suspects, then the subjects should be criminal suspects. Second, the cases used in the study should be selected by some random process without reference to the accuracy of the original examiners' decision or to the quality of the physiological data. Third, the decisions used for the data analysis should be based on independent reviews of only the physiological data. Information about the case facts and the overt behavior of the subjects should be withheld from the evaluators. (This criterion holds only if the goal of the study is to determine the ability of the physiological data to discriminate the innocent and guilty. If the goal of the study is to determine the utility of the procedure for some applied goal, admissibility in court for example, the data from the original examiners may be more valuable, see Honts & Quick, 1995.) Fourth, the independent evaluators should be experienced in the independent evaluation of PDD data and they should use techniques that are representative of those actually used in the field. Finally, the truthfulness of the subjects must be confirmed by some criterion that is independent of the outcome of the polygraph examination. Confessions, although problematic, are generally considered to be the best criterion, especially if they are supported by corroborating evidence.

A recent review (Honts & Quick, 1995), found four field studies of the CQT (Honts, 1994b, now in press; Honts & Raskin, 1988; Iacono & Patrick, 1991; and Raskin, Kircher, Honts & Horowitz, 1988) and two of the CKT (Elaad, 1990; Elaad, Ginton, & Jungman, 1992) that were able to meet the stringent requirements for a useful field study described above. Three of the field studies (Honts, 1994; Honts & Raskin, 1988; Raskin *et al.*, 1988) produced accuracy rates above 90%. The independent evaluators in the third study (Iacono & Patrick, 1991) produced a high false positive rate, although the accuracy rate of the original examiners exceeded 90%.

Recently, Patrick and Iacono (1991) have suggested that retrospective field studies may not be useful for estimating the accuracy of polygraph tests because of sampling biases built into the design of such studies. Their position is based on a theoretical analysis and an earlier thought experiment (Iacono, 1991). Fortunately there is no compelling data to support their analysis and many of the assumptions of that analysis are insupportable (*e.g.*, If a guilty person passes a polygraph test, there will be no further investigation of that suspect, and confessions are only obtained following failed polygraph tests). If these assumptions are altered or are invalid then very different conclusions can be suggested (Raskin, Honts & Kircher, in press). Moreover, recent work contradicts their position (Honts, in press) and indicates that confession results are very comparable with results based on other criteria.

Unfortunately, only Iacono and Patrick (1991) would have been readily available to the authors of the Introductory Psychology textbooks considered here and it would have appeared in print as most of these texts would have been nearing completion. It is not fair to expect that the authors of Introductory Psychology textbooks should know about unpublished reports in an applied area. However, there were a number of other field studies that were available to these authors at the time these books were written. All of those studies were reviewed in a study commissioned by the United States Congress and conducted by the Office of Technology Assessment (OTA, 1983). The OTA report was subsequently summarized in the *American Psychologist* (Saxe, Dougherty & Cross, 1985). OTA concluded that there were ten field studies of the Control Question Test that met minimum scientific standards (although none would unambiguously meet all of the criterion described above [Barland & Raskin, 1976; Bersh, 1969; Davidson, 1979; Horvath, 1977; Horvath & Reid, 1971; Hunter & Ash, 1973; Kleinmuntz & Szucko, 1982; Raskin, 1976; Slowik & Buckley, 1975; Wicklander & Hunter, 1975]). Over these ten studies, the average accuracy with guilty subjects was 90% and the average accuracy with innocent subjects was 80%. In those eights studies that used a confession criterion, the accuracy of decisions with guilty subjects ranged from 98.6% (Wicklander & Hunter, 1975) to 75% (Kleinmuntz & Szucko, 1982). With innocent subjects the accuracy rates ranged from 100% (Davidson, 1979) to 51.1% (Horvath, 1977).

At present, there are only two published field studies of the CKT. Both of those studies would meet the criteria described above for a useful field study of the detection of deception. The two studies were reported by Elaad and his colleagues (Elaad, 1990; Elaad, Ginton & Jungman, 1992). The average accuracy rate for guilty subjects in those studies was 47% while the average accuracy with innocent subjects was 98%. These results suggest that in the field the CKT produces extremely high numbers of false negative errors. This finding has been discussed in the light of what we know about eyewitness memory, and may not be surprising (see the discussion in Raskin *et al.*, in press).

Thus, like the laboratory studies, the high quality field studies also seem to paint a relatively positive picture of the accuracy of the CQT, although one could argue that the literature is mixed in both venues. The picture for the CKT is clearer, both the laboratory and the field studies indicate that the CKT is prone to false positive errors and that in field settings the false negative rate may be extreme.

Attitudes of the Scientific Community Toward PDD. Another index of the scientific community's view of PDD testing could be found in surveys. The members of the Society for Psychophysiological Research (SPR) were polled on this topic by The Gallup Organization (1984). At that time, 63% of the respondents said that they believed polygraph tests were useful diagnostic tools when used with other available information, while only 1% of the respondents stated a belief that polygraph tests were without value. More recently, the members of SPR were again surveyed about the attitude toward polygraph testing (Amato & Honts, 1994). The results of the Amato and Honts study showed that 60.2% of the respondents believed PDD tests were useful diagnostic tools when used with other available information. Moreover, 80.5% of the respondents who claimed to be familiar with the PDD literature believed that polygraph tests were useful diagnostic tools. Only 1.7% of the respondents stated that polygraph tests were without merit.

Discussion of the Present Results

Although there is controversy, the empirical and review literature concerning PDD suggests the following conclusions: There is little support for the Relevant-Irrelevant Test, but this test is in frequent use only in employment settings. The laboratory and field data concerning the Control Question Test are mixed. However, when the ecologically valid laboratory studies and the high quality field studies are considered, both indicate high validity for the CQT. The ecologically valid laboratory studies and the high quality field studies of the Concealed Knowledge Test converge on a conclusion that the CKT is prone to false negative errors. Moreover, in the field the CKT seems to produce extreme numbers of false negative errors.

Given the generally favorable findings of both the empirical laboratory and field literature on the CKT, our review of Introductory Psychology textbooks appears to have revealed a distressing lack of balance. None of the textbooks accurately noted the important distinctions in the literature concerning the validity of the three techniques. Moreover, the general negative tone of the textbooks appears to be unjustified by the literature. This lack of balance is typified by the fact that the most commonly cited field study of PDD was the study by Kleinmuntz and Szucko (1984). Of all the field studies available in the literature, regardless of quality, this study is the one of two confession studies (the other is Horvath, 1977) that produced notably lower accuracy estimates. Of the eight confession confirmation studies in the OTA report, these are the two with the worst accuracies.

Given that it is so frequently cited, it may be illustrative to describe the methodology of the Kleinmuntz and Szucko (1984) study at this time. Unfortunately, the most cited form of this study is a 1984 publication in the journal *Nature* which is only about one page in length. Very few details are provided in that publication. However, the study has been described in detail elsewhere (OTA, 1983, and in Kleinmuntz & Szucko, 1982). From those descriptions we can determine the following facts about the Kleinmuntz and Szucko (1984) study. The subjects of this study were individuals who were tested by a private company regarding employee theft as a condition of their employment. None of the subjects was under critical investigation at the time of testing. The psychological data were evaluated by students of a polygraph school who had not completed their training. The polygraph school these students were attending is one that stresses the evaluation of the case facts and the subject's overt behavior. The independent quantitative analysis of the physiological data is not stressed. Finally, the student evaluators were given only 1/9th of the data they would usually have in making an evaluation and they were forced to use an unfamiliar rating scale with which they had no prior experience or training. That rating scale is never used in the field, and the students were not allowed to arrive at an inconclusive outcome, as they would be allowed to do in the field. The cases used were confirmed by confession, but the method of case selection was not specified in the report. There is no indication that any additional confirmatory information was sought or obtained. If the criteria for a useful field study described above are consulted, it can readily be seen that the Kleinmuntz and Szucko (1984) study fails on almost every count. However, none of these methodological shortcomings were mentioned by any of the Introductory Psychology textbook authors who referenced this study.

Another problematic field study that is frequently cited is one by Horvath (1977). One problem with

that study is that the cases were selected for inclusion in the study on the basis of the quality of the recordings, not on some random sampling basis. Moreover, although it is not indicated in the *Journal of Applied Psychology* publication, the dissertation (Horvath, 1974) upon which it is based states that some of the innocent subjects were crime victims who were being tested to verify their statements to the police. Subsequent analyses indicated that all but one of the false positive errors occurred with innocent victims, not suspects (see Raskin, 1986).

We realize that the authors of Introductory Psychology textbooks do not have the time to read each dissertation upon which an empirical report is based, or to read all the available overlapping sources. However, the critical information about the Kleinmuntz and Szucko (1984) and Horvath (1977) studies discussed above was available to the Introductory Psychology textbook authors discussed here through several published reviews (notably, Raskin, 1986; 1987; 1989). The 1987 review by Raskin would have been readily revealed by even a cursory search on PsycLit.

Unfortunately, similar biases are evident in the descriptions of laboratory studies. One of the two most frequently cited laboratory studies (Szucko & Kleinmuntz, 1981) was the only study in the Kircher *et al.* (1988) meta-analysis that produced chance discrimination. As such, it was an extreme outlier in the negative direction. The other frequently cited laboratory study was by Honts *et al.* (1985). Although this study produced moderate discrimination rates in its control conditions, it was cited in the Introductory Psychology textbooks because it demonstrated that under certain circumstances PDD tests could be distorted and/or defeated by countermeasures. Thus, this article was also used to paint PDD testing in a negative light. Numerous laboratory studies published in readily available first tier journals were available to the Introductory Psychology authors, but were ignored or overlooked in favor of an outlier in the negative direction.

Through their choice of citations, the authors of Introductory Psychology textbooks have painted a very negative picture of the science of PDD testing. Our review of the scientific literature shows that this extreme negative view is not justified. Although there is controversy, we strongly believe that the empirical literature supports the validity of polygraph testing with the Control Question Test. Moreover, scientific surveys indicate that the majority of psychophysiologists agree. We believe that most of the current treatments of PDD in Introductory Psychology textbooks are doing an injustice to newcomers to psychology by painting a distorted and biased view of this important applied psychology. At the worst, it could be argued that Introductory Psychology textbook authors should note that there is controversy and describe data from both sides. If studies such as Kleinmuntz and Szucko (1984) are cited, the criticism of such studies should always be mentioned. Such a neutral position would seem to be defensible.

It would appear that Introductory Psychology textbook authors would do well to actually examine the research literature in controversial areas they write about, rather than relying on secondary sources that may have been written by extreme proponents for one side or the other in an ongoing controversy. Truth, rather than bias, should be the criterion for inclusion in this important format that introduces most people to scientific psychology.

References

Amato, S.L., & Honts, C.R. (1994). What do psychophysiologists think about polygraph tests? A survey of the membership of SPR. *Psychophysiology*, 31, S22. (Abstract).

Atkinson, R.L., Atkinson, R.C., Smith, E.E., & Bem, D.J. (1993). *Introduction to Psychology* (11th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Balloun, K.D., & Holmes, D.S. (1979). Effects of repeated examinations on the ability to detect guilt with the polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology*, 64, 316-322.

Barland, G.H., Honts, C.R. & Barger, S.D. (1989). *Studies of the Accuracy of Security Screening Polygraph Examinations*. Fort McClellan, AL: Department of Defense Polygraph Institute.

Barland, G.H., & Raskin, D.C. (1976). *Validity and Reliability of Polygraph Examination of Criminal Suspects* (Contract No. 75-N1-99-0001). Washington, D.C.: National Institute of Justice, Department of Justice.

Baron, R.A. (1992). *Psychology* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Benjamin, L.T., Hopkins, J.R., & Nation, J.R. (1990). *Psychology*. New York: Macmillan.

Ben-Shakhar, G., & Furedy, J.J. (1990). *Theories and Applications in the Detection of Deception*. New York: Springer-Verlag.

Bersh, P.J. (1969). A validation study of polygraph examiner judgments. *Journal of Applied Psychology*, 53, 399-403.

Bernstein, D.A., Clarke-Stewart, Roy, E.J., Srull, T.K., & Wickens, C.D. (1994). *Psychology* (2nd ed.). Boston: Houghton Mifflin.

Bootzin, R.R., Bower, G.H., Crocker, J., & Hall, E. (1991). *Psychology Today: An Introduction* (7th ed.). New York: McGraw-Hill.

Bourne, L.E., Jr., Ekstrand, B.R., & Dunn, W.L.S. (1988). *Psychology: A Concise Introduction*. New York: Holt, Rinehart and Winston.

Bradley, M.T., & Ainsworth, D. (1984). Alcohol and psychophysiological detection of deception. *Psychophysiology*, 21, 63-71.

Bradley, M.T., & Warfield, J.F. (1984). Innocence, information, and the guilty knowledge test in the detection of deception. *Psychophysiology*, 21, 683-689.

Carlson, N.R. (1993). *Psychology: The Science of Behavior* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Churchill, S.D. (1988). Humanistic psychology in introductory textbooks. *Humanistic Psychologist*, 16, 341-357.

Crooks, R., & Stein, J. (1991). *Psychology: Science, Behavior, and Life* (2nd ed.). Orlando, FL: Holt, Rinehart and Winston.

Darley, J.M., Glucksberg, S., & Kinchla, R.A. (1991). *Psychology* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Davidson, P.O. (1968). Validity of the guilty knowledge technique: The effects of motivation. *Journal of Applied Psychology*, 52, 62-65.

Davidson, W.A. (1979). Validity and reliability of the cardio activity monitor. *Polygraph*, 8, 104-111.

Dawson, M.E. (1981). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. *Psychophysiology*, 17, 8-17.

Department of Defense (1991). *Fiscal Year 1990 Report to Congress on the DoD Polygraph Program.* Washington, D.C.: Department of Defense.

Douglass, H.J. (1990). Transactional analysis in American college psychology textbooks. *Transactional Analysis Journal*, 20, 92-110.

Doyle, C.L. (1987). *Explorations in Psychology.* Monterey, CA.: Brooks/Cole.

Dworetzky, J.P. (1991). *Psychology* (4th ed.). St. Paul, MN: West.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, 75, 521-529.

Elaad, E., Ginton, A. & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, 77, 757-767.

Employee Polygraph Protection Act of 1988. Public Law 100-347, 29, U.S.C. 2001 (1988).

Feldman, R.S. (1993). *Understanding Psychology* (3rd ed.). New York: McGraw-Hill.

The Gallup Organization (1984). Survey of members of the Society For Psychophysiological Research concerning their opinions of polygraph test interpretation. *Polygraph*, 13, 153-165.

Gay, J. (1988). The incidence of photographs of racial minorities in introductory psychology texts. *The Journal of Black Psychology*, 15, 77-79.

Gerow, J.R. (1992). *Psychology: An Introduction* (3rd ed.). New York: Harper Collins.

Giesen, M., & Rollison, M.A. (1980). Guilty knowledge *versus* innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality*, 14, 1-11.

Ginton, A., Netzer, D., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67, 131-137.

Gleitman, H. (1991). *Psychology* (3rd ed.). New York: W.W. Norton.

Goldstein, E.B. (1994). *Psychology.* Pacific Grove, CA: Brooks/Cole.

Gray, P. (1991). *Psychology*. New York: Worth.

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

Honts, C.R. (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. *Forensic Reports*, 5, 215-218.

Honts, C.R. (1994a). The psychophysiological detection of deception. *Current Directions in Psychological Science*, 3, 77-82.

Honts, C.R. (1994b). Field validity study of the Canadian Police College polygraph technique. *Psychophysiology*, 31, S57. (Abstract).

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

Honts, C.R., Hodes, R.L., & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, 70, 177-187.

Honts, C.R., & Perry, M.V. (1992). Polygraph admissibility: Challenges and changes. *Law and Human Behavior*, 16, 357-379.

Honts, C.R., & Quick, B.D. (1995). The polygraph in 1996: Progress in science and the law. *North Dakota Law Review*, 71, 997-1020.

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.

Horvath, F.S. (1974). *The Accuracy and Reliability of Police Polygraphic ("Lie Detectors") in Examiner's Judgments of Truth and Deception: The Effect of Selected Variables*. Unpublished Doctoral Dissertation, Michigan State University.

Horvath, F.S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 62, 127-136.

Horvath, F.S., & Reid, J.E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. *The Journal of Criminal Law, Criminology and Police Science*, 62, 276-281.

Huffman, K., Vernoy, M., Williams, B., & Vernoy, J. (1991). *Psychology in Action*. (2nd ed.). New York: Wiley.

Hunter, F.L., & Ash, P. (1973). The accuracy and consistency of polygraph examiners' diagnoses. *Journal of Police Science and Administration*, 1, 370-375.

Iacono, W.G. (1991). Can we determine the accuracy of polygraph tests? In P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.) *Advances in Psychophysiology* (vol. 4). Greenwich, CT: JAI Press.

Iacono, W.G., & Patrick, C.J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.). *Clinical Assessment of Malingering and Deception*. New York: Guilford. (205-233).

Kalat, J.W. (1993). *Introduction to Psychology* (3rd ed.). Belmont, CA: Brooks/Cole.

Kircher, J.C., Horowitz, S.W., & Raskin, D.C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12, 79-90.

Kircher, J.C. & Raskin, D.C. (1988). Human *versus* computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.

Kleinmuntz, B., & Szucko, J.J. (1982). On the fallibility of lie detection. *Law and Society Review*, 17, 84-104.

Kleinmuntz, B., & Szucko, J.J. (1984). A field study of the fallibility of polygraphic lie detection. *Nature*, 308, 449-450.

Laird, & Thompson (1992). *Psychology*. Boston: Houghton Mifflin.

Lefton, L.A. (1994). *Psychology* (5th ed.). Needham Heights, MA: Allyn and Bacon.

Lehr, E., & Spilka, B. (1989). Religion in the introductory psychology textbook: A comparison of three decades. *Journal for the Scientific Study of Religion*, 28, 366-371.

Leong, F.T., & Poynter, M.A. (1991). The representation of counseling *versus* clinical psychology in introductory psychology textbooks. *Teaching of Psychology*, 18, 12-16.

Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385-388.

Lykken, D.T. (1981). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York: McGraw-Hill.

McConnell, J.V., & Philipchalk, R.P. (1992). *Understanding Human Behavior* (7th ed.). Orlando, FL: Holt, Rinehart and Winston.

Meyers, D.G. (1992). *Psychology* (3rd ed.). New York: Worth.

Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation – A Technical Memorandum* (OTA-TM-H-15). Washington, D.C.: U.S. Government Printing Office.

Ornstein, R., & Carstensen, L. (1991). *Psychology: The Study of Human Experience* (3rd ed.). New York: Harcourt Brace Jovanovich.

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

Paul, D.B., & Blumenthal, A.L. (1989). On the trail of Little Albert. *The Psychological Record*, 19, 547-553.

Peterson, C. (1991). *Introduction to Psychology*. New York: Harper Collins.

Pettijohn, T.F. (1992). *Psychology: A Concise Introduction* (3rd ed.). Guilford, CT.: Dushkin Publishing.

Podlesny, J.A., & Raskin, D.C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-358.

Raskin, D.C. (1976). *Reliability of Chart Interpretation and Sources of Errors in Polygraph Examinations* (Report No. 76-3, Contract No. 75-NI-99-0001). Washington, D.C.: National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice.

Raskin, D.C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding the application and acceptance of polygraph evidence. *Utah Law Review*, 1986, 29-74.

Raskin, D.C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioral Science*, 19, 389-404.

Raskin, D.C. (1989). Polygraph techniques for the detection of deception. In D.C. Raskin (Ed.) *Psychological methods in Criminal Investigation and Evidence*. New York: Springer, (247-296).

Raskin, D.C. & Hare, R.D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 16, 126-136.

Raskin, D.C., Honts, C.R. & Kircher, J.C. (In press). The scientific status of research on polygraph techniques: The case for polygraph tests. Chapter to appear in D.L. Faigman, D. Kaye, M.J. Sakes, & J. Sanders (Eds.) *The West Companion to Scientific Evidence.*

Raskin, D.C., Kircher, J.C., Honts, C.R., & Horowitz, S.W. (1988). *A Study of the Validity of Polygraph Examinations in Criminal Investigation.* (Grant No. 85-IJ-CX-0040, National Institute of Justice). Salt Lake City: University of Utah, Department of Psychology.

Roediger, H.L. III, Capaldi, E.D., Paris, S.G., & Polivy, J. (1991). *Psychology* (3rd ed.). New York: Harper Collins.

Roig, M., Icochea, H., & Cuzzucoli, A. (1991). Coverage of parapsychology in introductory psychology textbooks. *Teaching in Psychology*, 18, 157-160.

Rubin, Z., Peplau, L.A., & Salovey, P. (1993). *Psychology.* Boston: Houghton Mifflin.

Santrock, J.W. (1988). *Psychology: The Science of Mind and Behavior* (2nd ed.). Dubuque, IA: Wm. C. Brown.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist*, 40, 355-366.

Shaver, K.G., & Tarpy, R.M. (1993). *Psychology.* New York: Macmillan.

Shepard, R.N. (1983). "Idealized" figures in textbooks *versus* psychology as an empirical science. *American Psychologist*, 38, 855.

Slowik, S.M. & Buckley, J.P. (1975). Related accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings. *Journal of Police Science and Administration*, 3, 305-309.

Smith, R.E. (1993). *Psychology*. St. Paul, MN: West.

Soper, B. & Rosenthal, G. (1988). The number of neurons in the brain: How we report what we do not know. *Teaching of Psychology*, 15, 153-156.

Steller, M., Haenert, P., & Eiselt, W. (1987). Extroversion and the detection of information. *Journal of Research in Personality*, 21, 334-342.

Stern, R.M., Breen, J.P., Watanabe, T., & Perry, B.S. (1981). Effects of feedback of physiological information on response to innocent associations and guilty knowledge. *Journal of Applied Psychology*, 66, 677-681.

Suedfeld, P., & Coren, S. (1989). Perceptual isolation, sensory deprivation, and rest: Moving introductory psychology texts out of the 1950s. *Canadian Psychology*, 30, 17-29.

Szucko, J.J., & Kleinmuntz, B. (1981). Statistical *versus* clinical lie detection. *American Psychologist*, 36, 488-496.

Wade, C., & Tavris, C. (1993). *Psychology* (3rd ed.). New York: Harper Collins.

Waid, W.M., Orne, E.C., Cook, M.R., & Orne, M.T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of deception. *Journal of Applied Psychology*, 63, 728-733.

Weiten, W. (1992). *Psychology: Themes and Variations* (2nd ed.). Belmont, CA: Brooks/Cole.

Weiten, W. (1994). *Psychology: Themes and Variations* (2nd ed., briefer version). Belmont, CA: Brooks/Cole.

Wicklander, D.E., & Hunter, F.L. (1975). The influence of auxiliary sources of information in polygraph diagnoses. *Journal of Police Science and Administration*, 3, 405-409.

Winton, W.M. (1987). Do introductory textbooks present the Yerkes-Dodson Law correctly? *American Psychologist*, 42, 202-203.

Worchel, S., & Shebilske, W. (1992). *Psychology: Principles and Applications* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Wood, E.R.G., Wood, S.E. (1993). *The World of Psychology*. Needham Heights, MA: Allyn and Bacon.

Wortman, C., Loftus, E., & Marshall, M. (1992). *Psychology* (4th ed.). New York: McGraw-Hill.

Zimbardo, P.G., & Weber, A.L. (1994). *Psychology*. New York: Harper Collins.

* * * * * *

Appendix A

Factual Errors and Misdescriptions in the Text Excerpts

Text	Errors and Misdescriptions **Correct Information**
Atkinson *et al.*	States that a relaxed baseline is taken for comparison to later responses. **No polygraph tests do this.** Person may be able to beat the test by causing reactions during the neutral questions. **This would have no impact on the evaluation of a polygraph.** The recording shown in the figure is referred to as a heart rate recording. **It is a relative blood pressure recording.** Persons who are less socialized may be less aroused and harder to detect. **All of the empirical evidence suggests that this is not the case.**
Baron	Control questions are described as name, place of birth, where someone works. **These are neutral not control questions.**
Bernstein *et al.*	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** For the polygraph to be effective, the person being tested must believe that the machine is infallible in its ability to detect lies. **No one who does research in this area states this position. There is no empirical research to support it and a great deal of research to refute it.**

Bootzin *et al.*	The text suggests that you can beat the test with countermeasures to neutral questions. **Countermeasures against neutral questions would have no effect.**
Carlson	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** The text describes a directed lie control test, but calls it a control question test. The text states that the chance of a false positive error on a 3 key 5 item GKT is 8/1000. **The correct value is 1/125, *i.e.*, 1/5 x 1/5 x 1/5, if the items are truly independent.**
Crooks and Stein	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Darley *et al.*	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Doyle	No errors.
Dworetzky	States that there are separate channels for respiratory rate, heart rate, blood pressure and GSR. **Heart rate is not measured unless it is derived from the blood pressure recording.** The text indicates that subjects will be monitored while giving narrative answers to questions like, "Where were you last night?" **In actual tests all questions are answered "Yes" or "No."** The date for Marston supporting the polygraph was given as 1932. **Marston testified in *U.S. v. Frye* in 1923.**

	The text states that most polygraph tests are given by employers and gives an example of a grocery store employee taking a screening test. **Such tests were outlawed by the U.S. Congress in 1988.** The text states that Honts, Hodes, and Raskin (1985) showed that it was "quite easy" to beat the polygraph by creating responses to truthful questions. **Honts, *et al.*, instructed their subjects to increase their response to deceptively answered control questions in the context of a training session where subjects were fully informed about the nature and scoring of the test. With this intensive training only about half of the subjects could beat the test. Without training, none of the subjects were able to beat the test.** States that Floyd Faye failed two polygraph tests. **Faye failed one polygraph, the other was so distorted by Floyd's deliberate movements that it was not able to be scored.**
Feldman	Polygraph measures irregularity in breathing pattern and increases in heart rate. **The polygraph measures respiration, but irregularities are not scored. Heart rate is not scored.** Biofeedback can be used to defeat the polygraph. **There is no evidence in the studies cited to support this assertion. Moreover, there are no credible data to support it in any source.** States that Honts, Hodes, and Raskin (1985) indicates that pressing on a tack in the shoe will allow people to beat the test. **No such manipulation was included in Honts *et al.*, (1985).**
Gleitman	No errors.
Huffman *et al.*	States that polygraph tests can be fooled by people who take tranquilizers, who have consumed high levels of alcohol, or who are psychopathic.

	No cites are provided to support these statements. The empirical literature does not support any of them. The data on psychopaths is particularly clear. They have no special ability to fool the polygraph.
Kalat	No errors.
Laird and Thompson	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** Faye's story about teaching other inmates how to beat the test is presented as fact. **In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.**
Lefton	Habitual liars show little or no autonomic reactivity when they lie. **The cite provided does not address this issue empirically. The literature indicates that psychopaths are just as detectable as normals.**
Meyers	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Peterson	No errors.
Pettijohn	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Roediger *et al.*	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Rubin *et al.*	Heart rate is described as a dependent measure.

	Heart rate is not used as a dependent measure in the field.
Santrock	Polygraph relies on heart rate. **Heart rate is not used.** Drugs and biofeedback can be used to beat the test. **The Waid *et al.* study failed to replicate. ALL other drug studies have failed to find effects. The Corcoran *et al.* study addresses the guilty knowledge test which is not in use in the field. There is no evidence to suggest that biofeedback can be used as a countermeasure against actual field techniques.** Honts, *et al.* is reported as showing that 80 percent of physical countermeasures could be detected by examiners. **Honts *et al.* actually reported that most physical countermeasures could NOT be detected.**
Shaver and Tarpy	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.**
Smith	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** The recording shown in the figure shows one tracing as Pulse Rate Averaging. **PRA is not used in polygraph. The tracing shown is a relative blood pressure tracing.** Faye's story about teaching other inmates how to beat the test is presented as fact. **In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.**
Wade and Tavris	Increased heart rate used as an indicator. **Heart rate not used.**

	People can learn to beat the machine by tensing muscles or thinking about an exciting experience during neutral questions. **This would have no impact on the evaluation of a typical field polygraph.** States that there are problems with reliability. **The literature shows that the reliability of numerical scoring of the Control Question Test is very high, interrater reliabilities are almost always reported to be above 0.90.**
Weiten	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** States that critical questions are compared to nonthreatening questions. **Critical questions are compared to Control Questions that are probable lies.** Kleinmuntz and Szucko (1984) is described as an experiment. **Kleinmuntz and Szucko is an archival field study.**
Weiten (briefer version)	Heart rate is described as a dependent measure. **Heart rate is not used as a dependent measure in the field.** Text indicates that test questions have narrative answers. **In the field all questions must be answered with either a "Yes" or "No".** Kleinmuntz and Szucko (1984) is described as an experiment. **Kleinmuntz and Szucko is an archival field study.**
Wood and Wood	Text suggests that there is no pretest, that subjects are unaware of the wording of questions, and that subjects give narrative answers. **There is a lengthy pretest where the test is explained and all of the questions are reviewed. Subjects must give "Yes" or "No" answers.**

	The nature of the answer to the control questions is unimportant.
	The subject is maneuvered into answering the control questions with a deceptive response. The test is based on differential reactivity between relevant and control questions.
	Heart rate is listed as a dependent measure.
	Heart rate is not used in the evaluation of polygraph tests.
	Habitual liars are more likely to pass.
	There is no empirical evidence that this is true.
	Waid *et al.* (1981) cited as source of mental countermeasure study (counting backward by sevens).
	This study was actually Honts (1986).
	Lykken's (1981) popular book is cited as the source for drug and countermeasure studies.
	Although some countermeasure studies are discussed in Lykken (1981) no original data by Lykken are presented.
	Countermeasures during neutral questions are described as effective.
	Countermeasures during neutral question would have no effect.
Worchel and Shebilske	Heart rate is described as a dependent measure.
	Heart rate is not used as a dependent measure in the field.
	Operators avoid asking did-you-do-it questions.
	***Relevant questions are did-you-do-it questions.* They are asked in virtually all tests.**
	The guilty knowledge test described as if it is the most common in the field.
	The GKT is rarely used in the field.
	Faye's story about teaching other inmates how to beat the test is presented as fact.

	In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.
Wortman *et al.*	Neutral questions are described as control questions. The control question test and the guilty knowledge test are mixed together in the general description of the techniques.

* * * * * *

Errata

In the recent Janniro & Cestaro article (volume 27, number 1) the last sentence of the Data Analysis section should have included the following:

"Scoring reliability (in the form of interrater agreement) was assessed by a multiple rater kappa statistic (Fleiss, 1981)."

The last two sentences of the Results section should have been:

"Interrater reliability for all decisions rendered by the evaluators was high (kappa = .33, SE = .055, $p < .0001$). These evaluators obtained a correct unanimous agreement rate of 26%, and a correct majority (2 or 3) agreement rate of 46%."

We regret the error.

* * * * * *

THE ART OF INVESTIGATIVE INTERVIEWING
A Human Approach to Testimonial Evidence

By

Charles L. Yeschke

Boston: Butterworth-Heinemann, 313 Washington Street, Newton, MA 02158-1626.

BOOK REVIEW

By

Dan Weatherman

According to the author, the intent of the 242-page book is to teach proper interviewing skills to law enforcement officers. The author suggests that most law enforcement agencies are ill-equipped to conduct proper investigative interviews because of lack of time and little formal training in interview techniques. The first four chapters of the book contain an excellent summarization of varied rapport building skills and behaviors required by the investigator for a successful investigative interview. Each chapter is written in such a fashion that it can be used as a teaching outline. The end of each chapter includes *Review Questions*. The questions do a commendable job highlighting all of the important aspects of the chapter. The author acknowledges the late John E. Reid as being a positive influence in his affective interviewing techniques. That influence reveals itself in the authors "semi-structured, non-accusatory questions." As part of his "interview process" the author has developed a *Polyphasic Flowchart*. The chart identifies the different phases of an interview and the degree of "intensity" used in each phase. The flowchart is difficult to follow as one reads the book. However, the last chapter pulls all of the pieces of the flowchart together into a logical investigative process. The strong points of this book are the information regarding rapport building, positive attitude, and active listening.

* * * * * *