The Police Polygraph Digest

January 2013

Table of Contents

Research & Information Chair: Mark D. Handler

Insurance is about Peace of Mind

Complete Equity Markets, Inc. has been insuring polygraph examiners for over 25 years.

We are backed by an insurance company that's been in business for over 300 years and has defended a suit against a polygraph examiner all the way to the Supreme Court.

Limits up to 1M/3M

* Duty to Defend
* Polygraphists Misconduct Endorsement available (includes coverage for Sexual Abuse and Molestation)
* Disciplinary Proceedings Coverage included
* Personal & Advertising Injury included
* Up to 30% in discounts
* Discount for those insured with us for more than 7 years.
* Private Investigation Coverage included (except surveillance) if less than 50% of your time and gross receipts from Private Investigation work under $50,000.
* No deductible
* Recently lowered rates in CA, NY, NJ, HI, FL, TX & AK
* Covers polygraph exams done as a police officer
* Discovery Demand Defense Coverage included
* Optional Data Breach coverage available

Complete Equity Markets, Inc. in CA dba
Complete Equity Markets Insurance Agency, Inc.
CASL 0D44077
Lake Zurich, IL 60047
www.cemins.com (800) 323-6234

1994-HARLEY DAVIDSON/Dyna Wide-glide

or choice of ***$10,000 Cash***

AMERICAN ASSOCIATION OF POLICE POLYGRAPHISTS, INC

William "Buddy" Sentner Scholarship Fund

Classic Motorcycle Raffle

Tickets are $20 each or 6 for $100

Drawing to be held June 5th, 2013
during AAPP Annual Business Meeting
NEED NOT BE PRESENT TO WIN

OR

OVER $30K INVESTED!

- CUSTOM "SHOW" BIKE RARELY DRIVEN 'LIKE NEW' ONLY 4,400 TOTAL MILES- EXCELLENT CONDITION
- CUSTOM "FLAME" PAINT W/ SILVER LEAF LOGOS
- CUSTOM LIGHTING BY ARLEN NESS AND OTHERS
- CUSTOM FRONT FORK SET UP BY ARLEN NESS
- CUSTOM WHEEL SPOKES BY LANDMARK
- CUSTOM MADE "FLAME STITCH" LOW-PROFILE LEATHER SEAT W/BACKREST
- CUSTOM CHROME HANDLEBARS
- CUSTOM 11.5" ROTORS AND CALIPERS BY ARLEN NESS

- HI-PERFORMANCE ENGINE:
  1. S & S 89" HI-PERFORMANCE STROKER KIT
  2. HI-PERFORMANCE H-D "MILLED" HEADS
  3. CRANE HI-PERFORMANCE CAM
  4. CRANE HI-PERFORMANCE IGNITION
  5. S & S SUPER "E" RACING CARBURETOR
  6. EAGLE ONE HI-PERFORMANCE VOLTAGE REGULATOR
  7. UNIQUE CUSTOM OIL COOLER KIT
  8. HI-PERFORMANCE EXHAUST
  9. PINGEL "HI-FLOW" CHROME FUEL COCK

To enter, complete & follow instructions on the back of this flyer or contact:_____

Should a minimum of $10,000 in tickets not be sold, each purchaser will be refunded the respective full purchase amount. Winner responsible for taxes, tags, fees and delivery (from Las Vegas, Nevada) ***The AAPP is a non-profit organization under Section 501(C) (6) of the Internal Revenue Code. Void where prohibited.***

CLASSIC HARLEY DAVIDSON MOTORCYCLE RAFFLE

1994 Customized Harley Davidson Dyna-Wide Glide or $10,000 Cash !

Drawing to be held June 5th, 2013 during AAPP Annual Business Meeting NEED NOT BE PRESENT TO WIN

Sponsored by the American Association of Police Polygraphists, Inc. To support the William "Buddy" Sentner Scholarship Fund
A non-profit organization
P. O. Box 657 Waynesville, OH 45068

Mail orders must be received by May 29, 2013.

$ _____ Check #_____

This is your receipt. You do not need to have a ticket stub to win. Void where prohibited. Contest rules available at www.policepolygraph.org. All information will be used by the AAPP only. We do not sell, rent or give away information we obtain from our supporters. Should a minimum of $10,000 in tickets not be sold, each purchaser will be refunded the respective full purchase amount. Winner responsible for taxes, tags, fees and delivery.

ORDER FORM (This form may be copied)
$20 per Raffle Ticket or Six for $100
Make checks or money order payable to
AAPP
P. O. Box 657, Waynesville, OH 45068

☐ Check ☐ Visa ☐ MasterCard ☐ American Express
Card # _____
Card three-digit security code #:_____ Exp. Date: ____ / ____
Name on Card: _____
Signature:_____

Please Print

Name: _____
Address: _____
City/Town:_____ State: ___ Zip Code: _____
Phone: (_____)-_____-_____ Email: _____

IMPORTANT: Please check your choice of prize:
☐ **CASH** ☐ **MOTORCYCLE**

* WHO IS WILLIAM "BUDDY" SENTNER

Special Agent Buddy Sentner, a 44 year old AAPP member, was shot and killed in the Tallahassee Federal Correctional Institution while serving arrest warrants on six federal corrections officers on June 21st, 2006. The officers had been charged with smuggling contraband to prisoners in exchange for money and other favors. As agents served the warrants in the lobby, one of the six corrections officers opened fire with a weapon he had smuggled into the prison. Agent Sentner returned fire while he shielded the other agents in the room from the gunfire. Before he was fatally shot in the chest, Agent Sentner's shots fatally wounded the assailant. A corrections lieutenant, who assisted the agents serve the warrants and make the arrests, was also shot and wounded by the assailant. The other five corrections officers were taken into custody. Agent Sentner had served in law enforcement for 17 years. He was assigned to the U.S. Department of Justice - Office of the Inspector General, Orlando Field Office. He had formerly served as a special agent with the United States Secret Service and as an officer with the United States Secret Service Uniformed Division.

AAPP 2013 SEMINAR

Omni Charlotte Hotel
Charlotte, NC
June 2, 2013 – June 7, 2013

CLASS/EVENT SCHEDULE

SUNDAY, June 2, 2013

TIME	*EVENT*	*Room*
2:00p – 5:00p	Early Registration	Cypress
5:15p – 5:45p	Worship Service—Barry Cushman AAPP Chaplain	Juniper
6:15p – 8:00p	President's Reception * Bring your registration drink tickets!*	Willow/Birch

MONDAY, June 3, 2013

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
6:15a – 7:30a (Board Members 6:00a)	Registration	Julie & Mark Gerspacher *National Office Manager* *Crew: AAPP BOD*	Cypress
8:00a – 9:00a	Opening Ceremony	Opening Prayer Honor Guard- National Anthem- Pledge of Allegiance President's Welcome Housekeeping Rules Special Announcements	Grand Ballroom

9:00a – 10:00a	*Spouse Breakfast*	*Breakfast for Member Spouses and Guests*	*Oak*
9:00a – 12:00p	**Operation Lie Busters**	**John Schwartz, Fred Ball-US Customs & Border Protection**	**Grand Ballroom**
12:00p – 1:00p	***Lunch- on your own***		
1:00p - 5:00p	**Nailing The Pre-test Interview**	**Skip Webb US Army CID**	**Willow/Birch**
1:00p - 3:00p	**How To Deal With Voice Stress**	**Jim Wygant Polygraph News & Views**	**Poplar/Dogwood**
3:00p - 5:00p	**Polygraph Examinee Suitablity**	**Ray Nelson Lafayette Instrument Co.**	**Poplar/Dogwood**
3:00p – 5:00p	**QC Your Charts** No Appointment Necessary-Just Stop In	**Elmer Criswell**	**Oak**
5:00p – 6:00p	**Meet Your Regional Director**	**Regional Directors, I - V**	**TBA**

TUESDAY, June 4, 2013

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00p	**Managing Conversation in the Screening Environment**	**Dan Weatherman NCCA**	**Willow/Birch**
8:00a – 12:00p	**Polygraph Law: Recent Cases/Polygraph and the Law of Confessions**	**Gordon Vaughan APA General Counsel**	**Poplar/Dogwood**
12:00p – 1:00p	***Lunch- on your own***		
12:00p – 1:30p	***State & National Leadership Luncheon*** *Invitation Only*	**Karen Clark** *AAPP- Vice President State Association Leadership*	**Pomodoro**
1:00p – 5:00p	**Semi-Structured Interviewing**	**Matt Hicks Texas DPS**	**Willow/Birch**
1:00p – 5:00p	**Developing A Police Applicant Screening Program**	**JW Gee JW Gee Consulting (Retired Army CID)**	**Poplar/Dogwood**
1:00p – 3:00p	**ASTM Meeting**	**Chris Fausett**	**Pine**
3:00p – 5:00p	**QC Your Charts** No Appointment Necessary-Just Stop In	**Elmer Criswell**	**Oak**
6:00p	**Tuesday Night Function:** ***NASCAR Hall of Fame***	**General Membership**	

WEDNESDAY, June 5, 2013

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00p	***Annual Business Meeting***	***Everyone Welcome!***	***Grand Ballroom***
12:00p – 1:00p	***Lunch- On your own***		
1:00p - 5:00p	**Applicant R/I Screening with Rank Order Scoring**	Sam Braddock Retired Army CID	**Willow/Birch**
1:00p – 5:00p	**Employment Law Issues Related to Polygraph**	David Nagle AAPP Legal Counsel	**Poplar/Dogwood**
1:00p – 5:00p	**QC Your Charts** No Appointment Necessary-Just Stop In	**Elmer Criswell**	**Magnolia**
1:00p – 3:00p	**Vendor Instrument Training**	**Limestone Technologies**	**Poplar**
3:00p – 5:00p	**Vendor Instrument Training**	**Stoelting**	**Pine**
4:00p	**School Director's Meeting**	**Karen Clark**	**Pomodoro**

THURSDAY, June 6, 2013

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00p	**Polygraph and Internal Affairs Investigations**	**Joel Montezinos Roswell (Ga.) Police Dept. Internal Affairs**	**Willow/Birch**
8:00a – 12:00p	**Social Networking Considerations for the Polygraph Examiner**	**Lee Malkow Snowhomish Co. Sh. Office-Washington**	**Poplar/Dogwood**
12:00p – 1:00p	***LUNCH – on your own***		
1:00p – 5:00p	**The Prevalence of Gangs In America**	Chuck Hastings/Steve Parker – CMPD Gang Unit	**Poplar/Dogwood**
1:00p – 3:00p	**Vendor Instrument Training**	**Lafayette Instruments**	**Juniper**
3:00p – 5:00p	**Vendor Instrument Training**	**Axciton Systems**	**Oak**
1:00p - 5:00p	**Countermeasures**	**Kerry White Las Vegas Metro P.D.**	**Willow/Birch**
6:00p – 7:00p	***Cocktail Hour***	***General Membership***	
7:00p – 9:00p	***Annual Awards Banquet***	***General Membership***	**Grand Ballroom**

FRIDAY, June 7, 2013

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a - 12:00p	**The Psychology of Interviewing**	**Stuart Sentor NCCA**	**Grand Ballroom**
12:00p – 1:00p	***LUNCH – on your own***		
1:00p – 5:00p	**Social Networking, Sexting & Cyberstalking**	**Lee Malkow Snowhomish Co. Sh. Office-Washington**	**Grand Ballroom**

Closing Remarks – AAPP President
Immediately following last speaker

NOTE: Although seldom done, the AAPP reserves the right to change class times, topics and speakers without advanced notice

Please plan on joining us in Las Vegas, NV in 2014

SENSITIVITY AND SPECIFICITY IN POLYGRAPH TESTING - What If? (Ten Reasons to Use a Zone Technique)

Raymond Nelson, Mark Handler, Barry Cushman, and Marty Oelrich*

Last November a 55 year old man followed the advice of his physician and underwent a radical prostatectomy (surgical removal of his entire prostate gland) after a positive PSA blood test result. The removal required an abdominal incision and a 10 day hospitalization because of complications secondary to the surgery. Side effects of the surgery include impotency and urinary incontinence. Prior to the surgery the man was employed as lineman for an electrical company and earned over $100K per year with overtime. The incontinence, however, has made it impossible for him to remain out of doors for lengths of time and he was forced to resign. He now works from home selling used clothing on E Bay to allow him easy access to the bathroom. Prior to the surgery he endured radiation therapy as a precautionary measure. The side effects included loss of appetite, fatigue, skin reactions such as redness and irritation, rectal burning or injury, diarrhea, an inflamed bladder, and blood in his urine.

Following the surgery, pathologic examination found the removed prostate gland to be cancer-free. Unfortunately the physician who diagnosed the cancer relied on a test with high sensitivity but low specificity. The highly sensitive test was able to predict when cancer was present (true positive). However, weak specificity meant the test was poor at ruling out or predicting when the problem was not present, resulting in a substantial percentage of false-positive results. The patient in this case was rendered impotent, lost his job and became depressed as a result of a procedure undertaken following a false positive test result. The patient sued the physician for lost income and asked for punitive damages citing the physician should have known better than to rely on results from that particular test.

During the ensuing civil law-suit, the plaintiff's attorneys brought to light that a better test was available, one with higher sensitivity and specificity. Under oath, the physician was compelled to admit that he had in fact read about the examination that was better able to diagnose when cancer was not present but stated he had not been "trained" in the newer test. The attorney asked the doctor if he recalled whether there was research to support the test he chose, and the good doctor had to concede that there was little empirical support for "his" test. The attorney went on to produce numerous data in support of the technique with higher sensitivity and specificity, all of which the doctor had to admit he was aware of. The doctor stood fast in his defense of the "statistically inferior" test. He told the court that when he was trained 15 years ago, this was the test his medical school taught.

At the conclusion of the hearing the judged summed the case facts up this way: The doctor was originally trained in a technique that he has used for the past 15 years. The doctor is aware of the specificity problem with the technique he is using. The doctor

knows there is at least one technique with substantial scientific data to support increased specificity. The doctor has chosen to ignore the benefits of the increase in specificity derived from this test simply because he has not "been trained" in the newer technique. The doctor admits that he has not sought training and that such training would likely require little more than attending a four-hour continuing education seminar. The actions taken as a result of the false-positive result of this "statistically inferior" test resulted in irreparable harm to the plaintiff. The doctor tried to blame the surgeon for conducting the surgery, but the judge reminded the doctor it was he who made the diagnosis. The judge stated the doctor had a professional responsibility to be able to defend the test he chose and the doctor's defense of his choice was unpersuasive. Judgment for the plaintiff: ten million dollars!

What if the man in this story had been you, a friend or a loved one? Would you feel the judge correctly summarized the physician's professional responsibilities? Should the physician have been allowed to claim a defense of being tied to the technique he had learned a decade-and-a-half earlier? Once he learned a more scientifically supported test existed, did the doctor have a professional responsibility to learn and use that test? Can the ethical dilemma of this hypothetical case inform the polygraph profession? We think so.

Faced with the need to conduct a polygraph examination on a known allegation or known incident, examiners have the choice of two celebrated and established approaches: a single issue Zone Comparison Technique (ZCT) stemming from the important work of Cleve Backster, or a Modified General Question Technique (MGQT) following the example set by John Reid. Though originally developed as discrete techniques themselves, the ZCT and MGQT have evolved into families of related ZCT and MGQT methods, based on the principles defined by Reid and Backster. Differences in these two families of comparison question polygraph techniques include the use of additional test questions, originally intended to address or stabilize some of the dynamic psychological issues that were thought to be at play during polygraph testing, and the use of time-barred comparison questions.

As with many matters of science, some hypothesis work and some prove only to be fancy ideas with apparent face validity but no real empirical usefulness. For example, we now know that outside issue questions do not contribute in their intended ways. Also, we have learned that time-barred comparison questions do not accomplish their desired objective, and that Reid-type comparisons are as good, and possibly superior in some ways, to time-barred comparison questions. The collective result of this knowledge is that our modern application and the empirical basis of the ZCT and MGQT families of polygraph techniques is more similar today than may have been assumed by their progenitors and proponents in the past.

What appears to be well established is the basic principle of numerically evaluating the examinee's perceived salience of, and differential reactivity to, relevant and comparison questions. While Reid may not have originally advocated for numerical scoring of his techniques, no present day scientist would ever attempt to submit or argue a test as scientifically sound or empirically valid without numerical scores, statistical decision support, and normative data. Both the ZCT and MGQT methods employ relevant questions, comparison questions and other questions which are not numerically scored, and both methods are always numerically scored in present usage. As a result of the paring of essential principles to those which have proven themselves over and over, through decades of field experience and numerous empirical studies, contemporary differences in the ZCT and MGQT families of polygraph techniques amount to the selection of examination targets.

The basic principles of formulating question language remains similar for these two families of comparison question techniques, but important differences continue to exist. Field examiners must choose thoughtfully when considering the best approach to a diagnostic or investigative examination referral.

The ZCT approach to target selection is to formulate questions that describe a single behavioral aspect of a known allegation or incident, with the goal of stimulating and observing or measuring the examinee's aggregated responses over multiple iterations of the stimulus. In contrast, the MGQT approach to target selection attempts depends more heavily on verbal and conceptual precision regarding a known allegation or incident, and emphasizes the selection of examination targets that describe distinct behavioral roles of involvement in the allegation or incident. Target selection is the essential difference between a single-issue and multi-facet approach to investigative or diagnostic polygraph testing; the ZCT has historically been a single issue approach, while the MGQT is characteristically a multi-facet approach. We are aware that in practical terms, ZCT formats are also used in multi-issue and multi-facet exploratory circumstances, but emphasize the value of the ZCT family of techniques as primarily stemming from their effectiveness as single-issue techniques. For the purpose of this paper, ZCT exams are regarded as single issue exams.

These two approaches in target selection produce different approaches to the decision theoretic problem, including decision policies employed by field examiners and the calculation of levels of significance and the probability of an erroneous test result, as required by legal standards for acceptance of scientific test data (e.g. Daubert). Another important difference is the volume of research and empirical support describing the accuracy characteristics of these two approaches to the comparison question polygraph test. However, it is the decision policies and calculation of the potential for an erroneous test result, using inferential statistics, that will ultimately establish or refute polygraph as a form of scientific testing among its professional consumers .

While the MGQT family of techniques might initially impress one with its versatility and flexible capabilities as either a diagnostic/investigative technique or a screening technique, the overwhelming volume of polygraph accuracy and validity research has involved single-issue ZCT techniques. There is little information regarding the diagnostic accuracy of multi-facet diagnostic or investigative polygraphs. Moreover, there is a consistent stream of empirical evidence over the last 20 to 30 years, along with statistical decision theory, suggesting that single issue testing approaches and grand-total (scoring) decision policies will produce the highest rates of decision accuracy.

Statistical theory is consistent with these findings and suggests that multi-facet testing approaches, if the results are not calculated using statistical decision theory, can be expected to produce high levels of test sensitivity with higher rates of false-positive errors and inconclusive results for truthful persons. The reasons behind this have to do with the widening of the perceived sensitivity of the test questions themselves, to a broader range of conceptual behaviors. However, sensitivity is also a feature of simple statistical inference. This is because the results of individual questions in multi-facet examinations are based on the total scores for each relevant question. More simply, results are based on the total spot score of *each* relevant question – not the grand total score of all relevant questions (collectively).

Unless one is willing to pretend that the polygraph is a perfect or near perfect test, every result of every distinct question includes the probability of a Type-I (false-positive) error and the probability of a Type-II (false-negative) error. Polygraph decision policies involve “any-or-all” conditions in which “any” indication of a significant reaction will result in a positive test result, while “all” questions must meet the thresholds for a negative result. Increasing the number of distinct scores will inevitably increase the potential for a Type-1 error. Another result of this will be the inevitable mathematical increase in the potential for inconclusive results among truthful persons. This is because the practical result of a multi-question test, scored using traditional scoring rules, will be that the examinee has many opportunities to produce a reaction that meets the threshold for a positive (deceptive) result, and

simultaneously has many opportunities to produce a reaction that fails to meet the threshold for a negative (truthful) result.

Said differently, if positive results are decided at a specified level of statistical significance (e.g. alpha (α) = .05, or a desired false-positive error rate at less than 5 percent) or if decided at individual questions, the cumulative probability of a false-positive (Type-1) error is *inflated* to the specified level of significance multiplied by the number of distinct questions. Therefore, a specified level of significance of α = .05 applied to an MGQT exam with four relevant questions will result in a condition in which up to 20 percent of truthful persons may produce a positive test result. (Four relevant questions times the alpha level of .05 equals .20 or 20 %.) Similarly, a specified level of significance of α = .05 for truthful decisions (meaning there is a 5 percent probability persons will produce similar scores), applied to a four question MGQT exam will result in a condition in which the observed error rate becomes mathematically *deflated* to approximately 1.2 percent, calculated as the inverse of one minus the desired tolerance for error raised to the number of distinct questions. (One divided by ((1 minus .05)^4). This will constrain false-negative errors to desirably low levels, as it becomes highly unlikely that a deceptive person will produce truthful results to all questions. However, this will also result in an increased rate of inconclusive results for truthful examinees.

Single-issue ZCT exams are not subject to the same mathematical complications as multi-facet and multi-issue exams, and error rates are can be calculated much more simply. Moreover, there is emerging evidence that the practice of increased iterations (or presentations) of a single stimulus has the effect of ensuring a sensitivity rate that is equivalent to that of a broader multi-facet investigative exam. In other words, there is no loss of specificity to deception with a single-issue investigative or diagnostic polygraph compared to a multi-facet exam in which each distinct behavioral/role stimulus is presented fewer times. The real advantage of a single issue testing approach becomes evident when looking at specificity rates, with no increase in inconclusives among truthful persons with single issue exams.

These issues become obscured when we placate the desire for simplistic discussions about test accuracy and hold onto naïve hopes for a single numerical index to describe the array of factors that make up the accuracy profile of any test. Test accuracy is properly understood as indicating many things, including sensitivity (the ability to notice or detect the issue of concern when it is present), specificity (the ability to reject cases that do not include the issue of concern), inconclusive rates (which include separate rates of inconclusive results for true-positive and true-negative groups), overall decision accuracy (which is non-resistant again changes in incidence rates), positive-predictive-values (a non-resistant estimation that a positive result is correct), and negative-predictive-values (a non-resistant estimation that a negative result is correct), and unweighted accuracy (which may be more generalizable to circumstances with unknown or variable incidence rates. Traditional discussions about accuracy with inconclusives and accuracy without inconclusives provide a limited and incomplete picture of the accuracy characteristics of the polygraph test.

A recent series of Monte Carlo experiments provided the following accuracy profile for single issue ZCT exams using 10,000 iterations of a Monte Carlo space of 100 single-issue ZCT exams and an incidence rate that varied randomly around a specified mean of .5, while decisions were made using statistically optimized cut-scores and an empirically based simplified manual scoring procedure. (See Table 1.)

Table 1 - ZCT Monte Carlo Accuracy Profile w/ confidence intervals

Dimension	Estimate	95% CI
Correct	0.92	(.86 to .97)
INC overall	0.10	(.05 to .17)
INC Truthful	0.1	(.02 to .19)
Inc Deceptive	0.1	(.03 to .19)
Sensitivity	0.81	(.70 to .91)
Specificity	0.84	(.73 to .93)
FP	0.07	(.02 to .13)
FN	0.08	(.02 to .17)
PPV	0.93	(.84 to .98)
NPV	0.91	(.82 to .98)

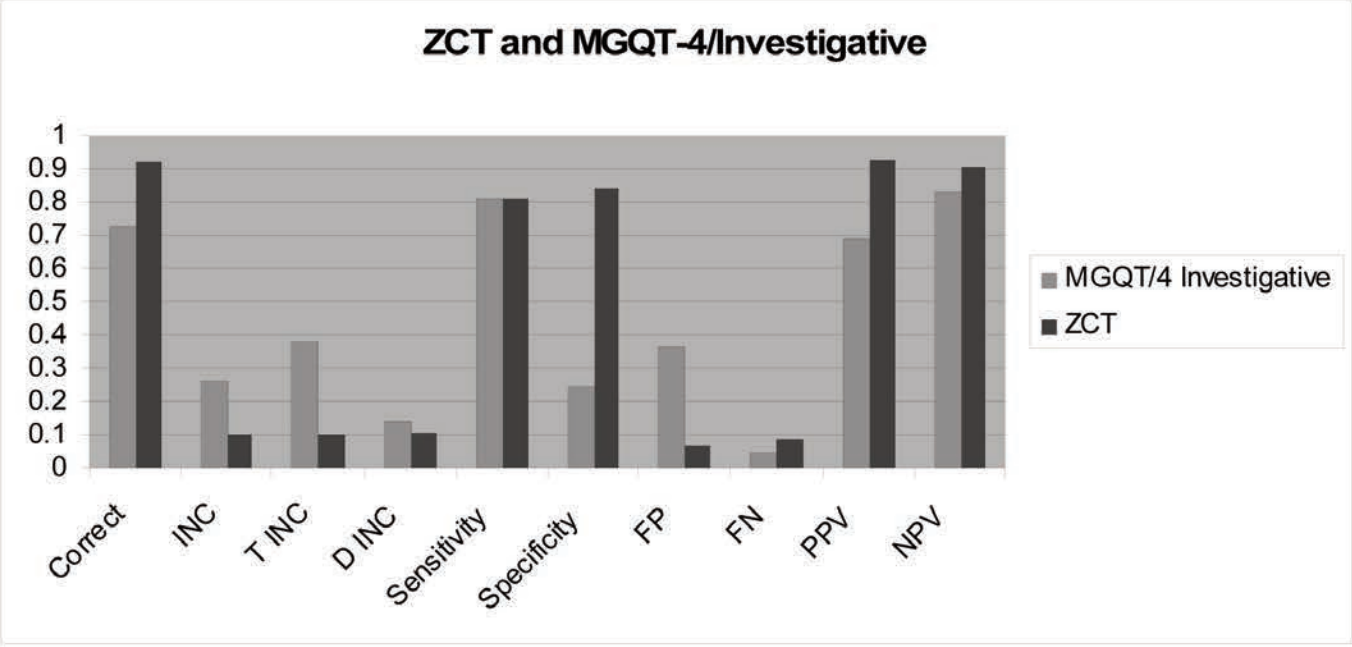
In comparison, MGQT exams produced the following accuracy profile, using 10,000 iterations of a Monte Carlo space of 100 multi-facet MGQT exams and an incidence rate that varied randomly around a specified mean of .5, using traditional cutscores and the same empirically based simplified manual scoring procedure. (See Table 2.)

Table 2 - MGQT Monte Carlo Accuracy Profile w/ confidence intervals

Dimension	Estimate	95% CI
Correct		0.74 (.62 to .83)
INC overall		0.26 (.17 to .35)
INC Truthful		0.38 (.24 to .53)
Inc Deceptive		0.14 (.05 to .25)
Sensitivity		0.82 (.70 to .91)
Specificity		0.26 (.13 to .38)
FP		0.37 (.24 to .51)
FN		0.05 (.02 to .10)
PPV		0.69 (.59 to .81)
NPV		0.84 (.64 to .94)

Figure 1 shows a graphical comparison of the three relevant question ZCT and the four relevant question MGQT results from the Monte Carlo experiment.

Figure 1- ZCT and MGQT investigative accuracy profiles.



These data demonstrate that an empirically based use of the single-issue ZCT approach to diagnostic or investigative polygraphs is capable of providing sensitivity to deception that is equivalent to the multi-facet MGQT. Evaluation of the confidence intervals surrounding the specificity levels of these two

approaches indicates that the single-issue ZCT will also provide levels of specificity to truthfulness that are superior, at statistically significant levels, to a multi-facet MGQT investigative approach using traditional cutscores. Careful evaluation of the confidence intervals surrounding the inconclusive rates for these two testing approaches reveals that the overall accuracy and inconclusive rates will be superior, at statistically significant levels, for single-issue testing approaches. Equally important is the volume of research supporting and describing the accuracy of single-issue ZCT techniques and the conspicuous absence of a similar volume of data for MGQT investigative techniques. To paraphrase Kurt Vonnegut, it is the difference between "oodles and oodles" vs. "diddly-squat." With all of this in mind, we offer the following 10 reasons for recommending a single-issue, three-question zone comparison examination technique for all diagnostic or investigative polygraph circumstances:

1. Less uncertainty about the particular issue to which the examinee is deceptive or truthful,
2. Greater overall accuracy at statistically significant levels,
3. Greater specificity to truthfulness at statistically significant levels,
4. Fewer inconclusive results at statistically significant levels,
5. Easier to explain to non-polygraph professional consumers of polygraph results,
6. Easier for the examinee to pick one type of question to assign greater levels of salience,
7. Easier to defend, with numerous studies on the accuracy of the ZCT and few on the MGQT,
8. More standardized as a result of simpler structure and fewer issues,
9. Fewer demands on the examinee's capacity for precise verbal understanding, and
10. Smaller errors of measurement with more iterations of a single stimulus - easier to control the level of statistical significance, estimate the level of accuracy, and calculate the probability of a false-positive or false-negative error.

As we consider our present knowledge about polygraph techniques, it begins to appear very difficult to endorse any use of a multi-facet MGQT examination approach in a diagnostic or investigative setting. There are simply no advantages to doing so. However, the family of MGQT techniques will likely remain a primary tool in the arsenal of every field examiner as they are inherently suited to multi-issue screening test circumstances.

LX5000 COMPUTERIZED POLYGRAPH SYSTEM

The LX5000 provides **superior physiological data recording** and the **most advanced electrodermal solutions** available. Backed by a team of hardware and software engineers with **decades of experience**, our LX5000 is the **most advanced and flexible** polygraph system on the market!

LX5000 HARDWARE FEATURES

Designed as a robust system that is significantly smaller in size, our basic LX5000 System records nine channels at a time, and provides you with many additional benefits including:

- Data transfer rate up to 360 samples per second across all channels
- 24-bit analog to digital conversion
- Small, compact design making transport and storage easy
- Can add up to 9 additional channels (18 total)
- Extended measurement ranges
- Selectable GSR or GSC channel
- Dedicated PPG channel included
- Durable, yet lightweight design
- Operation with our proven, state-of-the-art LXSoftware
- 3 year warranty and lifetime technical support

LXSOFTWARE v11.1.4 FEATURES

Our Windows®-based software offers unparalleled ease-of-use and proven reliability, and is Windows® 7 compatible. LXSoftware comes with POLYSCORE® and OSS Scoring Algorithms, as well as, the following features:

- Updated User List and Audit Trail
- Ability to "Snap" an Individual Trace to Baseline
- Customizable Personal History
- Detailed Exam/Series Information
- Customizable Reports and Question Templates
- Configurable Audio/Video Options
- Integrated Multi-Language Support for English, Spanish, and Russian languages
- Six EDA choices (GSR or GSC - manual, detrended, and automatic)
- Multi-Camera Support: providing multiple views of the subject
- Customizable Personal History and Exam/Series forms
- Custom Scripting Capability
- Save Polygraph Files and all other documents as PDF formats
- No Software Maintenance Agreement (SMA) or fees for updates!

Call us or visit our website today for more information or to request updated software!

Lafayette Instrument®

3700 Sagamore Pkwy N
Lafayette, IN 47904

Phone: (765) 423-1505
Fax: (765) 423-4111

sales@lafayetteinstrument.com
www.lafayettepolygraph.com

A Primer on Cognitive Dissonance and its Application to Polygraph Testing

Mark Handler and Raymond Nelson

Reprinted with permission of the American Polygraph Association
This article originally appeared in the APA Publication, Polygraph, 41, (3)

Introduction

Cognitive Dissonance (CD) theory is sometimes used to describe underlying causes of arousal resulting in reactions observed during various phases of psychophysiological detection of deception (PDD) testing. This primer explores the origin, history and maturation of CD theory and attempts to make a logical nexus into the realm of interviewing and interrogation.

Basic History

CD was the brainchild of social psychologist, Leon Festinger (1957) whose original proposal was that humans are social creatures and we don't like inconsistency. He suggested the greater the inconsistency, the more agitated we become and the more motivated we are to reduce the inconsistency causing that discomfort. Festinger saw CD as a drive state and believed people consciously and unconsciously sought ways to reduce the state (Greenwald & Ronis, 1978; Harmon-Jones & Mills, 1999).

Cognition comes from Latin meaning "knowledge or thinking" and is commonly used to refer to thought processes. It can be loosely defined as 'any piece of knowledge' a person possesses and it is our brain's psychological representation of anything that can be thought, which can include knowledge about one's behavior or attitude. *Dissonance*, on the other hand, is simply inconsistency or disagreement. Festinger (1957) described the percept of CD as a negative tension state accompanied by a drive to reduce the discomfort. The original CD model ascribed this negative affective state to inconsistent cognitions stating CD occurs when people feel two of their psychological representations differed. Festinger's original hypothesis was that the mere inconsistency created the dissonance and the resulting physiological manifestations. (Cooper, 2007).

Festinger left dissonance research and other scientists continued his work, extending our understanding of the concept. Much of the subsequent CD research focused on the isolating aspects of CD that tease out the causes of dissonance as well as the dissonance reduction strategies. The results of the CD experiments were initially heretical to the way many psychologists thought about social processes (Cooper, 2007). This radical concept was totally inconsistent with the learning and reinforcement (reward and punishment) theories that existed at the time CD was introduced. It showed there was an inverse relationship between the magnitude of incentive and the degree of *attitude change*, one metric by which CD is measured. The more negative the consequences of the behavior (pledging a fraternity, charging a hill full of enemy soldiers, studying hard to complete a training) the more people changed their attitudes towards the value of the achieved goal (they gave it greater value). The harder they fought, the more painful the experience, the greater the worth of the achievement. Other experiments consistently showed the less someone was paid for engaging in attitude-discrepant behavior, the more they changed their attitude towards liking the behavior. It just did not make sense at the time, given the prevailing teachings about punishment and reward.

CD Experimental Paradigms

Scientists developed many clever ways to study the phenomenon of CD. Experiments led to fruitful ideas spurning new hypotheses that were tested, resulting in a multitude of paradigms in which CD was observed. They included:

A. The Free Choice Paradigm - liking what you choose. CD follows decisions we make and is reduced by spreading the attractiveness of the choices (Brehm, 1956). Our choice becomes more attractive and the

one you did not pick becomes less attractive. The more difficult the choice, the greater the dissonance aroused and the greater the spread of the attractiveness. The car you ended up buying is more attractive than the one you chose not to buy.

Perhaps you were taught two test data analysis (TDA) methods and have come to prefer one over the other. You may find yourself coming up with many reasons to use your preferred method and seek faults in the non-preferred method. Another example is during a post-test interrogation, you provide a subject with two alternative reasons for engaging in a behavior, one being less egregious. Should the subject agree to engaging in that less deplorable choice, CD predicts they will feel less dissonance about that behavior. This could help alleviate some stress and allow for more follow up discussion. Post-admission narrative is crucial to ensuring the admission is not false.

B. The Induced Compliance Paradigm – when covertly induced behavior is dissonant with attitude. If you were persuaded into doing something, in which you don't believe, dissonance occurs (Festinger & Carlsmith, 1959).

An often repeated experimental example for this paradigm involves inducing (covertly persuading) a subject to lie to someone about a boring task by telling them it is great fun, and giving them a low or a high reward ($1 or $20) for the lie. Those who are given the $1 will feel more dissonance about their lie and change their attitude more by rating the task less onerous than those paid the $20. The $20 people might think to themselves "yeah that task was boring and it was not nice to lie, so that's why they paid me the big bucks."

This aspect of the CD phenomena has also been shown to work in the political realm. Experimenters asked college students to write counter-attitudinal essays supporting the local police (whom the college kids disliked). The students were offered incentives of 50 cents, $1, $2, and $10 for writing the essays. Those who agreed to write the essays were surveyed afterwards and there was an inverse relationship between magnitude of incentive and attitude change; the lower the incentive, the more positive the attitude towards the police.

Suppose you are required to give a speech on how countermeasure training on the Internet *benefits* PDD testing, and you hate those anti-polygraph websites. As you prepare for, and give the speech, your negative attitude towards countermeasure training will be inconsistent with your behavior of speaking on the benefits of those websites. Dissonance will result and there will be a drive to reduce the dissonance. Now, you can change your behavior (and not give the speech) or you can change your attitude to be more in-line with what you have publicly said. CD theory predicts that you will ultimately come to have a more favorable attitude about the websites. Perhaps you come to value their "raising the bar" on our instrumentation by inducing manufacturers to build better activity sensors. Maybe you feel the anti-polygraph sites have forced us to empirically defend our testing protocols and test data analysis methods.

Another example is what can occur during a posttest interrogation of a suspect who had a significant response to relevant questions pertaining to an alleged sexual assault on a child. The suspect's personal attitude towards their own behavior (if guilty and lying) may be one of disgust, shame or vilification. The resultant attitudinal disposition is not likely to be one consistent with admitting to what they did. However, if they come to believe that others around them see their transgressions as a manifestation of an illness that can be controlled through proper intervention, CD theory predicts they can alter that attitude. Modification of their attitude to one of guilt, compassion or understanding, will increase the likelihood of behavior consistent with repatriation, resulting in an admission.

C. The Effort Justification Paradigm – liking what you suffer for. The prevailing learning theory at the time CD was proffered, predicted that people would be attracted by positive rewards and would reject punishments. As logic would lead us to believe, that suffering would produce a negative attitude about what was causing it.

CD theory predicts the suffering we experience is inconsistent with our desire to

not suffer (Aronson & Mills, 1956). Let's say we enlist in the United States Navy and suffer mental and physical stress at the hands of a recruit company commander, or "drill instructor," throughout our time in boot camp. One way to reduce the dissonance is to change our attitude towards that for which we suffered by increasing its positivity. So enduring suffering in the furtherance of a goal would increase the overall attitude about the value of that goal. We come to value the boot camp experience as something "worth fighting for."

If a posttest interrogation were to ensue for a long time, any admission the suspect makes may be perceived by us, as a worthwhile reward for the effort we expended. The value of that admission can become greatly inflated because of how hard we worked for it. This is why it is of paramount importance to not let an over-valuation of an admission result in stopping short of obtaining post-admission narrative. In defense of a claim of a false confession, "I did it" should not be a stopping point, but rather a *starting point*. CD can be a double-edged sword, capable of cutting both ways, and understanding how this could affect us can increase our personal and collective professionalism.

D. The High versus Low Threat Paradigm – forbidden items are devalued more in a low threat than in a high threat situation. Let's say you have some kids playing in a room and there is an obvious "best" toy. You tell one group of kids if they touch the toy while you are gone you will be really angry and leave with all of your toys (high threat). You tell the other group of kids you will be disappointed with them if they touch that toy and ask them not to do so while you are out of the room (low threat). Later, when asked how "nice" the special toy is, those in the low threat group will rate the toy as less desirable than those in the high threat group. Both groups of kids wanted to play with the toy and none did. The high threat group needed no justification for why they didn't play with the toy; they were told they would face serious consequences for breaking the rules. The low threat group had to reduce the dissonance of not playing with the toy by changing their attitude about the value of the toy. They adjusted their attitude such that they did not like the toy as much. This would make sense in a posttest interrogation setting as well.

If the suspect believes the consequences of admission are commensurate with the high threat of social or moral ostracization, they are unlikely to change their attitude towards making an admission. However, if they can change their attitude about how they will be perceived, the resultant behavior of denying will then become inconsistent with their new attitude, CD will result, and a change in behavior (an admission) made to reduce the dissonance. In other words, if they believe the threats of moral reprisals are high, CD theory predicts they will be less likely to make an admission. Lowering the perceived threat level should induce the subject to being more amenable to making an admission.

E. The High versus Low Choices Paradigm - freedom to choose is a necessary condition for dissonance arousal. When free (high) choice is removed, dissonance is not aroused and the behavior is justified because there was no choice. For example, in a Directed Lie Comparison Question (DLCQ) test, I am required to lie to the DLCQ. That requirement serves as an important cognition, consonant with my behavioral act. I lied because the examiner directed me to lie (it is a low choice condition). If, on the other hand, I have free (high) choice to engage in a lie or not (e.g. a Probable Lie Comparison Question), CD can result. My behavior, lying to the Probable Lie Comparison Question (PLCQ), is not consistent with my attitude that I should not lie to any question on this test in order to pass the examination.

High-low choice conditions have been studied extensively and have conclusively shown that free (high) choice is a necessity to create dissonance and the lower the incentive (i.e. $1 versus $20), the greater the attitude change. When low choice conditions (DLCQ) are set up, experimental results followed the prediction of learning and reinforcement and attitude change was positively correlated with the magnitude of incentive. The higher the incentive, the more attitude changed. This is inconsistent with CD theory and so the arousal associated with DLCQ testing is unlikely a result of CD, and probably more

consistent with goal achievement. When high choice conditions are established (PLCQ), arousal follows the predictions of CD theory.

F. The New Look Model of Dissonance – developed by Joel Cooper (Cooper, 2007) and colleagues at Princeton University in the early 1970's. One of the founding principles was that CD arousal reduction was in the service of mitigating negative valance, aversive, or unpleasant feelings. People tend to identify the affective reaction to CD as discomfort. If you ask people to describe what they feel, they usually describe it along the lines of "tension" or "uncomfortable." People experiencing CD are motivated to reduce the unpleasant state and do so by changing their attitudes about the perceived cause (usually their behavior). Several studies have shown that dissonance reduction via attitude change reduces the level of discomfort the subjects report. Alternatively, if experimenters remove the discomfort, dissonance goes away.

The New Look experiments also focused on *modifiers*, they called 'but-onlys.' They found;

1. CD was aroused, *but only* when decision freedom was high (high choice). Little choice results in little dissonance.

2. CD was aroused, *but only* when people are committed to their statements, beliefs or initial attitudes about their behavior. Also, if there is a chance to 'take back' what you did or said at a later time, CD is mitigated.

3. CD depends on perceived consequences, *but only* if those consequences are aversive. Aversive consequences play a pivotal role in the New Look model. For example, in one study students told a another student (actually a confederate) that a dull task was exciting. If the to-be-duped confederate acted excited as if they believed the student, CD resulted. However, if the confederate acted unconvinced, no CD resulted.

4. The unwanted consequences of a high choice behavior will create CD, *but only* if those consequences are foreseeable (not necessarily foreseen) for CD to occur.

Dissonance Arousal Turns to Motivation

As described earlier, CD is experienced as an unpleasant arousal state, which people become motivated to reduce. The reduced motivation can be mitigated by redirecting the subjects' attribution of the cause of the arousal. Subjects can be fooled into "blaming" the physiological arousal on something else and CD will not result. During several experiments, subjects were tricked into attributing CD arousal to a drug (actually a placebo) and this false attribution reduced or eliminated dissonance. In one experiment, 1/3 of the students given a "memory enhancing" drug (actually a placebo), were told they would feel excited, tense or aroused (excited group), 1/3 were told the drug had no side effects (none group) and the remaining 1/3 were told they would feel calm and relaxed (calm group). Each third was further sub-divided into high-low choice groups and all engaged in an induced compliance experiment. CD theory predicts those in the high choice group would experience dissonance and those in the low choice group would not. A further hypothesis of this experiment was that the excited group-high choice participants would attribute the tenseness to the pill and not experience CD as measured by attitude change.

CD theory predictions and the hypothesis held true. There was a significant difference in the change in attitude in the calm and none groups between the high and low choice participants. There was no significant difference between the high and low choice participants in the excited group.

Measuring Physiologic Arousal

Using an induced compliance paradigm experiment, autonomic arousal has been measured via non-specific electrodermal responses (NS-EDR). High choice participants exhibited heightened physiological arousal, as measured by NS-EDR compared to the low choice participants.

Summary

Behavior inconsistent with attitudes produces CD, but only;

a. if behavior is freely chosen,
b. if the subject is committed to the behavior,
c. if agency is attributable to the self ("I caused this thing"),
d. if the behavior leads to aversive or unwanted consequences, and
e. if those unwanted consequences were foreseeable (not necessarily foreseen) at the time of the behavior.

To reduce the dissonance one can change their attitude about the behavior or change the behavior itself. There are a number of attitude changing strategies that could be employed in a posttest interrogation to reduce the arousal associated with CD resulting from withholding information. CD theory predicts a desire to reduce arousal can foster an admission that can be developed appropriately into a true confession. One way to change attitude is to spread the choice. This can be done by emphasizing the moral and psychological benefits of an admission. The non-selected choice becomes less valuable and the selected choice increases in perceived value. Additionally, if the subject chooses the explanation for their behavior, they will tend to value that choice more than any alternative. This can result in a further reduction of dissonance which could encourage further post-admission narrative. Perhaps a "Wow, that feels better, I will keep going" kind of self-talk.

If the subject's behavior of concern is not perceived as freely chosen, or not attributed solely to them, or was not foreseeable, then CD may be mitigated and they may be less inclined to admit to the behavior. If the moral and psychological consequences of not admitting to the behavior can be perceived as less aversive, CD arousal will be reduced, and again, an admission is less likely to occur.

In summary then, there are many ways to reduce dissonance once it has been aroused. The greater the discomfort, the greater the reduction motivation the subject feels. The easiest way to decrease the CD would be to reduce or eliminate the discrepancy. If my knowledge of my behavior and my attitude towards that behavior are consonant (they agree), there will be no dissonance. If I become aware they are dissonant, I will feel a tension state and be motivated to move them more in line with one another.

CD theory, then, is really a theory about attitude change. It is often easier, and more likely, to change my *attitude* about my behavior than to change the behavior itself. It is difficult to alter the reality of my behavior (especially once the behavior has already occurred) but not so difficult to change my attitude towards that behavior. If I can be caused to change my attitude about the information I am withholding, it will serve to reduce the CD. If CD is reduced, the resulting tension state keeping me from divulging can be reduced and I am more likely to admit to what I did. An additional and important aspect of CD theory is that not all cognitions are made equal. If one cognition is assessed as more important than another, it will create more CD than one of perceived lesser value. Examiners can capitalize on this by motivating a suspect to change their attitude about *the importance* of a cognition to reduce the dissonance.

Applying CD theory to PDD testing and posttest interrogations can provide insight into causation of arousal and arousal reduction strategies. CD theory can inform our profession on ways to apply social science, but with one important caveat. "The wise prevail through great power and those who have knowledge muster their strength" (Proverbs 24:5, New International Version). Knowledge is said to provide strength and those empowered by this learning have an ethical obligation to use it for benevolent purposes. We should strive to use newly gained knowledge for the betterment of all members of society, without regard for our personal stakes. We should not only embrace those aspects of social science that agree with our personal beliefs and objectives, we should expand our professional horizons by opening our minds to alternative theories. Changing our own attitudes about new, novel or different theories can reduce our dissonance. There are a number of areas in which social science can improve our collective knowledge (e.g. false confession research) that we should explore as well. The march towards a science based PDD profession is inexorable, and it begins with an open mind.

References

Aronson, E. & Mills, J. (1956). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 59, 177–181.

Brehm, J. (1956). Post-decision changes in desirability of alternatives. *Journal of Abnormal and Social Psychology*, 52(3), 384–389.

Cooper, J. (2007). *Cognitive Dissonance: 50 Years of a Classic Theory*. London, England: Sage publications.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Festinger, L., & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58(2), 203–210.

Greenwald, A.G. & Ronis, D.L. (1978). Twenty Years of Cognitive Dissonance: Case Study of the Evolution of a Theory. *Psychological Review* (American Psychological Association) 85 (1): 53–57.

Harmon-Jones, E., & J. Mills. (Eds.) (1999). *Cognitive Dissonance: Progress on a Pivotal Rheory in Social Psychology*. Washington, DC: American Psychological Association.

Replication and Extension Study of Directed Lie Screening Tests: Criterion Validity with the Seven and Three Position Models and the Empirical Scoring System

Raymond Nelson, Mark Handler, Benjamin Blalock and Nayeli Hernández

Reprinted with permission of the American Polygraph Association
This article originally appeared in the APA Publication, Polygraph,41, (3)

Abstract

Two experienced examiners completed blind scoring tasks on 49 Directed Lie Screening Tests (DLST), also known as the Test for Espionage and Sabotage (TES), conducted by seven inexperienced examiners on 8 non-naive examinees who participated in a mock espionage scenario at a forward operating base in Iraq. Seven-position scores, using the US Federal test data analysis (TDA) model, were transformed to three-position and Empirical Scoring System (ESS scores. Monte Carlo models were used to calculate the distributions of seven-position, three-position and ESS scores, and the results were analyzed using multivariate ANOVAs. Unweighted decision accuracy and inconclusive rates using the seven-position scores did not differ significantly from previous studies of the TES at the U.S. Department of Defense. Criterion accuracy for the seven-position, three-position and ESS TDA models was significantly greater than chance. Only the ESS model produced both test sensitivity to deception and test specificity to truth-telling that were significantly greater than chance. The three-position TDA model produced significantly more inconclusive results that were loaded on deceptive cases. The seven-position and ESS scores were found to extract similarly useful diagnostic information from the raw data. Pairwise decision agreement was significantly greater than chance for all models. Results support the criterion validity of the DLST and suggest continued interest in this technique.

Introduction

The directed lie screening test (DLST) (Handler, Nelson & Blalock, 2008; Nelson & Handler, 2012; Nelson, Handler & Morgan, 2012) was based upon the Test for Espionage and Sabotage (TES) (Department of Defense, 2006; Research Division Staff,

---

We are extremely grateful to the USF-I ITAM-MoD Intelligence Division and ITAM Police Missions who sponsor training for Akram Sabri Jwad Al NDawi, Mohammed Ahmed Mufeed Kider, Rabea Minhal Araf Al Rubaii, Mohammed Abdul Jabar Al Dulaymi, Mahmood Shaker Raheem, Mohammed Ali Kader, Mina Khadim Al Juburi, Baydaa Hammood Al-Hadeethi, Hassan Falih Hatim (Algaboory), Noor Ismaeel (Al Rubaee), Mohammed Khames Dhari (al Delemi), Asaad Kazim Hassan. Without the commitment of these dedicated professionals none of this work would have been accomplished.

Raymond Nelson is a research specialist with the Lafayette Instrument Company and an elected member of the APA Board of Directors. The views expressed in this work are those of the author and not the LIC or the APA. Mr. Nelson is a psychotherapist, polygraph field examiner, developer of the OSS-3 scoring algorithm, and is the author of several publication on various polygraph topics. Unrestricted use of this work is granted to polygraph training programs accredited by the American Polygraph Association, or recognized by the American Association of Police Polygraphists or the National Polygraph Association. For information contact raymond.nelson@gmail.com.

Mark Handler is an experienced police examiner and polygraph researcher who helped develop the Objective Scoring System, version 3 and the Empirical Scoring System. Mr. Handler is the research chairperson for the American Association of Police Polygraphists, and has published numerous articles and studies on many aspects of the polygraph. He can be reached at polygraphmark@gmail.com.

Benjamin Blalock is a former federal examiner and APA certified primary instructor, now in private practice. Mr. Blalock can be reached at ben@polygraphtoday.com.

Nayeli Hernández Pimentel is a polygraph examiner in private practice, a member of the APA, and a licensed psychologist in Mexico. Ms. Hernández can be reached at nayeli.psic@gmail.com.

1995a; 1995b) and has been adapted to screening use in public safety selection and post-conviction supervision programs. Prior to the development of this format, Psycho-physiological Detection of Deception (PDD) screening formats consisted primarily of the family of Modified General Question Techniques (MGQT), General Question Techniques (GQT), and the Relevant/ Irrelevant Technique (R/I), which did not include comparison questions and is scored globally or impressionistically and not numerically. The DLST is conducted in the absence of any known incident, known allegation, or known problem and is designed for use with multiple independent targets for which it is conceivable that an examinee may be involved in one or more target behaviors while remaining uninvolved in other investigation targets.

The DLST is similar to other PDD formats in its use of test questions, including the use of multiple presentations of a thoroughly reviewed sequence of relevant questions (RQs), comparison questions (CQs), and other procedural questions. Unlike other PDD screening formats, the DLST was designed to maximize testing efficiency with several presentations of all test stimuli within a single test question sequence and is always conducted using directed-lie comparison (DLC) questions.

Development studies on the TES/DLST were based on the seven-position manual test data analysis (TDA) method taught at the Department of Defense during the 1990s (Department of Defense, 2006). Although results of the TES/DLST studies by the Research Division Staff (1995a, 1995b) have been published, neither the data nor any statistical description of the sampling distribution is available for comparison with other samples. Nelson and Handler (2012) used Monte Carlo methods to show that DLST examinations can be interpreted using the Empirical Scoring System (Blalock, Cushman & Nelson, 2009; Handler, Nelson, Goodson & Hicks, 2010; Nelson, Blalock, Oelrich & Cushman, 2011; Nelson & Handler, 2010; Nelson & Krapohl, 2011; Nelson et al., 2012) and Objective Scoring System version 3 (Nelson, Krapohl & Handler, 2008) with criterion accuracy that is significantly greater than chance. The present study was designed to replicate the results of the seven-position studies conducted by the Department of Defense, and extend our knowledge of criterion accuracy of DLST with the three-position TDA models. The hypothesis was that blind scored results of confirmed DLST examinations from a laboratory study, including results using the seven-position, three-position, and ESS TDA models, can differentiate deception from truth-telling at rates that are greater than chance.

Method

Eight polygraph examiner trainees, employed with the Ministry of Defense and Ministry of the Interior in Iraq, participated in this study during their ninth week of training. Three of the participants were female. Ages of the participants ranged from 28 to 42 years. All of the participants had completed four-year college degrees. None of the participants were taking medications for chronic pain, cardiovascular illness, or mental health reasons. Participation in the study was voluntary, and had no effect on the training or employment status of the participants. No harm came to any of the participants as a result of participation in this study.

This study took place in Iraq, in an area known as Forward Operating Base (FOB) Union III. All participants in this study functioned as both PDD examiner and examinee. A laboratory scenario was developed in which study participants were randomly assigned to guilty and innocent groups, with four participants in each group.

Guilty participants were assigned to commit a mock espionage scenario, in which they were told to open an envelope and follow the instructions inside. Instructions required the guilty participants to leave the training room individually at predetermined times and walk to a nearby location where they were to hand an envelope, marked "secret information" to a man wearing a blue shirt with the number "3" on his sleeve. The man identified himself as a member of an anti-government group. The man wearing the blue shirt was a confederate in the study, and a linguist contractor working in support of U.S. forces and the Iraqi government. The envelope marked "secret information" contained a blank business card, and no secret

information was actually released to persons associated with anti-government groups as a result of this study. In exchange for the envelope the confederate gave each guilty participant a token that could be exchanged for merchandise at the post exchange (PX). Innocent participants were provided identical envelopes, that contained information instructing them to leave the training room individually at predetermined times, walk to a nearby location, and then return to the training room. Innocent participants were instructed to answer that they were taking a break for some exercise if questioned by anyone regarding their presence outside the training room.

Following the completion of the scenario, each participant was tested by each of the other participants using the DLST format. Examination questions, including investigation target questions, directed lie comparison questions, and procedural questions were standardized for all participants. All examinations were conducted in Arabic. Examination targets pertained to providing secret information to persons belonging to anti-government groups, and having unauthorized contact with persons belonging to anti-government groups. Testing activities took place over two days. Because the examinees were already familiar with the polygraph technique and instrumentation, all examinations were conducted without the use of an acquaintance test after ensuring the proper adjustment and functioning of the instrument. Participants were required to repeat examinations that resulted in inconclusive results. Nine inconclusive examinations were repeated. Four of those examinations resulted in a deceptive classification after retesting. No post-test discussion was completed following any of the examinations. However, the participants were provided an opportunity to debrief the experience individually and as a group following the completion of all study activities. Participants were required to maintain secrecy regarding their role during study, and there were no discovered lapses or breaches of information for the roles of the participants.

Study participants were given one day of instruction and practice using the DLST before beginning the study activities. The original design was for the eight study participants to conduct seven examinations on the other participants, for a total of 56 examinations. However, one participant became sick during the study. This participant was tested by the other participants but was not able to function adequately to participate effectively as an examiner. The participant was released from the study and the remainder of the field PDD training requirements due to the illness. Forty-nine examinations were completed, including 24 examinations of guilty participants and 25 examinations of innocent participants.

Blind scores were obtained from two examiners using the seven-position model for TDA (Department of Defense, 2006), the third and fourth authors (BB and NH). Both blind scorers were trained at schools accredited by the American Polygraph Association (APA). One blind scorer was trained at the U.S. Department of Defense and is an APA Primary Instructor with five years of experience. The second blind scorer is a native Spanish speaking bilingual examiner, an APA member working in Mexico, and with less than two years of field experience in police polygraph screening programs. Blind scorers worked independently from each other. Seven-position scores were transformed to their corresponding three-position values, and the electrodermal scores were weighted to produce ESS scores.

Cutscores and decision rules for seven-position and three-position scores were those specified by the Department of Defense (2006). All subtotals were required to be positive and the grand total score must equal or exceed four to be considered a No Significant Response (NSR) result. Any examination with a subtotal of -3 or less or a grand total of -4 or less would be classified as Significant Response (SR). Examinations meeting neither of those conditions would be classified as Inconclusive (INC).

The decision rule for the automated ESS model was the spot-score-rule (SSR) (Light, 1999; Swinford, 1999). Alpha was set at .05 for deceptive classifications and alpha = .1 for truthful classifications. ESS cutscores corresponding to these alpha levels were -3 and +1, using the normative data shown by Nelson, Handler and Morgan (2012). Any subtotal score of -3 or lower would be

statistically significant for deception ($p < .05$), while test results in which all subtotal scores are +1 or greater would be statistically significant for non-deception ($p < .1$). Bonferonni correction to the alpha cutscore for deceptive classifications was not used with the DLST examinations because the SSR is premised on the assumption that the criterion variance of individual questions is not affected by and does not affect the other questions.[1] However, an inverse of the Šidák correction for independent issues is used to correct for the deflation of alpha that occurs when calculating the normative probability that an examinee would produce a statistically significant truthful result to all investigation targets while lying to one or more of the independent issues.

Means, standard deviations, and statistical confidence intervals were calculated for a dimensional profile of criterion accuracy, including: sensitivity, specificity, inconclusive results for deceptive and truthful cases, false-positive and false-negative errors, positive predictive value, negative predictive value, percent of correct decisions for the deceptive and truthful cases, and the unweighted means of the percentage of correct decisions and inconclusive results for deceptive and truthful cases. A three-way ANOVA (study x status x criterion dimension) was calculated to compare the decision accuracy and inconclusive rates to those reported by the U.S. Department of Defense (Research Division Staff, 1995a, 1995b). Post-hoc analyses were completed as necessary. A second three-way ANOVA (TDA model, x status x criterion dimension) and post-hoc analyses was completed to compare the unweighted means of the percentage of correct decisions and inconclusive results of the seven-position, three-position, and ESS scores.

Results

All statistical results were evaluated with a level of significance set at alpha = .05.

Sample distributions

Seven-position scores from the two blind scorers produced a mean deceptive subtotal score of -1.833 (SD = 4.099) and a mean truthful subtotal score of 3.670 (SD = 3.443). Three position scores resulted in a mean deceptive subtotal score of -1.458 (SD = 2.784) and a mean truthful subtotal score of 2.470 (SD = 1.853). ESS scores produced a mean deceptive subtotal score of -1.781 (SD = 4.437), and a mean truthful subtotal of 3.636 (SD = 2.917).

Interrater reliability of numerical scores

The proportion of decision agreement was significantly greater than chance for both seven-position and three-position scores. Seven-position scores resulted in a pairwise proportion of decision agreement of .722 (95% CI = .580 to .864). Three-position scores resulted in a pairwise proportion of decision agreement of .761 (95% CI = .602 to .919). Decision agreement did not differ significantly for the two TDA models. ESS scores produced a pairwise proportion of decision agreement of .796 (95% CI = .652 to .940). Decision agreement did not differ significantly for the three TDA models.

Replication

Decision accuracy and inconclusive rates of the seven-position scores were compared to the results reported for seven-position scores in studies reported by the U.S. Department of Defense (Research Division Staff, 1995a; 1995b). Table 1 shows the unweighted average accuracy and unweighted inconclusive rates from the replication and U.S. Department of Defense studies.

Figure 1 shows the mean plots and 95% confidence intervals for the proportions of correct decisions excluding inconclusive results, and inconclusive results of the deceptive and truthful seven-position scores from the present replication and U.S. Department of Defense studies. Table 2 shows the results of a three-way ANOVA (study x

[1] It is often the case that the behavioral details of the investigation target questions are not completely independent.

criterion status x criterion dimension). The three-way interaction between study, status, and criterion dimension was significant, in addition to the two-way interaction between status and criterion dimension, and the main effect for criterion dimension. The main effect comparing the studies was not significant in the three-way analysis, nor was the interaction of study and criterion status.

Table 1. Means, (standard errors), and {statistical confidence intervals} for DLST exams

	7 Position Replication	7 Position DoD Studies
D Correct	0.681 (0.089) {0.507 to 0.856}	0.821 (0.082) {0.66 to 0.982}
T Correct	0.978 (0.021) {0.937 to 1.02}	0.845 (0.082) {0.685 to 1.006}
D Inc	0.145 (0.048) {0.05 to 0.24}	0.087 (0.056) {-0.022 to 0.198}
T Inc	0.039 (0.026) {-0.013 to 0.092}	0.073 (0.051) {-0.028 to 0.175}

Figure 1. Means and 95% confidence intervals for correct decisions and inconclusive results from seven-position replication and US Department of Defense studies

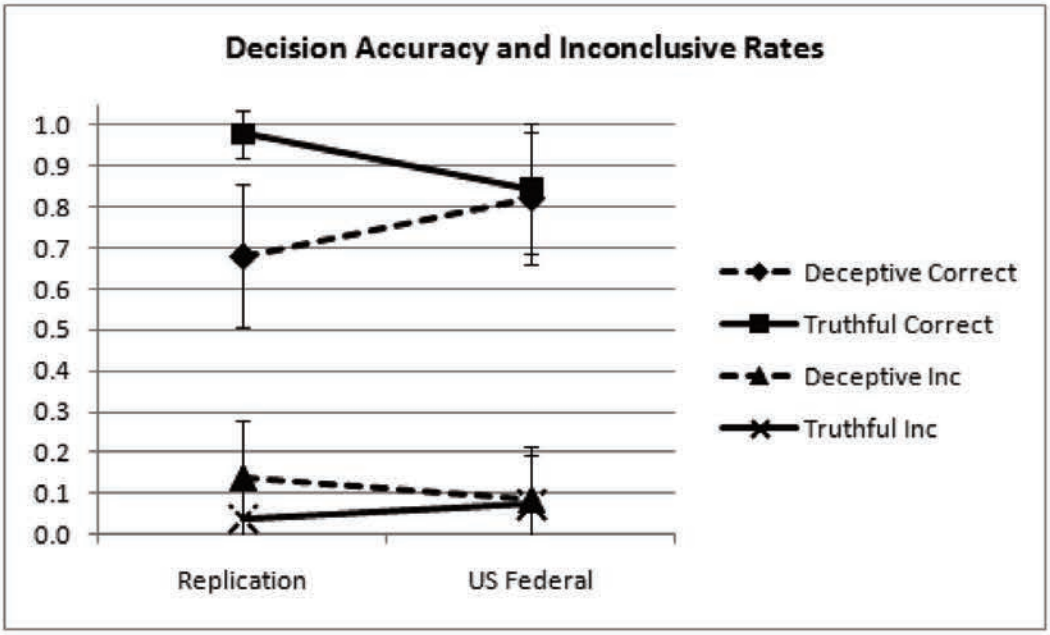


Table 2. Three-way ANOVA summary for Replication and U.S. Department of Defense studies

Source	SS	df	MS	F	p	F crit .05
Criterion dimension	57.761	1	57.761	12708.527	0.000	3.862
Status	0.088	1	0.088	19.297	0.000	3.862
Study	0.001	1	0.001	0.124	0.725	3.862
Criterion dimension x Status	1.441	1	1.441	317.144	0.000	3.862
Status x Study	0.162	1	0.162	35.734	0.000	3.862
Criterion dimension x Study	0.004	1	0.004	0.890	0.346	3.862
Criterion dimension x Status x Study	6.817	1	6.817	1499.793	0.000	3.862
Error	2.036	448.000	0.005			
Total	68.310	455				

A series of one-way unbalanced post-hoc ANOVAs was used to investigate differences between the study results. Differences in decision accuracy were not significant for deceptive cases [F (1,65) = 0.008, (p = .276)] or for truthful cases [F (1,81) = 2.320, (p = .132)]. Differences in inconclusive rates were also not significant for deceptive cases [F (1,65) = 0.357, (p = .552)] or for truthful cases [F (1,81) = 0.280, (p = .598)]. Table 3 shows the unweighted mean decision accuracy and unweighted mean inconclusive rates from the replication and development studies.

Under many circumstances the significant interaction between case status and criterion dimension, as shown in Table 2, would limit the evaluation and reporting of accuracy and inconclusive rates to the separated deceptive and truthful groups, as shown in Figure 1. However, because PDD field examiners and program administrators may be interested in a measure of combined test effectiveness, and because none of the one-way ANOVAs was significant, decision accuracy and inconclusive rates for the combined deceptive and truthful cases are shown in Table 3 and Figure 2. The two-way ANOVA summary (criterion dimension x study) is shown in Table 4, and indicates that the interaction of study and criterion dimension was not significant for the combined groups. The main effect for study was not significant, and the significant main effect for criterion dimension is expected.

Figure 2 shows the interaction plot of the unweighted mean decision accuracy and unweighted mean inconclusive rates from the present replication and the U.S. Department of Defense studies.

Table 3. Means, (standard errors) and {95% CI} for unweighted accuracy and inconclusives

	7 Position Replication	7 Position DoD Studies
Unweighted average accuracy	.829 (.051) {.727 to .930}	.833 (.056) {.723 to .944}
Unweighted average inconclusive results	.090 (.040) {.011 to .169}	.080 (.037) {.006 to .154}

Figure 2. Mean and 95% confidence intervals for decision accuracy and inconclusive rates

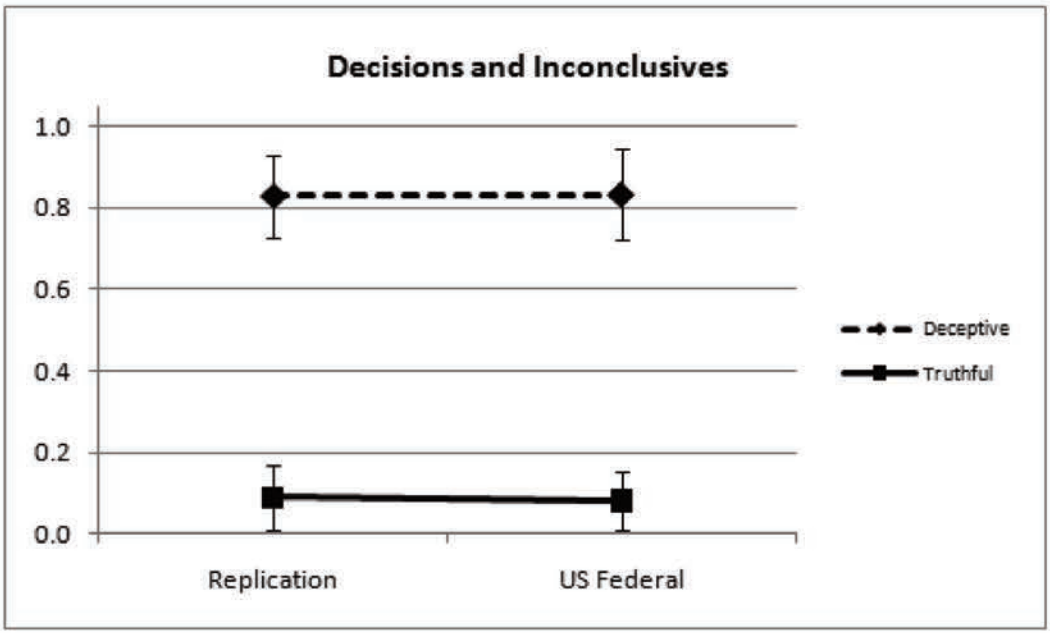


Table 4. Two-way ANOVA summary for decision accuracy and inconclusives of combined deceptive and truthful groups

Source	SS	df	MS	F	p	F crit .05
Study	0.000	1	0.000	0.002	0.964	3.890
Criterion Dimension	27.269	1	0.278	127.847	0.000	3.890
Interaction	0.002	1	0.002	1.103	0.295	3.890
Error	0.418	192	0.002			
Total	27.272	195				

Table 5. Means, (standard errors), and {statistical confidence intervals} for DLST exams

	7 Position Scores	3 Position Scores	ESS Scores
Unweighted Accuracy	.830 (.037) {.756 to .903}	.816 (.043) {.731 to .902}	.858 (.036) {.786 to .929}
Unweighted Inc	.092 (.027) {.038 to .146}	.248 (.042) {.164 to .331}	.123 (.033) {.057 to .188}
Sensitivity	.583 (.070) {.444 to .722}	.415 (.071) {.276 to .555}	.665 (.068) {.532 to .799}
Specificity	.94 (.032) {.876 to .999}	.848 (.050) {.750 to .947}	.839 (.050) {.740 to .938}
FN Error	.271 (.061) {.150 to .392}	.228 (.060) {.110 to .347}	.207 (.058) {.092 to .322}
FP Error	.020 (.020) {.001 to .060}	.009 (.013) {.001 to .036}	.040 (.027) {.001 to .094}
D Inc	.145 (.048) {.050 to .240}	.355 (.069) {.219 to .491}	.126 (.046) {.035 to .217}
T Inc	.039 (.026) {.001 to .092}	.141 (.047) {.047 to .235}	.119 (.045) {.029 to .209}
PPV	.966 (.033) {.900 to 1.032}	.977 (.032) {.913 to 1.041}	.942 (.038) {.866 to .999}
NPV	.775 (.052) {.673 to .878}	.789 (.054) {.683 to .896}	.800 (.057) {.689 to .912}
D Correct	.681 (.089) {.507 to .856}	.644 (.086) {.476 to .813}	.762 (.066) {.631 to .892}
T Correct	.978 (.021) {.937 to .999}	.988 (.016) {.956 to .999}	.954 (.031) {.893 to .999}

Criterion accuracy of seven-position, three-position and ESS scores of DLST exams.

Table 5 shows the dimensional profile of criterion validity for DLST exams when scored with seven position and three position models.

Figure 3 shows the mean plots and 95% confidence intervals for decision accuracy excluding inconclusive results and inconclusive rates for the DLST cases when scored with the seven-position, three-position, and ESS TDA models. Table 6 shows the results of a three-way ANOVA (criterion status x criterion dimension x TDA model) for decision accuracy and inconclusive rates. The three way interaction was significant, along with a significant two-way interaction between criterion dimension and TDA model. Main effects for TDA model, criterion dimension, and status were also significant in the three way analysis. The main effect for criterion dimension is expected as there is no reason why decision accuracy rates should not be different than inconclusive rates.

Figure 3. Means and 95% confidence intervals for correct decisions and inconclusive results

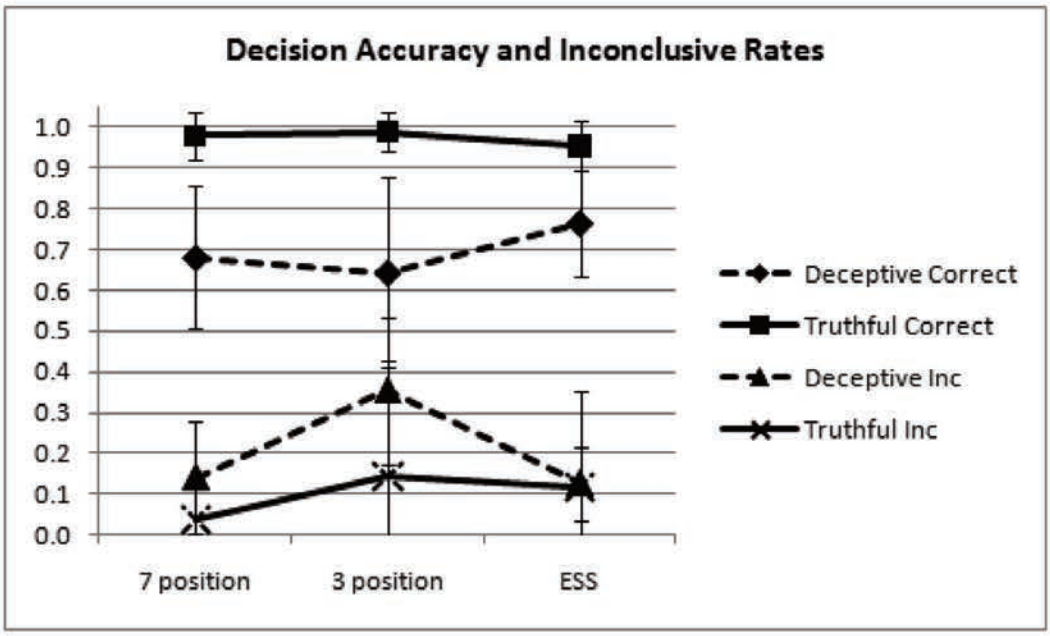


Table 6. Three-way ANOVA summary seven-position three-position and ESS scores

Source	SS	df	MS	F	p	F crit .05
Criterion dimension	33.956	1	33.956	7707.410	0.000	3.875
Status	0.243	1	0.243	55.193	0.000	3.875
TDA Model	0.263	2	0.131	29.847	0.000	3.028
Criterion dimension x Status	3.015	1	3.015	684.452	0.000	3.875
Status x TDA Model	0.013	2	0.007	1.503	0.224	3.028
Criterion dimension x TDA Model	0.481	2	0.240	54.534	0.000	3.028
Criterion dimension x Status x TDA Model	0.402	2	0.201	45.619	0.000	3.028
Error	1.242	282.000	0.004			
Total	39.615	293				

A series of post-hoc one way ANOVAs showed that the differences in decision accuracy, excluding inconclusive results, was not significant for deceptive cases [F (2, 69) = 0.423, (p = .657)] or for truthful cases [F (2, 72) = 0.358, (p = .700)]. Differences in inconclusive rates were not significant for truthful cases [F (2, 72) = 1.025, (p = .367)]. However, differences in inconclusive rates were significant for deceptive cases [F (2, 69) = 3.123, (p = .050)].

Under many circumstances the significant interaction between case status and criterion dimension, as shown in Table 6, would limit the evaluation and reporting of accuracy and inconclusive rates to the separated deceptive and truthful groups, as shown in Figure 3. However, PDD field examiners and program administrators may be interested in a measure of combined test effectiveness. For this reason, and regardless of the significant one-way effect for inconclusive results with deceptive cases, decision accuracy and inconclusive rates for the combined deceptive and truthful cases, are shown in Table 5 and Figure 4. The two-way ANOVA summary (criterion dimension x study) is shown in Table 7.

Figure 4. Mean and 95% confidence intervals for unweighted average decision accuracy and unweighted inconclusives for the seven-position, three-position and ESS TDA models

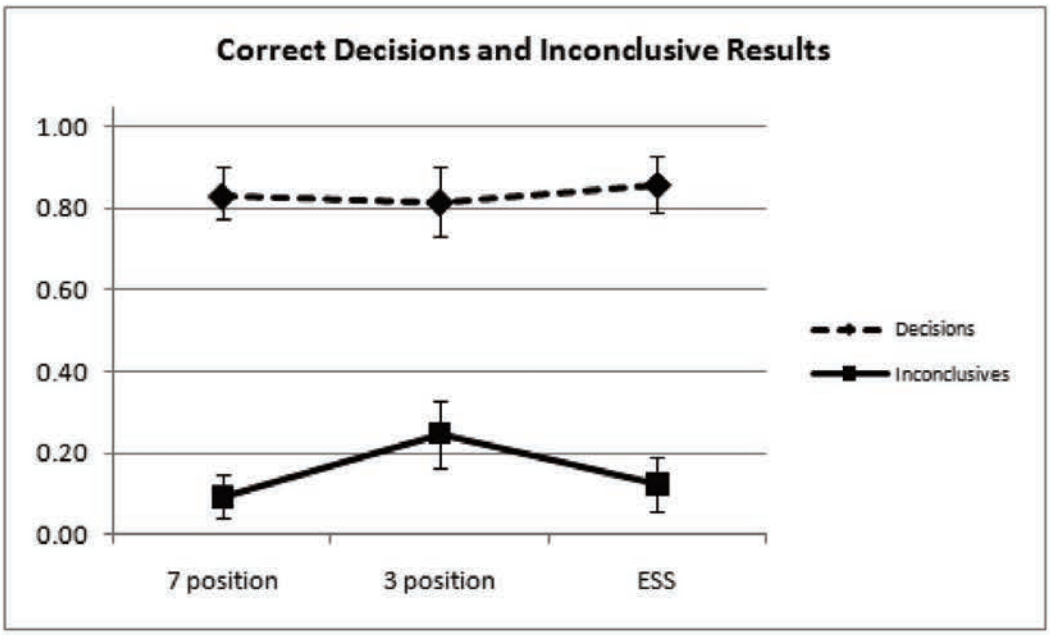


Table 7. Two-way ANOVA Summary

Source	SS	df	MS	F	p	F crit .05
TDA Model	0.249	2	0.003	1.886	0.171	3.874
Criterion Dimension	34.020	1	0.231	171.512	0.000	3.874
Interaction	0.464	1	0.464	343.763	0.000	3.874
Error	0.389	288	0.001			
Total	34.733	291				

The interaction of study and criterion dimension was significant for the combined groups. The significant main effect for criterion dimension is expected and uninteresting. However, main effect for TDA Model could not be interpreted due to the significant interaction. One-way post-hoc ANOVAs showed that the difference in decision accuracy was not significant for the three TDA models [F (2,144) = 0.304, (p = .738)], while the difference in inconclusive rates was significant [F (2,144) = 5.712, (p = .004)]. The three position TDA model produced more inconclusive results than the other models. Pairwise ANOVA contrasts of the TDA models revealed that the difference in inconclusive rates was significant for the three-position and ESS models [F (1,46) = 4.913, (p = .032)]. The three-position model produced significantly more inconclusives. The difference in inconclusive rates between the seven-position and three-position models was approaching a significant level [F (1,46) = 3.359, (p = .073)]. Differences in inconclusive rates for the seven-position and ESS models were not significant [F (1,46) = 0.314, (p = .860)].

Discussion

Results of this study replicate the unweighted DLST decision accuracy and inconclusive rates of DLST examinations, as reported in previous studies by the U.S. Department of Defense. All three TDA models, seven-position, three-position, and ESS, produced criterion accuracy that was significantly greater than chance. Interrater decision agreement, excluding inconclusive results, was significantly greater than chance for all three TDA models. Overall decision accuracy, excluding inconclusive results, did not differ significantly for the seven-position, three-position, or ESS models. However, only the ESS model produced both test sensitivity to deception and test specificity to truth-telling that were significantly greater than chance. The three-position model produced significantly more inconclusive results than the ESS, and the difference in inconclusive results was approaching a significant level for the seven-position and three-position models. Differences in inconclusive rates were significant only for the deceptive cases, suggesting that the component weighting achieved by the seven-position and ESS models may increase test sensitivity, making these models more effective at extracting diagnostic information indicative of deception. The absence of significant differences between the seven-position and ESS models suggests that these models are similarly effective at extracting and using diagnostic information. The inconclusive rates produced by the three-position model may be considered excessive for use in field PDD programs. It is possible that the use of normative data, optimal cutscores and improved decision rules could improve decision accuracy for the three-position TDA model. Additional research is recommended in this area.

Although the sampling distributions of scores are not available for direct comparison, these results suggest that differences may exist between the distributions of scores from this study and those of previous studies (Research Division Staff, 1995a, 1995b). Scores from the present replication were more effective with truthful cases, while scores from the previous studies appear to be more effective with deceptive cases. However, differences in results are not statistically significant, and the exact cause of these observed differences will remain unknown without further study. It is possible that these differences are the result of differences in study design, or to differences in internal and external motivation for the study participants. It is also possible that these differences are the result of differences in language and culture, classroom or professional relationships among the study participants, examiner experience, or the degree of naivety of the examinees regarding the PDD examination. Previous studies on the DLST involved experienced examiners and presumed naïve examinees, while the present replication was conducted under adverse circumstances, with inexperienced examiners testing examinees who were explicitly non-naive.

Despite the acknowledged differences, these results replicate the results of earlier studies, which showed that information regarding the PDD examination does not substantially degrade accuracy. In addition to the presumed adverse condition of testing non-naive examinees, the environmental conditions of the data collection were also adverse. The study location known as FOB Union III is located in a war zone. Indeed the

study and training facility was subject to an explosive rocket attack during the period of the study. The exact effect of these stressors on the performance of the study participants cannot be known.

An interesting aspect of these results, and the fact that examinations were conducted in the Arabic language, is that they demonstrate the ability of a highly standardized PDD test format, including RQs and DLC questions, to transcend language and cultural barriers and remain effective.

Limitations of the present study include the small cohort of scorers, small sample size, and the unknown degree to which the results of laboratory examination data obtained by inexperienced examiners and non-naive examinees are generalizable to field settings. One noteworthy result is that the pattern of inconclusive results in this study is contrary to that of the general trend in the literature. This study resulted in more effective performance with truthful cases and more inconclusive results with deceptive cases. Most previous studies produced the opposite pattern, and the exact reasons for these differences are unknown. Replication of this study and continued research is recommended.

An additional limitation to this study was that no attempt was made to investigate decision accuracy at the level of the individual RQs. Previous research has not supported the hypothesis of highly accurate decisions at the level of the individual questions, and decisions in this study were made at the level of the test as a whole when evaluating subtotal scores for individual questions. Future research should investigate DLST decision accuracy at the level of the individual question subtotals.

Regardless of these differences and limitations, the results of this study differ minimally from those of previous studies on the DLST, and support the validity of the DLST as capable of differentiating deception and truth-telling at rates that are significantly greater than chance under adverse circumstances. Continued interest in the DLST is recommended.

Acknowledgments

Special thanks to Chip Morgan and Sabino Martinez without whose assistance and professionalism this study would not have been possible.

References

Blalock, B., Cushman, B., & Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph*, 38, 281-288.

Department of Defense (2006). *Federal Psychophysiological Detection of Deception Examiner Handbook*. Reprinted in *Polygraph*, 40(1), 2-66.

Handler, M., Nelson, R., & Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph*, 37(2), 100-111.

Handler, M., Nelson, R., Goodson, W., & Hicks, M. (2010). Empirical Scoring System: A cross-cultural replication and extension study of manual scoring and decision policies. *Polygraph*, 39(4), 200-215.

Light, G.D. (1999). Numerical evaluation of the Army zone comparison test. *Polygraph*, 28, 37-45.

Nelson, R., Blalock, B., Oelrich, M., & Cushman, B. (2011). Reliability of the Empirical Scoring System with expert examiners. *Polygraph*, 40.

Nelson, R., & Handler, M. (2012). Monte Carlo Study of Criterion Validity of the Directed Lie Screening Test using the Empirical Scoring System and the Objective Scoring System Version 3. *Polygraph*, 41(3), 145-155.

Nelson, R., Handler, M., & Morgan, C. (2012). Criterion Validity of the Directed Lie Screening Test and the Empirical Scoring System with Inexperienced Examiners and Non-naive Examinees in a Laboratory Setting. *Polygraph*, 41(3), 176-185.

Nelson, R., & Krapohl, D. (2011). Criterion Validity of the Empirical Scoring System with Experienced Examiners: Comparison with the Seven-Position Evidentiary Model Using the Federal Zone Comparison Technique. *Polygraph*, 40(2), 79-85.

Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Cushman, B., Russel, C., & Oelrich, M. (2011). Using the Empirical Scoring System, *Polygraph*, 40(2), 67-78.

Nelson, R., & Handler, M. (2010). *Empirical Scoring System: NPC Quick Reference*. Lafayette Instrument Company. Lafayette, IN.

Nelson, R., Krapohl, D., & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37, 185-215.

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

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

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph*, 28, 10-27.

A Respiration Primer for Polygraph Examiners

Mark Handler, Joel Reicherter, Raymond Nelson, and Chris Fausett

Reprinted with with permission of the American Polygraph Association
This article originally appeaed in the APA Publication, Polygraph, 38 (2).

"Divide each difficulty into as many parts as is feasible and necessary to resolve it."
Rene Descartes

Introduction

The analysis of Psychophysiological Detection of Deception (PDD) data relies on the mathematical combination of scores assigned to a fixed set of physiological parameters. These parameters can be recorded and measured for their differential response to various stimulus questions describing past behavioral acts, and then inferentially and empirically correlated with attempts to deceive (ASTM International, 2002; Bell, Raskin, Honts & Kircher, 1999; Handler, 2006; Kircher & Raskin, 1988, 1999, 2002; Krapohl & McManus, 1999).

One of the required physiological signals in PDD testing is that of movement associated with pulmonary ventilation (breathing) (ASTM, 2002; Bell, Raskin, Honts & Kircher 1999; Department of Defense Polygraph Institute [DoDPI], 2006). Respiratory data are generally obtained via a pneumograph transducer placed around the thorax and abdomen of the test subject. Breathing movement is graphically displayed for evaluation by computer algorithms or traditional hand-scoring methods. When recorded electronically, the data are digitized and stored in numerical form.

PDD examiners have historically evaluated breathing movement data through a subjective approach that relies on the presence or absence of various signature patterns indicative of deception (DoDPI, 2006). Timm (1982a; 1982b) introduced the concept of the Respiration Line Length (RLL) as an objective, though general, measure of increases or decreases in respiration activity. RLL is obtained by measuring the length of the graphical respiratory wave form for a specified period of time and is evaluated by comparing one line length to another. Empirical studies have shown that greater decreases in breathing movement as measured by RLL are correlated with increased salience of stimuli during PDD testing (Harris, Horner & McQuarrie, 2000; Kircher, Kristjansson, Gardner, & Webb, 2005). Empirically-based hand-scoring methods emphasize only those patterns that have been reliably correlated with deception and most of those patterns result in decreased RLL (ASTM International 2002; Bell, Raskin, Honts & Kircher, 1999; Handler, 2006; Kircher & Raskin, 1988, 1999, 2002; Krapohl & McManus, 1999).

This paper is offered as a primer for the prospective and practicing polygraph examiner with respect to the "respiration channel" in PDD testing. It will provide descriptions of the physical organs and the mechanical properties involved in pulmonary ventilation. The paper will discuss some of the putative neural substrates that contribute to rhythmic breathing and those involved in altering that rhythm. It will offer evidence contrary to the historically taught "fight or flight" concept as the underlying cause of arousal and suggest the Orienting Response as an alternative hypothesis. Finally, the paper offers suggestions as to how it may be possible to extract more diagnostic information from the respiration channel and calls for further research.

Brief description of respiratory system and the mechanics of breathing

The primary function of the respiratory system is to supply the cells of the body with oxygen and to vacate the body of carbon dioxide. Pulmonary ventilation or breathing describes the collective actions that move air into and out of the lungs. External respiration describes the exchange of oxygen for carbon dioxide in the alveoli. Internal respiration describes the exchange of oxygen for carbon dioxide between the tissues and blood. Cellular respiration describes the cellular metabolic reactions that consume oxygen and produce carbon dioxide. PDD testing relating to this phenomena focuses on the first aspect, pulmonary ventilation (breathing), or the moving of air into and out of the lungs (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

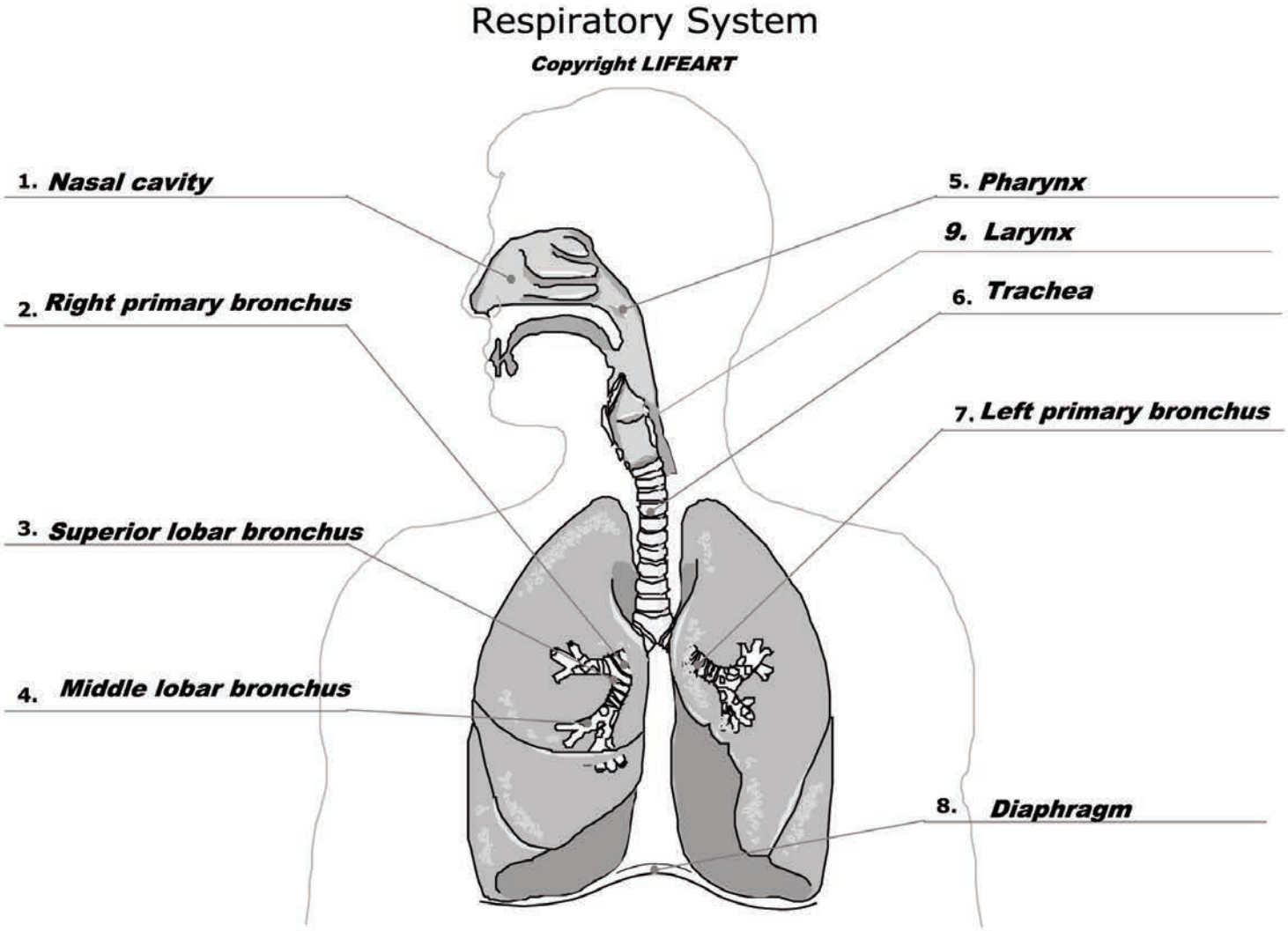


Figure 1 - A basic diagram of the airway path. Copyright LIFEART and reprinted with permission of LIFEART and SmartDraw, Inc.

Breathing involves moving air through the airway (dead air space) composed of the nasal cavity, pharynx, larynx, trachea, bronchi bronchial tree then into the lungs. The airway, through which the air travels, warms, humidifies and cleans the air before directing it to the lungs. The nasal passageway contains olfactory receptors which are unusual in that their input bypasses the

thalamus and is sent directly to cortical and limbic system areas of the brain that stimulate memory. The pharynx connects the nasal cavity and mouth to the larynx. The larynx is composed primarily of cartilage, vocal cords and other connective tissue, and connects the pharynx to the trachea. The trachea, composed of C shaped cartilaginous rings, is a flexible tube that connects the larynx to the bronchi. The bronchi enter the lungs and branch out to form secondary and tertiary bronchi leading to terminal bronchioles and finally into alveoli air sacs. Pulmonary capillaries surround the alveoli sacs providing the pathway for blood flow to and away from them. It is at this junction the exchange of oxygen for carbon dioxide takes place (Harver & Lorig, 2000; Marieb, 1999; Tortora & Grabowski, 1993). (See Figure 1.)

Smooth muscle of the airway from the trachea through the bronchioles is under control of the autonomic nervous system. Stimulation of adrenergic sympathetic fibers causes dilation of the bronchial muscle and decrease in mucous secretion. Evolutionary adaptations have apparently resulted in the bronchial smooth muscles being normally under greater parasympathetic control. This results in a tendency towards constriction of and increased secretion of mucous.

Ventilation Dynamics

The mechanics of breathing generates a pressure differential between the inside and outside of the lungs, causing air to move one direction or the other. Air, as with fluids, moves from areas of higher pressure to lower pressure regions. Just before inspiration, the differential pressure between the inside and outside of the lungs (intrapulmonary pressure) is zero and there is no air movement. The act of breathing causes the pressure inside of the lungs to be lower than that outside and thus air flows inward (Boyle's Law), similar to the concept of drawing a fluid up into a syringe. This negative intrapulmonary pressure is made possible by the expansion of the lungs resulting from the ventilation dynamics of the diaphragmatic and intercostal muscles (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

The muscles of normal, quiet inspiration (eupnea) include the diaphragm and the external intercostals. The diaphragm is a large, domed shaped muscle that separates the abdominal cavity from the thoracic cavity. The diaphragm is attached to the sternum and is the muscle most responsible for eupneic breathing. During normal quiet breathing the diaphragm contracts, causing it to descend about one half inch into the abdominal cavity. This results in stretching the thoracic cavity downward, increasing its volume. The diaphragm is innervated by the bilateral phrenic nerves which leave the spinal cord at the third, fourth and fifth cervical segments (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

Simultaneously, contraction of the intercostal muscles lift the rib cage and pull the sternum outward, like a handle on a bucket. The external intercostal muscles are innervated by nerves leaving the first through the eleventh thoracic segments of the spinal column.

Since the lungs are passive, they have no capacity to expand or contract on their own but rather are subject to external forces; much like a sponge absorbs and releases water. Each lung is encased by one continuous serous tissue folded over itself called the pleural membrane. The

parietal pleura portion is attached to the outer wall of the thoracic cavity with the visceral pleura bonding directly to the lungs. This creates a small space between the two pleurae which is called the interpleural space or pleural cavity. Both pleurae secrete a fluid into the cavity which reduces friction between them. Just prior to inspiration, the pressure within the pleural cavity is about 4mmHg below atmospheric pressure. This negative pressure between the pleura membranes keeps the lungs sucked to the chest wall thus preventing them from collapsing inward. As the thoracic cavity expands, the lungs are pulled into an expanded mode, reducing the pressure in the alveoli (intrapulmonic pressure), resulting in air pulled into the lungs (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993). (See Figure 2.)

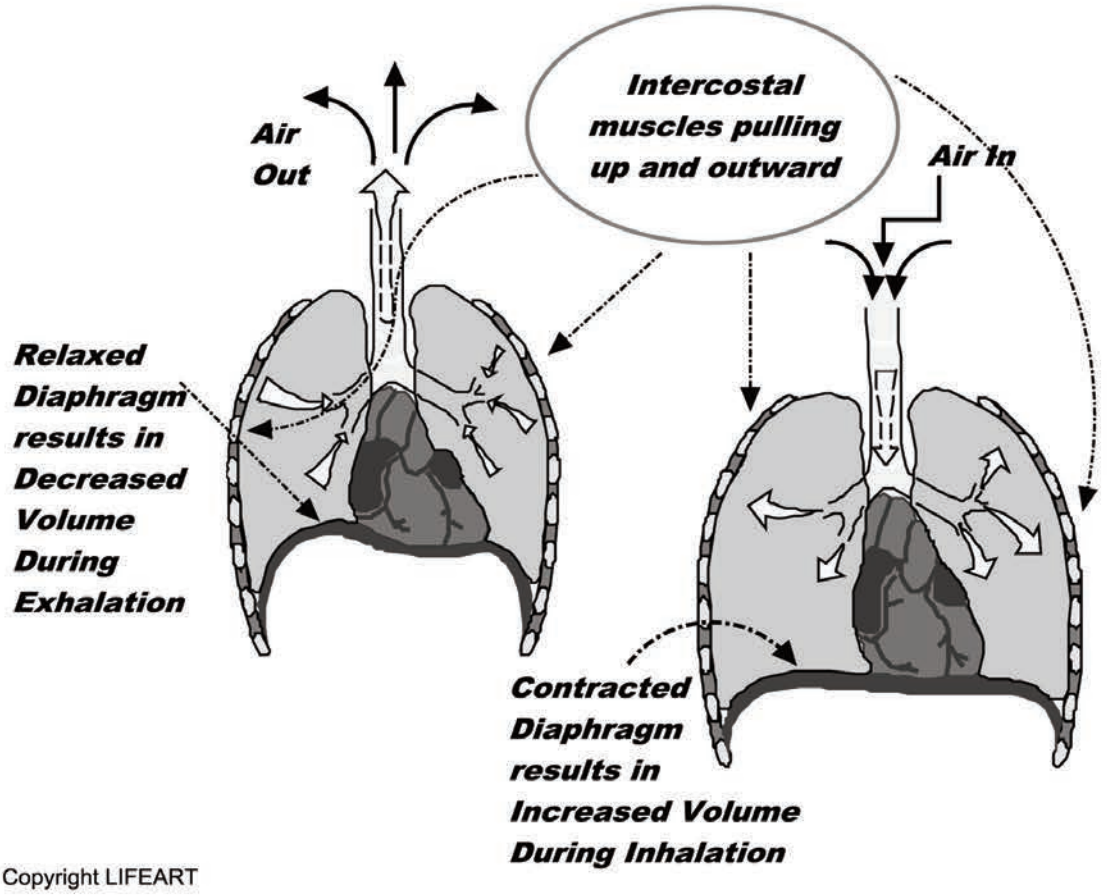


Figure 2 - Diagram showing how thoracic volumetric changes move air into and out of the lungs. During inhalation, the increase in the volume of the thoracic cavity results in air being drawn into the lungs. Copyright LIFEART and reprinted with permission of LIFEART and SmartDraw, Inc.

The combination of the contractions of the diaphragmatic and intercostal muscles results in an action that increases the thoracic cavity by approximately 500 milliliters. This increase causes a drop of intrapulmonary pressure of about 1-2 mmHg and air rushes into the lungs.

Expiration during eupnic breathing is passive and is accomplished through the elastic nature of the lungs and relaxation of the inspiratory muscles. As the muscles relax and the lungs recoil, the volume of the thoracic cavity decreases and there is no longer a difference in pressure

between the inside and outside of the lungs. Additionally, alveoli ducts and bronchioles have elastic fibers that recoil inward, expelling air. Finally, inward pull resulting from the surface tension of water vapor in the alveoli also contributes to lung volume decrease. The intrapulmonary pressure rises to about 1 mmHg above atmospheric pressure to force air out of the lungs (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

Regulatory Control of Breathing

Vegetative regulations of visceral body organs, including breathing dynamics, are controlled in part by nuclei and centers in the brain stem. The respiratory rhythmicity centers are located in the lower brain stem, medulla oblongata, with refining regulatory centers in the pons (see Figure 3). In the medulla, the rhythmic respiratory center is comprised of two distinct respiratory areas known as the dorsal respiratory group (DRG) and the ventral respiratory group (VRG). The DRG neurons are the primary innervators of the phrenic nerve and thus the diaphragm muscle (Harver & Lorig, 2000; Heimer, 1995; Janig, 2006).

The VRG, a column of individual nuclei stacked upon one another, contains mostly expiratory neurons and receives drive input from the DRG. The VRG is also involved in innervating the larynx and pharynx via vagal motoneurons which assists in maintaining airway patency. During inhalation, the VRG innervates the external intercostal muscles and has some connection to the phrenic nerve. Expiratory neurons originating in the VRG project to the internal intercostal muscles and abdominal muscles but these function mostly during intense and rapid exhalation such as during exercise when passive exhalation would take too long.

Modulatory centers such as the pontine respiratory group (formerly called the pneumotaxic) and a putative "apneustic center" located in upper area of the pons (see Figure 3) appear to be associated with phase-related activity. If nuclei exist that form an apneustic center it seems they may function as a "cut off switch" terminating inspiration. While this center has not been positively identified, it is presumed to be located at about the same level as the pontine respiratory group. Investigators who have experimentally transected the brain stem at this level have been able to produce apneusis (inspiratory spasms or cramps) but only if they also sever the vagus nerve. This suggests any "apneustic center" that exists receives input via the vagus nerves in order to prevent apneusis (Hlastala & Berger, 1996; Levitzky, 1999). While not well defined, the function of the respiratory related neurons in the pons seems to be to "fine tune" the action of eupneic respiration helping to provide a smooth transition between inspiration and expiration. The ponto-medullary respiratory rhythmicity center, however, can be influenced by the emotional limbic system centers as well as the cognitive cerebral cortical areas (Hlastala & Berger, 1996; Levitzky, 1999).

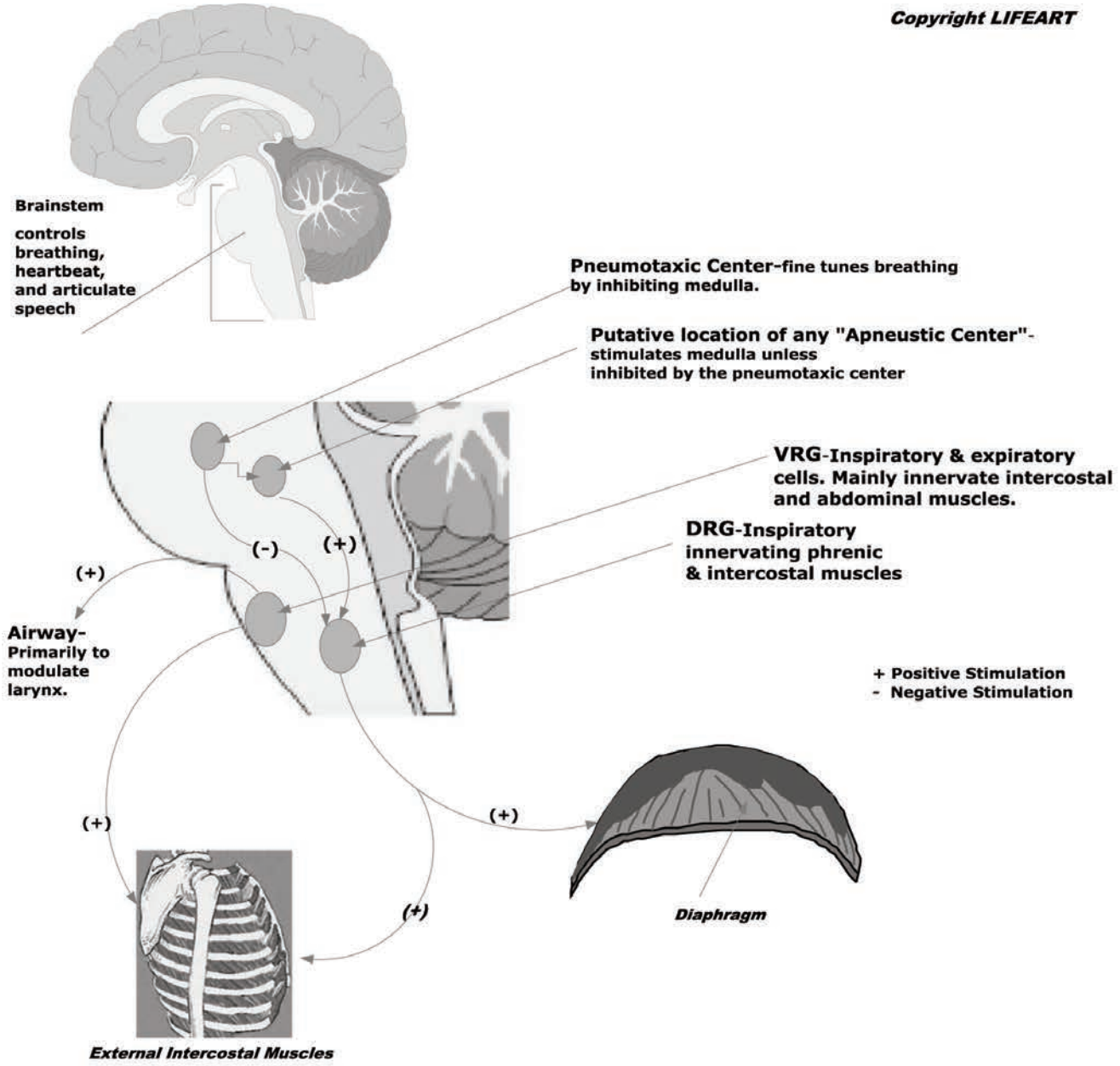


Figure 3 – General locations of central nervous system nuclei responsible for rhythmic regulatory control of breathing. DRG and VRG generalized location and effects on the diaphragm and intercostal muscles during eupneic breathing. Copyright LIFEART and reprinted with permission of LIFEART and SmartDraw, Inc.

Reflexes in the breathing cycle

Stretch receptors within the airways have the potential to influences the respiratory cycle. One such stretch receptor reflex, known as the Hering-Breuer inflation reflex, can result in decreased respiration drive. As the lungs expand through pulmonary inflation, it activates the sensors of these stretch receptors which project via the vagus nerve to the DRG and the pontine respiratory group. The end result is bronchial dilation and increased expiration time resulting in a decrease in respiration rate.

This seems to be a protective reflex which has developed to prevent the lungs from over-expanding (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

Irritant receptors are located throughout the airway and can be activated by such things as certain chemicals, gasses, smoke, dust and very cold air. Activation by these vectors is transmitted primarily by the vagus nerve and can result in bronchial constriction which functions to protect the airways from the noxious agent (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

Chemoreceptors are located centrally in the medulla and peripherally in the great vessels of the neck. The central chemoreptors are exquisitely sensitive to carbon dioxide, which is the most tightly controlled chemical factor. Carbon dioxide diffuses into the cerebral spinal fluid and forms carbonic acid which liberates hydrogen ions resulting in a drop in the pH of the cerebral spinal fluid. It is these hydrogen ions that actually excite the central chemoreceptors in the medulla which in turn stimulates ventilation. The peripheral chemoreceptors, however, are more responsive to oxygen levels in the blood. Chemoreceptors sensitive to oxygen are located in the aortic and the carotid bodies. If the circulating level of oxygen drops substantially, these act to stimulate respiration rate and depth. Under normal conditions, oxygen levels in the blood affects breathing only indirectly by enhancing the sensitivity of the central carbon dioxide sensors (Harver & Lorig, 2000; Hlastala & Berger, 1996; Levitzky, 1999; Marieb, 1999; Tortora & Grabowski, 1993).

Hypothalamic Integration

The hypothalamus is a collection of small nuclei located at the base of the forebrain in the ventral portion of the diencephalons (below the thalamus). The hypothalamus may be physically small, about 1% of the brain, but is metaphorically a grand maestro in the orchestration of bodily functions. The hypothalamus regulates a wide range of physiologic and behavioral functions and is the key controlling and integrating agent for homeostatic functions via the autonomic nervous system. The behavior control column of the hypothalamus is important to aspects of breathing control as it projects to somatomotor and autonomic control centers in the brain stem and spinal cord that determine respiratory behavior. Emotional, motor control and cognitive states generated in the telencephelon direct the hypothalamic activity, often via the behavior control column (Janig, 2006) (see Figure 4).

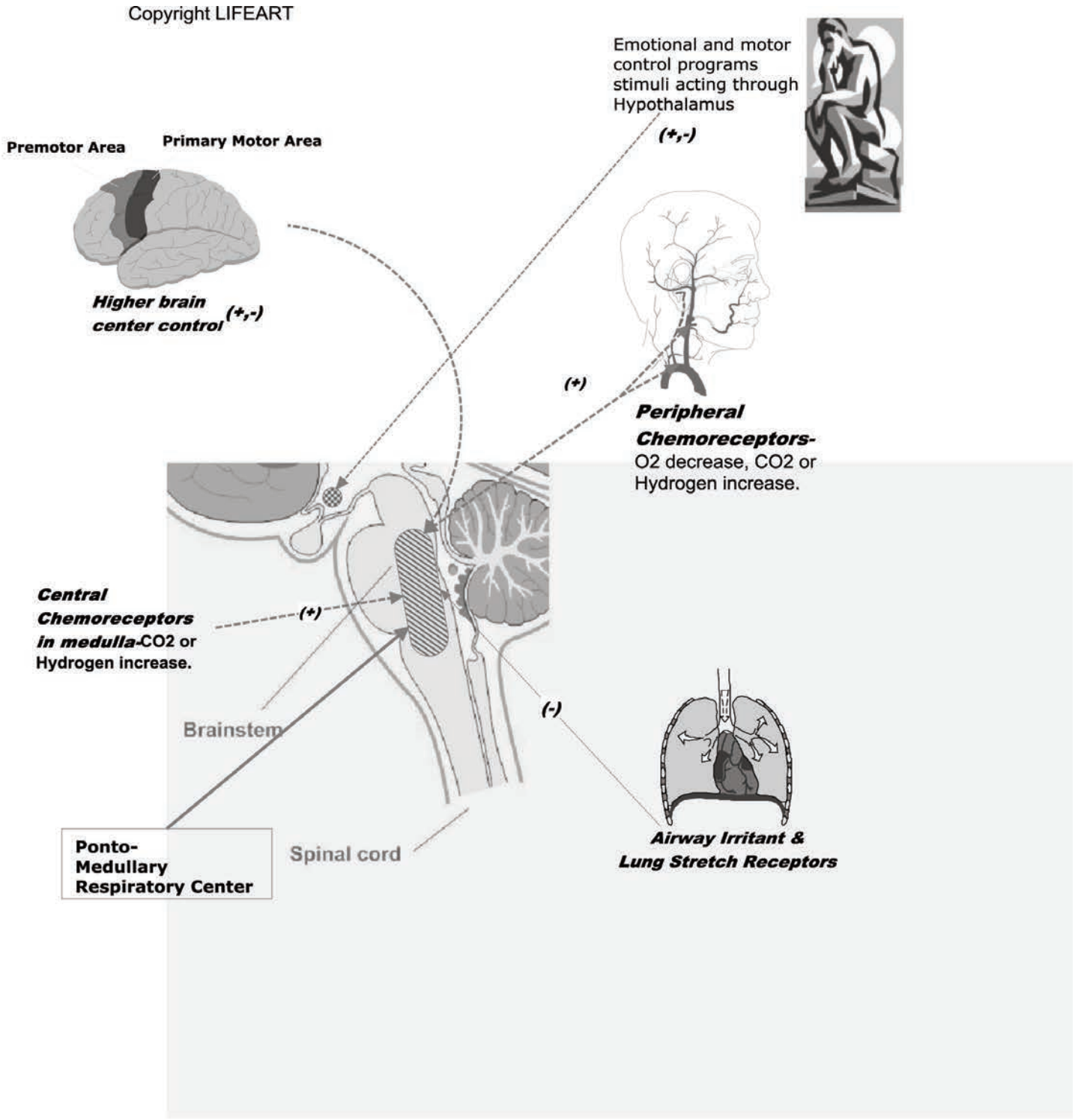


Figure 4 - Chemical and higher order regulation of breathing. Inhibitory influences are denoted with a (-) and excitatory influences by (+). Copyright LIFEART and reprinted with permission of LIFEART and SmartDraw, Inc.

Inhibition of breathing and a review of the Orienting Response

Breathing inhibition has been found to be a reliable indicator of arousal during polygraph testing (Harris et al., 200l; Kircher et al., 2005; Nelson et al., 2008). A number of theories have been proposed to explain the underlying cause of arousal during PDD testing (Handler & Honts, 2008, Handler & Honts, 2009; Kleiner, 2002; Lykken, 1974; Steller, 1987) and many of these incorporate some reference to the Orienting Response (OR).

The orienting response, first described by Pavlov (1927) as the "orienting reflex" was said to bring an immediate response in both human and animal to changes in their surroundings. Pavlov sometimes called it the "what is it" reaction, and noted it was of great significance for survival. Some of the stimuli that are known to cause an OR include: novelty, intensity, color, surprise, a conditioned stimulus, complexity, uncertainty or conflict (Pavlov, 1927).

The orienting response increased the probability of survival. Pavlov wrote "The biological significance of this reflex is obvious. If the animal were not provided with such a reflex, its life would hang at any moment by a thread" (Pavlov, 1927, p. 12). Pavlov's early description of the reaction discussed the postural changes and skeletal responses that seemed to be aimed at an investigatory and assessing response. These postural changes include: momentary cessation of motor activity (freezing), an orientation of the head towards the stimulus and an adjustment in receptors (pricking up the ears or a cocking of the head) towards the source of the stimulus. Pavlov believed the purpose of the OR was to prepare for better reception and response to a possibly threatening stimulus (Barham & Boersma, 1975) and he constructed the first known sound-proof room, "the silence tower", to study the OR.

Stimuli may be categorized as either signal or non-signal in nature. Signal stimuli are those that convey important information to the organism and may be regarded as significant (Sokolov, Spinks, Naatanen & Lyytinen, 2002). One example of a signal stimulus would be the sudden appearance of a deadly predator in the local area. Signal stimuli need not be physical in nature to solicit an OR. For example, verbal stimulus that captures the subject's attention can also invoke an OR. Non-signal stimuli, on the other hand, are those the organism considers neutral, and tend to convey no important information, such as different pure tones (Cacioppo, Tassinary & Berntson, 2000). Novel stimuli are initially signal stimuli as they convey to the organism that something new has happened and they reliably elicit an OR. If a novel stimulus is repeated but not paired with any meaningful consequence, the OR associated with it will decrease and eventually become extinct through a well-known process called habituation (Ohman, Hamm & Hugdahl, 2000). Habituated stimuli, though once novel, do not elicit ORs.

Significant stimuli are said to possess signal value, and can evoke an enhanced or greater OR (Gati & Ben-Shakhar, 1990). Sokolov (1963) determined that the significance or salience of a stimulus can affect the magnitude of an OR. He described *signal stimuli* as stimuli that are not novel but rather familiar and important. From a survival standpoint, it may be more beneficial to an organism to respond to a stimulus of known importance than one which is novel (Cacioppo, Tassinary & Bernston, 2000).

Sokolov found that an organism could self-assign salience to the particular stimulus and this salience may result from a previous experience or reflect an innate biologically programmed autonomic or behavioral response. While the OR can be an affectively neutral response, it may just as well be one that occurs concomitantly with an emotional stimulus (Ohman, Hamm & Hugdahl, 2000).

Signal value stimuli are thought to be associated with consequences which may be tied to memory (Ohman, 1979). An organism evaluates a stimulus and compares it to information stored in long-term or short-term memory. The current input is compared to active memory to determine if the stimulus is new and mismatched against previously encoded information, or if the stimulus matches an element of memory that has been primed to be significant (Cacioppo, Tassinary & Bernston, 2000). In either case, an assignment of novelty or significance can result in an OR. Both signal and non-signal ORs may have the initial cognitive function of information intake and processing of the stimulus. In the case of non-signal stimuli, a mismatch results in the OR occurring. The organism may compare the stimulus to information stored in memory and assign signal value, based on recognition and possible consequences, resulting in a signal value OR.

Data have shown stimuli with signal value will elicit larger and more slowly habituating ORs than non-signal ORs (Siddle, Stephenson & Spinks, 1983). While the response patterns for signal and non-signal ORs were similar, the underlying purpose may differ. Non-signal stimuli, novel stimuli, can evoke responses that may signal the organism that a potentially harmful or dangerous situation exists and prepare the organism to deal with that situation. Signal stimuli are those that can be elicited by recognized or familiar stimuli that have some degree of known significance (Sokolov, 1963).

Descriptions of the physiological responses associated with the OR in humans are well documented (Darrow 1936; Lynn 1966; Sokolov, 1963). These include; increased skin conductance, decreased heart rate, vasoconstriction in the limbs, an initial *delay in respiration rate and decrease in frequency*, and an increase in general muscle tonus. Possible benefits of the physiological response of OR are: Increased palmar perspiration allows for better tactile differentiation (Darrow, 1933), better hand grip (Boucsein, 1992; Darrow, 1933), and protection against injury (Adams & Hunter, 1969). Increased plantar perspiration allows for better footing; (Boucsein, 1992) an obvious benefit to bare foot runners and tree climbing primates. Vasoconstriction mobilizes reserve blood flow in preparation for fight or flight and may make the animal less likely to bleed as well as raise systemic blood pressure. Reduction in respiration results in quieting, making the animal less likely to be seen due to reduced movement and may result in increased olfactory intake. Dilation of the bronchioles reduces resistance which allows for a sustained level of oxygen intake with minimized movement associated with pulmonary ventilation.

Ideas surrounding breathing suppression relating to OR versus traditional discussions surrounding fight or flight

As early as the 19th Century, Darwin (1872, p.283) noted perception increased if body movement and breathing was quieted. Recent investigators (Stekelenburg & van Boxtel, 2001, 2002) have

reported findings of suppression of breathing during tasks that require attention which they relate to the OR. Recall that the OR is not limited to novel stimuli, but also occurs to stimuli the organism has determined to be important. An important consideration is the magnitude of reactions found to be commensurate with the salience or intensity of the stimulus (Gati & Ben-Shakhar, 1990; Siddle, Stephenson & Spinks, 1983; Stekelenburg & van Boxtel, 2001).

Stekelenburg and van Boxtel (2001) propose a centrally controlled prepotent motor control pattern that engages to enhance the body's sensitivity to stimuli. They concluded this pattern includes both inhibition of pericranial (facial) electromyographic (EMG) activity, and changes in respiration and heart rate when attending to external stimuli. Drawing a parallel with Sokolov's description of the OR (Sokolov, 1963), Stekelenburg & van Boxtel (2001) suggest the organism engages in a complex autonomic and somatomotor pattern of response that may have evolved to increase the chance for survival. One finding was that as attention increased, so did inhibition of breathing. Earlier studies also reported findings of breathing suppression and body movements during voluntary attention or involuntary orienting, particularly to low and moderate intensity stimuli (see, e.g. Graham, 1979; Sokolov, 1963; Turpin, Schaeffer & Bouscein, 1999).

The comparator theory proposed by Sokolov (1963) discussed the OR within a cognitive context. Sokolov proposed that repeated processing of sensory information gradually builds a "mental model" of the organism's surrounding world. Sokolov believed the organism assesses a stimulus and then compares that assessment to information already stored in memory. Sokolov stated that if a mismatch of information occurs between the incoming stimulus and the neuronal model, an OR will occur because of novelty, and thus is referred to as "novelty stimuli." If the organism detects no discrepancy between the stored and current input, an OR will not occur. For example, a rabbit eating grass in a field hears a rustling. It detects a difference or mismatch in local noises and it experiences an OR. This mismatch results in a non-signal OR that is nature's way of stopping the rabbit from eating to warn it there may be a threat nearby. If the rabbit were to keep eating the grass while a coyote was approaching, it may not survive the encounter. If the rabbit finds all is well it may well continue grazing.

Responses commonly associated with fight-or-flight reactions include increased heart rate, increased blood pressure, increased muscle tension, increased contractile force in the heart, vasoconstriction in the blood vessels supplying the skin and viscera (except the heart and lungs), vasodilatation in the blood vessels supplying the skeletal muscles and brain, transformation of glycogen into glucose which is released into the bloodstream for energy, sympathetic impulses to the adrenal medulla to cause the release of epinephrine and norepinephrine into the bloodstream, reduction in digestive actions, increase in respiratory passageways and an *increase in the rate of respiration* (Cannon 1929; Rathus, 2001;Tortora & Grabowski 1993).

The increase in respiration rate is incongruent with what has been empirically found to be diagnostic during polygraph testing. Deeper, faster breathing would result in an increase in RLL which is untenable given the results of recent studies of the respiration channel during polygraph testing.

The OR, RLL and Polygraph Testing

In the context of a polygraph test "significant stimuli" can be in the form of polygraph test questions to which the examinee assigns the greatest importance. A match between the stimulus and a mental representation can elicit a "significance-OR" (Verschuere, Crombez, De Clercq, & Koster, 2004). Lykken (1974), while discussing the Guilty Knowledge Test, stated "…for the guilty subject only, the 'correct' alternative will have a special significance, an added 'signal value' which will tend to produce a stronger orienting reflex than a subject will show to other alternatives"(p. 728). The measured strength of the OR has been correlated with stimulus intensity and can be produced at both low and high intensities (Lynn, 1966). Since perceived stimulus salience might well be linked to the memory that is associated with the event, it is reasonable to assume the crime related stimuli will produce significant ORs in a deceptive examinee. A person who is truthful to the relevant issue has no memory of the crime; however, they are likely to assign greater salience to the comparison questions.

There is an abundance of empirical data from polygraph testing supporting the idea that reduction of RLL infers salience and contributes to decisions of truthfulness or deception (Harris, Horner & McQuarrie, 2000; Kircher, Kristjansson, Gardner, & Webb, 2005). RLL has become the primary metric by which the respiration channel is evaluated and arguably, the OR contributes to some degree to this reduction in RLL.

RLL encompasses several breathing waveform patterns affecting depth and rate into a single metric. Overall when the subject perceives the question with greater jeopardy, the behavior pathway is to conserve energy until needed. Thus when greater jeopardy is perceived cognitively along with the emotional mix generated out of the limbic system, an increase in sympathetic arousal is launched. The receptor mix in the airway receives norepinephrine resulting in dilation with a consequent reduction in airflow resistance. Because air flow is increased through a dilated airway, diaphragmatic and intercostal muscle contraction can be reduced which is reflected in lower respiratory waveform amplitude. As a consequence not only is the ventilation amplitude decreased but the respiratory cycles are reduced. In sum, the RLL is shortened as question jeopardy perception is increased.

Attempts to separate OR from emotion

Recent attempts have been made to separate deception from the OR using a Concealed Information Testing paradigm. Ambach, Stark, Peper, and Vaitl (2008) reported results from attempts to uncouple deception from orienting. Ambach et al. (2008) concluded decreases in RLL measurements were a result of deception or preparation to deceive, as opposed to OR processes. We are not entirely convinced that deception and ORs are mutually exclusive. Perhaps an alternative explanation is the act of deception, including preparation and execution, will raise the salience of a stimulus. Such increases in salience have been shown to result in larger ORs (Gati & Ben-Shakhar, 1990; Siddle, Stephenson & Spinks, 1983; Stekelenburg & van Boxtel, 2001). It may well be that deception and salience are inextricably intertwined and may not be separable. In the Amach et al. experiment, the increase in signal value from deception clearly made a difference in the RLL measurements which they inferred resulted from different mental processes. This suggests there is a difference when deception is introduced but does not definitively prove *why* deception decreased RLL. A more parsimonious suggestion is that "deception" increased salience of those targets thereby enhancing the OR. An additional

consideration is that the RLL measurement is less sensitive to differences in salience. This has important implications in the weighting of RLL data in automated polygraph scoring systems. Published weighting of components in such algorithms consistently favors EDA over RLL (Nelson et al. 2008).

It is possible that similarly designed follow-up experiments using within-subject trials may shed light on whether dual mental processes are involved. By assigning test subjects to both truthful and deceptive conditions, it may be possible to determine if there are differences in how individuals' self-assign salience to objects when they are truthful compared to when they are deceptive. Greater individual arousal during the deceptive versus truthful condition would tend to confirm an increase in salience attributable to deception. Finally, increasing the sample size of the truthful group may increase the power to detect statistical significance with smaller effect sizes.

Conclusions and Recommendations

The purpose of this paper is to provide the practicing polygraph examiner with a primer of breathing mechanics and the OR as they may relate to polygraph. The article is part of a series directed to provide a more economical, defensible and satisfactory description of what we measure during polygraph testing. We discussed the current understandings of regulatory versus voluntary control of breathing as a foundation for understanding how the OR may contribute to breathing inhibition during polygraph. We suggest in an ideal polygraph testing setting changes to breathing occur unconsciously as a result of attending to the stimuli. We provided past and current research to support the notion of the OR contributing to breathing inhibition during polygraph testing. We provided some discussion surrounding why fight-flight-freeze behavior as described in the literature is difficult to reconcile with existing empirical data from polygraph studies. We suggest the idea of degrees of salience as a possible contributor to differences in breathing inhibition during deceptive response activity. To the extent that existing research in scoring feature development has demonstrated the contribution of the breathing inhibition to the detection of deception, we suggest that the OR provides a more satisfactory explanation for observable and measurable changes in respiratory activity during periods of deception, than does traditional fight-or-flight explanations. We hope continued research will attempt to further evaluate the role of orienting as a basis for response in the comparison question testing paradigm.

Based on current research surrounding the OR we offer several suggestions for improving respiration data quality which could ultimately lead to extracting more useful data from the respiration channel. These include; use of a Silent Answer Test (SAT) as described by Horvath Reid (1972) and Matte (1996), developing a signaling device to be activated by the examinee to indicate they have attended to the test question, and developing a method of evaluating differences between pre-stimulus and post-stimulus data.

In a SAT paradigm, the examinee is not required to voice an answer, but is instructed to simply listen to the question. By not requiring the examinee to answer, we would remove any artifact related to answering and/or preparing to answer. Some of the experiments we reviewed that found breathing inhibition associated with the OR did not involve vocalization of answers. Reactions are assumed to be a direct representation of the amount of salience the examinee

attributes to each individual stimulus as it is presented serially. This is consistent with the notion that the polygraph is clearly not a lie detector, but rather an instrument that measures physiological reaction to a stimulus. Differential salience is inferred from differential reactivity. Truthfulness and deception (or significant responses or no significant responses) are then opined from differential salience.

Some may express concerns of an examinee dissociating during a silent answer polygraph examination. Field remedies for this concern include having the examinee nod or shake their head slightly in answer to the test question. There may be a concern that motor related activity or mental preparatoration associated with head movements causes confounding reactions. To obviate this concern, we suggest providing the examinee with a signaling device used to inform the examiner they have attended to the test question. Optimally it would be something like an electronic touch device integrated into the polygraph that the examinee could activate with a finger. We would suggest having the examinee only activate the device to test questions answered in the affirmative. Most polygraph techniques call for test relevant and comparison test questions to be answered "No" and thus this procedure would potentially only minimally affect responses to non-scoring questions. Arguably the respiration channel would be more stable using this approach, possibly providing more diagnostic value because there is less noise in the signal. Another alternative might be using an eye-blink signal to alert the examiner that the examinee has attended to the test question.

Finally we suggest manufacturers develop data evaluating algorithms capable of comparing pre-stimulus and post-stimulus data. Several of the recent experiments we reviewed that support breathing inhibition being related to attention or OR use similar measurement approaches. One example might be to compare ten seconds of the post-stimulus line length with four seconds of the pre-stimulus line length. Considerations for comparison might include; percent differences, line length differences, peak amplitude differences or differences in respiratory period. Instrument software designers could use existing platforms and data measuring tools to develop software that may allow us to extract more useful data from the respiration channel.

Authors' Note

The authors are grateful to Chuck Slupski, Toby Mc Swain, Pam Shaw, Don Krapohl, Dr. Tim Weber, Dr. Charles Honts and Dr. Stuart Senter for their thoughtful reviews and comments to drafts of this paper. The views expressed in this article are solely those of the authors, and do not necessarily represent those of the Lafayette Instrument Company, the American Association of Police Polygraphists or the American Polygraph Association. The authors are grateful to the Lafayette Instrument Company, Lafayette, Indiana for their generous contribution which allowed the purchase of the software used to create the diagrams in this article. The authors and the APA grant unlimited reproduction rights to any polygraph school accredited by the AAPP or APA. Questions and comments are welcome at polygraphmark@gmail.com.

References

Adams, T., & Hunter, W.S. (1969). Modifications of skin mechanical properties by eccrine sweat gland activity. *Journal of Applied Physiology*, 26, 417-419.

Ambach, W., Stark, R., Peper, M., Vaitl, D. (2008). Separating deceptive and orienting components in a Concealed Information Test. *International Journal of Psychophysiology*, 2008 Nov;70(2):95-104.

ASTM International (2002). Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data (E 2229-02). West Conshohocken, PA: ASTM International.

Barham, R.M. & Boersma, F.J. (1975). *Orienting Response in a Selection of Cognitive Tasks*. Netherlands, Rotterdam*:* University Press.

Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.

Boucsein, W. (1992). *Electrodermal Activity*. New York*, NY,* Plenum Press.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of Psychophysiology, 2nd edition*. New York, New York: Cambridge.

Cannon, W.B. (1929). *Bodily Changes in Pain, Hunger, Fear and Rage, 2nd edition*. New York: Harper & Row.

Darrow, C. W. (1936). The galvanic skin reflex (sweating) and blood pressure as a preparatory and facilitative function. *Psychology Bulletin*, 33, 73-94.

Darrow, C.W. (1933). The functional *significance of the galvanic skin reflex and perspiration on the backs and palms of the hands. Psychological Bulletin*, 30, 712.

Darwin, C. (1872) *The expressions of the emotions in man and animals*. London: Murray.

Department of Defense Polygraph Institute (2006). *Test data analysis: DoDPI Numerical Evaluation Scoring System*. Ft. Jackson, South Carolina.

Gati, I. & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology*: General, 119, 251-263.

Graham, F. K. (1979). Distinguishing among orienting, defense and startle reflex. In H. D. Kimmel, E. H. van Olst, & J. H. Orlebeke (Eds.). *The orienting reflex in humans* (pp. 137-167). Hillsdale, NJ*:* Erlbaum.

Handler, M.D. (2006) The Utah PLC. *Polygraph*, 35(3), 139-148.

Handler, M.D. & Honts C.R. (2008). Psychophysiological mechanisms in deception detection: a theoretical overview, *Polygraph*, 36(4) 221-236.

Handler, M.D. & Honts C.R. (2009). You Can't Run, But You Can Hide: A Critical Look at the Fight or Flight Response in Psychophysiological Detection of Deception, *European Polygraph*, in press.

Harris, J.C., Horner, A., & McQuarrie, D.R. (2000). An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations. Johns Hopkins University, Applied Physics Laboratory. SSD-POR-POR-00-7272.

Harver, A. & Lorig*, T.S. (2000). Respiration. In Cacioppo, J.T., Tassinary, L.G. and Berntson, G.G. (Eds.), Handbook of Psychophysiology*. (p. 265-293). Cambridge: Cambridge University Press.

Heimer, L. (1995). *The Human Brain and Spinal Cord, 2nd edition*, New York, NY, Springer-Verlag.

Hlastala, M.P. & Berger, A.J. (1996). *Physiology of Respiration*, New York, NY, Oxford University Press.

Horvath, F.S, and Reid, J.E. (1972). The polygraph silent answer test. *Journal of Criminal Law and Criminology and Police Science*, 63(2), 285-293.

Janig, W (2006). *The Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis*. New York: Cambridge University Press.

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.

Kircher, J. C. & Raskin, D.R. (2002). Computer methods for the psychophysiological detection of deception. In Murray Kleiner (Ed.) *Handbook of Polygraph Testing*. :Academic Press: *San* Diego.

Kircher, J. C., Kristjansson, S. D., Gardner, M. K. & Webb, A. (2005). Human and computer decision-making in the psychophysiological detection of deception. University of Utah. Salt Lake City: Utah.

Kircher, J.C., & Raskin, D.C. (1999). The Computerized Polygraph System (Version 3.0) Manual. Salt Lake City, UT: Scientific Assessment Technologies.

Kleiner, M. (2002). Physiological detection of deception in psychological perspectives: a theoretical proposal. In M. Kleiner. (Eds.) *Handbook of Polygraph Testing*; San Diego, CA. Academic Press.

Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.

Levitzky, M.G. (1999). *Pulmonary Physiology, 5th edition*, New York, NY, Mc Graw Hill.

Lykken, D.T. (1974). Psychology and the lie detection industry. *American Psychologist*, 29, 725-739.

Lynn, R. (1966). Attention, Arousal and the Orienting Reaction. Oxford, UK, Pergamon Press.

Marieb, E.N. (1999). Human Anatomy & Physiology. Old Tappan, NJ: Benjamin.

Matte, J. A. (1996). *Forensic psychophysiology using the polygraph.* J.A.M. Publications: Williamsville, NY.

Nelson, R., Handler, M.D. & Krapohl, D.J. (2008). Brute-force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37(3), p. 185-215.

Ohman, A. (1979). The orienting response, attention, and learning: An information processing prospective. In H. D. Kimmel, E. H. van Olst & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 443–471). Hillsdale, NJ: Erlbaum.

Ohman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the Autonomic Nervous System; Orienting, Anticipation & Conditioning. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.). *Handbook of Psychophysiology, 2nd edition. New* York, NY: Cambridge.

Pavlov, *I.* P. (1927). *Conditioned Reflexes*. Oxford, Clarendon Press.

Rathus, S.A. (2001). *Essentials of Psychology, 6th Edition*. Fort Worth, Texas: Harcourt.

Siddle, D. A. T., Stephenson, D., & Spinks, J. A. (1983). Elicitation and habituation of the orienting response. In Siddle, pp. 109-82 (1983).

Sokolov, E. N., Spinks, J. A., Naatanen, R. & Lyytinen, H. (2002). *The Orienting Response in Information Processing*. Mahwah, New Jersey, Lawrence Erlbaum Publishers.

Sokolov. E. N. (1963). *Perception and the Conditioned Reflex.* New York, New York, MacMillan.

Stekelenburg, J.J. & Van Boxtel, A. (2001). Inhibition of pericranial muscle activity, respiration, and heart rate enhances auditory sensitivity. *Psychophysiology*, 38(2001), 629-641.

Stekelenburg, J.J. & Van Boxtel, A. (2002). Pericranial muscular, respiratory, and heart rate components of the orienting response. *Psychophysiology*, 39 (2002) 707–722.

Steller, M. (1987). Psychophysiologische Aussagebeurteilung - Wissenschaftliche Grundlagen und Anwendungsmöglichkeiten der "Lügendetektion". Göttingen: Hogrefe.

Timm, H. W. (1982a). Analyzing deception from respiration patterns. *Journal of Police Science & Administration*, 10, 47-51.

Timm, H. W. (1982b). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology*, 67, 391-400.

Tortora G.J. & Grabowski, S.R. (1993). *Principles of Anatomy and Physiology, 7th edition*. New York, NY Harper Collins.

Turpin, G., Schaefer, F., & Boucsein, W. (1999). Effects of stimulus intensity, risetime, and duration on autonomic and behavioral responding: Implications for the differentiation of orienting, startle, and defense responses. *Psychophysiology*, 36, 453–463.

Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. (2004). Autonomic and behavioural responding to concealed information: differentiating defensive and orienting responses. *Psychophysiology*, 41, 461-466.

Some Thoughts about Feelings: A Study of the Role of Cognition and Emotion in Polygraph Testing

Mark Handler, Pam Shaw and Michael Gougler

Reprinted with permission of the American Polygraph Association
This article originally appeared in the APA Publication, Polygraph, 39 (3)

When I was a child, I spoke and thought and reasoned as a child. But when I grew up, I put away childish things.

1 Corinthians 13:11 New Living Translation (2007)

Abstract

This manuscript is offered as a follow up to the work of Khan, Nelson and Handler (2009) that discussed emotion in psychophysiological detection of deception (PDD) testing. Our intent is to offer the interested reader a more in-depth discussion of the "cognitive appraisal theories" of emotion (Scherer, 2001) in hope of generating thought, debate and research. Our work here focuses on emotion, with the goal of expanding our profession's knowledge, vocabulary and appreciation of this complex concept. We suggest that emotional states, along with cognition and behavioral learning, work synergistically to create or produce measurable responses to stimulus questions during PDD testing. We further suggest that our subjects appraise or evaluate the test questions against some type of goal that is at stake, and that these appraisals serve a mediating function for valence and salience of emotional and physiological response. We will offer descriptions of how these appraisals intertwine and connect in the moment, or result from a reinstatement of a previous evaluated conflict. Our paper scaffolds on the Khan et al. (2009) work by integrating the writings of many modern emotional researchers whose works are listed in the reference section.

Describing versus *defining* Emotion

We find ourselves at a loss for locating a universally accepted definition of the word *emotion*. One reason for this is that definitions are generally used to describe tangible things or processes that can be observed. They do not include the need to describe systems before being able to describe what those systems do, as in the case of emotion. We will provide a basic concept of the constituent parts of an emotion, and then attempt to construct an explanation of emotion from that foundation. This description will be an amalgamation of the writings of several psychologists in the area of emotion research (Averil, 1994; Barlow, 2002; Bradley & Lang, 2000; Clore & Ortony, 2000; Coleman, 2001; Damasio, 1999 & 2000; Gray, 1994; Lane, Nadel, Allen & Kaszniak, 2000; Lazarus, 1994, 1991; Le Doux, 1994, 1996, 2000; Power & Dalgleish, 2008; Scherer, 2000; Scherer, Schorr, & Johnstone, 2001). We make no claim as to have encompassed all of the current offerings of what constitutes emotion, which would require a monumental effort. Rather, we attempted to include those aspects of modern writers that seem to aptly describe the interactions that are *most likely* observed during PDD testing. As with most

discursive attempts to account for the inner workings of the mind and subjective experience, we place an emphasis on cognitive contributions to reason to our conclusions.

Different emotional models / theories

We begin with an overview of many of the existing theories of emotion in current literature. Since there are many ways in which one could study what role emotions may play during PDD testing, we would be remiss to attempt such a study or discussion without consideration for the breadth and depth of existing work in the broad field of psychology. Scherer (2000) lists the following as some of the current psychological models of emotion. *Dimensional emotion models* place their major focus on subjective feelings and categorize emotions based on such concepts as valence (positive or negative value) and arousal. *Uni-dimensional models* (Duffy, 1941; Watson, Clark & Tellegan, 1988) stress the idea that one dimension is sufficient to distinguish emotional states. *Multidimensional models* (Plutchik, 1962; Russell, 1980) stress two-dimensional models assessing valence and activation. *Discrete emotion models* (Gray, 1990; Panskepp, 1982) include circuit models and basic emotional theories (Plutchik, 1980). *Circuit models* (Gray, 1990; Panskepp, 1982) approach understanding emotion by attempting to elucidate the neural circuits underlying the responses. *Basic emotional theories* suggest there are a limited number of core emotions that have developed over the course of human evolution. Each of these basic emotions has an associated antecedent or eliciting set of conditions and each has a specific response pattern. *Meaning oriented models* include lexical (Ortony, Clore & Collins, 1988) and social constructivist models (Averill, 1980; Harre, 1986; Shweder, 1993). *Lexical approaches* (Ortony, Clore & Collins, 1988) attempt to understand emotions using semantics to label emotional states. *Social constructivism models* (Averill, 1980; Harre, 1986; Shweder, 1993) claim socioculturally determined behavior, expectations and values impart the meaning that in turn generates emotional states.

Finally, we come to *componential models* (Ellsworth, 1991; Frijda, 1986; Lazarus, 1991; Roseman, 1984; Scherer, 1982) or *integrative models* (Barlow, 2002) that attempt to study emotion as a system of integrated components. These theories recognize that the complexity of emotions and emotional theories cannot be explained or understood by isolating them to a single field of study. A number of modern theorists now approach the study of emotion from an integrated perspective, one that includes affect, behavior, neurobiology and cognition (Barlow, 2002; Ellsworth, 1991; Frijda, 1986; Lazarus, 1991; Roseman, 1984; Scherer, 1982).

Emotional Syndromes, States and Reactions

We found the foundation Averill (1994a) provided for understanding the concept of emotion helpful, where the terms emotional syndromes, emotional states and emotional reactions all apply to emotion but in different ways. Having a definition of each will provide a common framework for our discussions.

Emotional *syndromes* are the hypothetical concept of what it means to experience an emotion, and the term *syndrome* refers to an expected pattern of co-occurring signs or symptoms that may indicate a common origin (Coleman, 2001). Emotional syndromes are what we think of when we "picture" an emotional state.

Emotional *states* are episodic experiences in the form of a short term disposition to respond in a manner consistent with the expectations of the equivalent emotional syndrome. In other words, emotional states are the condition we find ourselves in while experiencing an emotion and occur in response to a stimulus event. We will use the terms emotion and emotional states interchangeably throughout this paper.

Emotional *reactions* are responses to the emotional state. Emotional responses are characterized by the presence of four major components: a cognitive component, an affective component, a biological component and a behavioral component (Barlow, 2002; Bradley & Lang, 2000; Damasio, 1999 & 2000; Lazarus, 1991; Power & Dalgleish, 2008; Scherer, 2000). The cognitive component accounts for the conscious or unconscious perception and appraisal of the stimulus in terms of emotional significance or meaning to the subject. The affective component provides the subjective experience or feelings associated with a particular emotion, which humans and other animals have historically used to increase learning and enhance survival (Buck, 2000). The biological component includes the bodily effects resulting from activation of the autonomic and central nervous system. Finally, the behavioral component provides the impetus to engage in action or behavior, and is often the most useful point of observation when we seek to understand a person's motivation and goals. Emotional reactions result from a complex, integrated pattern of central and autonomic nervous system functions which manifest themselves in physiological response patterns and are intended to create circumstances that are advantageous to the organism. They are a product of evolution that have their biological origin in the brain, but use the body as their stage.

Emotionality in the polygraph context

E*motionality* is the measurable aspects of behavior resulting from emotion (Reber, 1995). During PDD testing we attempt to differentiate truthfulness from deception by assessing reactions to test questions and then making inferences about the salience of the questions from those reactions. Some of the changes we consider in PDD testing are likely the result of emotional states, which are also dependent on motivation, experience, memory, and cognition. Some of the physical manifestations of emotionality, as observed during PDD testing, include changes in respiratory, cardiovascular, vasomotor, and electrodermal activity. PDD testing theories hold that observed emotionality associated with the test question will contribute to the physiological reactions that can be measured and interpreted. We should be reminded at this point that we will never know, nor can we assume to know, precisely what emotion or emotions our examinees experience during PDD testing. Individual emotions are semi-predictable events that are assumed to be the direct result of the PDD test questions, but may vary depending on

who is being tested (personality), why they are being tested (pre-employment, sex offender, event-specific criminal test), question type (relevant, probable-lie comparison, directed-lie comparison, or neutral question), or veracity status (truthful or deceptive). Though currently we cannot identify the particular emotions measured during a PDD exam, there is no dearth of research to support PDD's ability to separate truthfulness from deception well above chance and quite reliably (National Research Council, 2003).

Purpose of Emotions

Emotional states are the result of evolutionary fine-tuning that is intended to ensure the survival of an organism (Smith & Kirby, 2001). This is accomplished by preparing and motivating the individual to contend with goal relevant stimuli, like PDD test questions. First, emotions serve to produce responses that enhance survivability of an encounter (Damasio, 1999; Ekman & Davidson, 1994a & 1994b; Lazarus, 1991). Emotions can be seen as mechanisms that regulate behavior in relation to patterns laid down through evolution.

Cannon (1927) described fear reactions as an overall sympathetic nervous system (SNS) arousal resulting behaviorally in what he called *fight-or-flight*. When presented with an emergency situation, Cannon felt an animal can choose to fight the danger or attempt to flee. Fighting and running away both involve an initiation of movement, where immobility is just the opposite. Alternatively, Gray (1988) introduced the term Behavioral Inhibition System (BIS) to describe a series of responses to fear stimuli that include increases in arousal, behavioral inhibition, and increases in attention. The *freeze* response became an integral part of Gray's early BIS hypothesis and described an inhibition of ongoing behavior. Updated descriptions of the BIS by Gray and McNaughton (2003) discussed behavioral inhibition as decreased motor activity when presented with fear or anxiety associated with an approach-avoidance based conflict. The updated theory separated pure "freeze" reactions, which were typically associated with the fight or flight response, from those that were behaviorally inhibited. This introduced the concept of higher brain functions being able to override programmed behavior. Such adaptive capabilities would serve to expand a response repertoire, thus increasing the chance for survival.

Gray and McNaughton (2003) noted that the freeze response and behavioral inhibition were physiologically so similar that they were very difficult to differentiate, especially in humans. The difference between the two in the causation of arousal was conceptualized as a difference between freezing proper and a defensive quiescence, or quieting. The freeze response (freeze proper) occurred when an animal was placed in the immediate proximity of a highly fearful stimulus and was followed immediately by a fight or flight reaction. Additionally, this freezing proper was insensitive to anxiolytic drugs and could be triggered by learned or innate stimuli. They noted that the freeze proper response functions in a manner similar to the startle response; it quickly disconnects the animal from ongoing behavior allowing it to attend to more important and immediate stimuli. Behavior inhibition, on the other hand, involved attention with

an emphasis on evaluation in the form of a *stop-look-and-listen* action. Behavior inhibition is said to result from anxiety and was *inhibited* by anxiolytics (Gray & McNaughton, 2003).

An additional biological function of emotions is to prepare the organism for a reaction, often in the form of a physical action, even though a reaction may not be needed and may not occur. Emotions, however, allow a head start towards a reaction, where a number of physiological changes will occur in anticipation of a potential negative encounter. This feed-forward type of physiological preparation is referred to as *allostasis* (Berntson, & Cacioppo, 2007; Handler, Rovner, & Nelson, 2008; Schulkin, 2003; Sterling, 2004; Sterling & Eyer, 1988). Allostasis can be described as a central nervous system mediated, integrated brain-body response geared towards viability or survival. It occurs in regulatory systems which have no fixed set point and all are the result of evolutionary tinkering. The evolutionary benefits of adopting a "Why wait?" response seems obvious.

Discrete versus Component-Process Theories of Emotion

Emotion literature is replete with arguments for and against the idea that there are "basic" or discrete categories of emotional syndromes. It would be an understatement to write that many well respected researchers have conflicting opinions about the existence of a prototypical core of emotion states and reactions (Ekman & Davidson, 1994a). Add to that the fact that there are conflicting definitions of what constitutes a *basic* emotion (Averill, 1994b), and we can begin to appreciate the ambiguity surrounding this concept. A number of researchers have agreed that in order for an emotion to be considered "basic," it must be able to be distinguished by distinct universal components, such as facial expressions, or have distinct physiological components (Power & Dalgleish, 2008). Power and Dalgleish (2008) made a cognitive case for distinguishing among emotions using the perceived appraisal as that which gives an emotion its distinctiveness. A number of discrete emotion theories suggest that these basic emotions can mix or blend to produce a variety of emotional states (Scherer, 2001). This approach, however, is rooted only in theory for the time being because it would involve testing mental states (appraisals) as opposed to physiologic arousal.

Khan et al. (2009) used the concept of basic emotion theory as a starting point for understanding emotional contributions to PDD testing, as there seems to be sufficient face validity for it to be considered plausible. They pointed out that among the diversity of lists of basic emotions, investigators have found evidence for six; fear, anger, happiness, sadness, disgust/contempt, and surprise. Several researchers point out that surprise does not always result in an emotion, and have dropped it from the list. They link surprise to the startle reflex and point out it is not indisputably an emotion state. Khan et al. (2009) and Power and Dalgleish (2008) listed those emotions that are listed in Table 1 along with their accompanying appraisals.

Basic Emotion	Appraisal
Anger	Frustration or perceived blocking of a role or a goal, directed at the perceived thwarting agent
Fear/Anxiety	Physical or social threat to self, goal or ego-type
Disgust	Something repulsive to oneself or society.
Sadness	Actual or potential loss or failure of a valued role (ego-type) or goal.
Happiness	Positive move towards a valued goal (ego-type) or role.

Table 1 The key appraisals for each of the five basic emotions, adapted from Power and Dalgleish (2008).

We want to take a moment and discuss fear and anxiety as they may relate to PDD testing. When discussing anxiety, we refer to state-anxiety, or that anxiety which occurs in a short-term phasic manner. This is not to be confused with trait anxiety which is a relatively stable characteristic of anxiety in a person. Fear and anxiety are sometimes used synonymously though their relationship has often been debated (Barlow, 2002). Many times it requires a set of semantics or a definition to clarify or settle any debate. The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV TR) is published by the American Psychiatric Association and provides diagnostic criteria for mental disorders. The current DSM-IV states the term anxiety denotes "apprehension, tension or uneasiness that stems from the anticipation of danger...". Ohman (1993) called anxiety "a state of undirected arousal following the perception of a threat," which seems to fit some of what we may observe during PDD testing. Epstein (1972) suggested fear relates to action potentials such as fight-or-flight while anxiety occurs when those action potentials are thwarted or prevented. Power and Dalgleish (2008) provide another description of anxiety that fits the PDD testing milieu. They propose that anxiety

> "…is conceptualized as a state in which the individual is unable to instigate a clear pattern of behavior to remove or alter the event/object/interpretation that is threatening an existing goal."

Again, from a PDD testing point of view this provides a plausible explanation for a portion of the emotionality we may be observing.

Contrary to a discrete emotion theory, the component process model (Scherer, 1984b) does not assume there are a set of hard-wired programs that mix or blend to produce a variety of emotional states. Rather, the component process model suggests that continuously changing patterns within the emotional components results in a large number of different emotions. Scherer (1984b) refers to these as "modal" emotions and writes they are potentially infinite in the possible combination of resulting emotions, though acknowledges there is some "bunching" of these elements around the "basic" emotions.

Ekman (1994) offers a possible middle ground for this argument under the concept of "emotion families". Each emotion is not itself a single affective state, but belongs to a family of states. These families share a number of characteristics that vary to produce individual differences based on the appraised circumstances. Any resulting emotion can be a variation of a theme, reflecting learning. Multiple variations of multiple themes can occur simultaneously.

The advantage of consciousness for our emotional reactions

Consciousness of emotions can be seen as adaptive responses to allow organisms a greater opportunity to succeed in life. Humans are gifted with extended consciousness (Damasio, 1999), which describes their ability to incorporate past memory and future planning into a current appraisal. This also allows humans to form and execute better adaptive strategies in the face of a perceived challenge to a goal. These strategies may include a response in preparation for a potentially aversive event, as well as inhibiting a prepotent response. Having an expanded repertoire of responses would seem to endow an organism with an increased chance of survival.

We are proposing that emotion results from an evaluative process, called an appraisal, and are always therefore about something. Feelings that are the conscious perception of the affective aspect of emotions, seems to have evolved to increase the value of the emotion. Being able to remember a prior emotional encounter serves to allow the organism to respond more quickly the next time it is in a similar circumstance (Damasio, 1999; Power & Dalgleish, 2008). In order to reap this benefit, that organism is best served by being endowed with an extended conscious.

Cognition

Cognitive activity broadly describes any conscious or unconscious thought process (Lazarus, 1991) that results in the appraisal of and response to a stimulus. For the purposes of this paper, the term *cognition* will include the route by which a stimulus, or test question, is perceived, attended to, processed, evaluated, compared to memory, encoded into memory, etc. It will also include the performance of the mental computations commonly referred to as *thinking*. Recent neuropsychological models have supported the notion that, at least in humans, emotional and cognitive functions are strongly reciprocally connected (Gainotti, 2000). We will make no attempt to discuss emotion outside of this relationship, lest we risk a semantics struggling match.

The Lazarus (1991) description of emotion from a *cognitive-motivational-relational* perspective applies easily to the PDD setting. Cognition can mediate a shift in attention from what was occurring to something new and can include planning and coping mechanisms. Motivation, which is essentially what we want, determines the power of the particular emotion as the encounter is appraised in terms of goals. Motivation can also be influenced by the valence (positive or negative value) of an emotion. The relational aspect of the Lazarus (1991) theory accounts for the idea that all emotion is a response to a person-environment relationship appraisal. The term environment here is used to describe anything the person may interact with that can result in an emotion. Lazarus (1991) discussed relational terms as either positive or negative emotion generating, depending on how the relation is appraised with regard to goal relevance, goal congruence or incongruence, and type of ego-involvement (Lazarus, 1991). He also offered the idea of secondary appraisals which are related to coping potential, agency (who is to blame) and future expectations. For example, if a person assesses the person-environment relationship to be harmful, goal incongruent, or have negative adaptational consequences, a negative emotion would likely be elicited with accompanying negative affective qualities. Conversely, an appraisal of a goal congruent relationship would foster an emotion that has a positive valence.

The appraisal process is not a one-shot circumstance where a stimulus is evaluated with regard to goal relevance, implication and coping potential, and a single response generated. Lazarus (1991) and Scherer (2001) point out the initial appraisal is followed up by multiple iterations of reappraisals. These reappraisals serve to update the organism on changes in the circumstances so that any appropriate adjustments, up or down, may be made. A series of "evaluation checks" are occurring in an ongoing manner as the signal terminates through extinction or is supplanted by a more salient stimulus (Scherer, 2001).

Cognition in the polygraph context

During PDD testing, examinees are presented with a number of stimuli, in the form of test questions, and are essentially asked to attend to each sequentially. Presumably, as the examinee attends to each test question he or she conducts an appraisal with respect to what that test question means. This appraisal relates to the examinee's goals, standards, and attitudes and how those may be affected within the PDD setting. Cognition and appraisal are a process of evaluating a stimulus for goal congruence within the examinee's motivational framework. While it is not feasible to attempt to state we know what particular meaning a particular examinee attributes to an individual test question, it is possible to discuss a number of possibilities of what the examinee *could be* thinking in terms of goal congruence. Appraisals are simply an evaluation that are assigned emotional meaning, value or salience (Barlow, 2002; Bradley & Lang, 2000; Clore & Ortony, 2000; Damasio, 1999; Gray, 1994; Lazarus, 1991; Lazarus, 1994; Le Doux, 1994; Le Doux, 1996; Le Doux, 2000; Power & Dalgleish, 2008; Scherer, 2000).

How do appraisals result in emotional states?

We have suggested that emotions are the response to appraisals of significance in a given situation with respect to goals. We offer that there are two routes of appraisal through which emotions may occur, and both are applicable to PDD testing. Both routes of appraisal involve a cognitive component and are equally capable of eliciting an emotion. One is a conceptual, computational or schematic route and the other is a reinstatement of a previously learned or evaluated situation (Power & Dalgleish, 2008). The former will be developed or computed through a situational analysis. The latter relies on memory of an earlier response and produces a faster, though potentially less accurate response. In either case, a situation that is appraised as having significance for a person's goals can elicit an emotional reaction either as a result of a reinstated prior emotion or because the person has perceived the situation to be one that will affect their goals.

A PDD related example of a conceptual, computation or schematic route for generating an emotional state.

This route of appraisal describes one that is essentially pieced together in a conceptual or story-like manner. In this case the appraisal is being conducted as the pieces of information become available. For example, take an examinee in a public safety pre-employment screening polygraph test that has been less than forthcoming about his past criminal activities. During the pre-test discussions of these issues he or she silently compares their personal involvement in criminal activities against what they believe are societal norms or what the hiring agency will accept. He or she concludes that telling the complete truth about what they have done may be incompatible with the hiring preferences of the agency to which he or she has applied. They may believe that in order to continue in the hiring process they must lie about these acts or minimize their admissions. This deception may then result in the activation of one or more emotional responses, all of which are in response to an appraisal. Perhaps the examinee is angry with himself or herself for having done these things, considering them stupid. Alternatively, they could be angry with the hiring agency for inquiring into what he or she feels is a private matter, or one that may have happened long ago and should no longer be relevant. He or she may feel some level of guilt for what they have done or possibly experience some degree of shame or embarrassment at the prospect of the polygraph examiner and hiring agency discovering this issue. There may also be some anxiety, or even fear, surrounding the idea of not getting the job or being labeled as someone who is not qualified for the job, thus ending their law-enforcement career. Some of these emotions could have occurred because, during the appraisal process, the examinee became concerned that past acts are incongruent with the goal of obtaining the job. Other emotions could result from the examinee being reminded of past transgressions which are socially objectionable. This is just one possible example of the multitude of ways the examinee could use a bottom-up or constructive approach to generate the emotional states.

A PDD related example of an associative route for generating an emotion state.

Imagine you are at the dentist having a cavity filled and the anesthesia is ineffective at masking the pain of the drill. As the dentist drills into your molar you experience a sharp pain coinciding with the sound of the drill. You hope your reaction to the pain will cause the dentist to stop and remedy the situation. But what about the next time you hear the sound of the dentist drill? It is possible that the sound of the drill would produce not only a cognitive response in the form of a memory, but may also result in an associative emotional reaction? This is an example of reinstatement of memory from an earlier-formed evaluation which generates emotions "as if" an appraisal is occurring. The appraisal and the emotion should not be confused for being the same thing. The appraisal work has already been done and the memory of the appraisal has been stored for this stimulus, allowing the emotion to more quickly and more reflexively occur. One need not stretch his or her imagination to appreciate the evolutionary benefits of such ability. Long term survival would seem more likely in an organism that does not have to perform a complete appraisal before generating an emotion and action in response to a threatening event. One that can activate responses because of a memory of a similarly appraised encounter can act faster and perhaps respond more effectively. In its extreme form, this feature of physiology results in pathological formations. In the case of Post Traumatic Stress Disorder, a person's reflexive cognitive, emotional, and physiological responses do not contribute to survival and effective response to the present environment, but rather, actually interferes with effective functioning.

Another example that offers a potential for a reinstatement of an emotion is the negative social connotations associated with lying. It is important to recognize that lying is both goal-directed and a common behavior, intended to reduce anxiety or threat associated with the truth about information for which a lie is told. This may occur in part because people are social creatures who often tend to seek approval and acceptance of their fellow humans, though they sometimes lie to achieve these goals. Most children are socialized from an early age to equate honesty with honor and goodness, that dishonesty is frowned upon, and that lying brings about punishment. We recognize that sometimes lying can also bring about reward when the deceptive behavior is not confronted. The decision to lie rests in the appraisal of whether lying will produce less internal anxiety or external consequences than would telling the truth. While lying is almost universally disapproved of, children are also socialized to understand the subtle boundaries surrounding verbal discretion, social politeness, and obsequiousness. In most societies lying in formal settings such as in discussions with a person in a position of authority is strongly discouraged, and in some cases, such lying is punished severely when it is discovered. For example, lying to a federal law enforcement officer during the course of an investigation is a felony in itself. It would seem there is a potential for anxiety to be associated with openly breaching such societal rules. There is also the potential for positive and conflicted emotions as the person hopes and seeks to obtain a desired result through telling a lie.

The act of having lied, for some people, may cause the test questions to function as a form of conditioned stimuli. Test questions pertaining to the act about which the examinee lied can produce a learned or associated internal anxiety state. This anxiety state is a consequence of a lifetime of conditioning experience resulting from accepting and rehearsing a system of socialized values that emphasize goodness and honesty. The possibility of getting caught in a lie and/or the punishment associated with being caught can generate a negative emotional state. Thus, even in a laboratory setting (where there is little jeopardy) the act of lying may create sufficient emotionality or conflicted response to produce measurable physiological reactions. Similarly, conditioned responses and any associated emotional states stemming from the behavioral act itself, independent of the act of lying about the event, may also play an additive role in the development of observable and measurable polygraph reactions, along with related neurobiological activity and mental effort.

Cognitive processes surrounding knowledge and memory of having engaged in an act can increase the salience of a test question about that act. Pretest discussion and review of the test question is thought to increase the salience of the test question for a person involved in the event, by stimulating thoughts, memory, and emotional experience pertaining to the event. Persons uninvolved in the event described by the test question have no associated memories, thoughts or emotional experience regarding the details of the incident. The memory tasks involved in lying can require additional mental effort or increased cognitive load. The subject must attempt to suppress a memory or thought and divert their attention to another matter when presented with the test stimulus question. Liars need to create their lie, assess that lie with regard to plausibility or believability, keep the lie straight during possibly numerous retellings and not confuse the lie with the truth. Liars also need to keep the lie separate from the truth and they need to monitor themselves more carefully in order to ensure they appear truthful and avoid giving away the falsehoods. In addition to the need to marshal sufficient mental ability to manage the content complexity and tell the lie in a convincing and coherent manner, liars must also try to conceal any emotional reaction which may occur in response to the either the event or the act of lying. This content complexity (Vrij, 2008) can add both emotional stress and cognitive demands to task requirements for dishonesty compared to similar requirements for truth telling.

A truthful person, uninvolved the event described by a polygraph question, is free of the burden of conditioned emotional responses to the act of lying and to the event described by the relevant questions. This person is also free of the complex demands on attention and cognitive systems, including any need to manage presentation or appearance while maintaining a separation of the truth from the development and presentation of a plausible alternative. The truthful examinee may devote attention and effort to assess the likelihood that the test will result in an error, and the potential consequences associated with an error. However, our position is that the emotional and cognitive demands relevant questions place on the truthful person are less than those required of someone who is involved in and chooses to lie about an event under investigation. The effectiveness of PDD stimuli would seem to be contingent upon whether there

is reference to both an event in question and the examinee's involvement in that event. For example, someone being investigated for a bank robbery might be asked, "Did you rob that bank?" This manner of questioning would more directly associate the examinee with the act of concern than would an indirect approach involving question about lying, in person or in writing, regarding the event in question (e.g., Were you truthful in your written statement about not robbing the bank?). We know from conditioning studies that the closer a stimulus is to the conditioned target stimulus, the larger the reaction (Kehoe & Macrae, 2002).

Emotion-Specific Physiology

The issue of whether there are distinct physiological measures specific to "an" emotion has been contemplated and investigated for some time. William James (1890) felt emotions were the result specific changes in skeletal muscle and other physiological changes that were read out to create each emotion. The notion that peripheral physiological measurements could parse out individual emotions began to be replaced by the Schacter and Singer (1962) two-factor emotional theory. Their theory emphasized that the cognitive factors were the main determinant of the specific emotion which resulted in a general state of arousal. Advances in instrumentation may have contributed to a renewed interest in psychophysiological differentiation among discrete emotions in recent years (Davidson, 1994).

Several experiments have attempted to find differences in autonomic nervous system (ANS) arousals. Sinha, Lovallo and Parsons (1992) found systemic differences among some emotions which have negative valence. Anger resulted in greater diastolic blood pressure and increased peripheral resistance when compared to fear. Levenson, Ekman and Friesen (1990) compared anger and fear using finger temperature and reported an increase in temperature for anger and a decrease for fear. Cacioppo et al. (1993) conducted an extensive review of studies comparing two or more emotions measured by two or more ANS responses and reported finding little consistency. Cacioppo, Petty, Losch and Kim (1986) reported increased electromyographic activation of corrugator muscles during negative affect stimuli and great activation of zygomatic activity with positive affect stimuli. Levenson, Ekman and Friesen (1990) reported the finding of four reliable differences among the negative affect emotions of fear, anger, sadness and disgust. They found: (a) anger produced a greater increase in heart rate when compared to disgust; (b) anger produced a greater increase in finger temperature when compared to fear; (c) fear produced a greater increase in heart rate when compared to disgust and (d) sadness produced a greater increase in heart rate when compared to disgust.

As Khan et al. (2009) point out, these findings suggest a weak support for an ability to differentiate specific emotions using ANS measurements (Davidson & Ekman, 1994; Levenson, 1994a & 1994b). From a PDD standpoint, this may seem like gloomy news if we were to be claiming to be able to pinpoint "fear" from among the many other potential emotional states an examinee may experience. Fortunately, a fear alone driven theory of PDD testing, is untenable. We suggest we do not know, and could not know, what specific emotion or emotions may be

contributing to ANS changes we measure during PDD in any particular individual. Instead, we are content to admit that whatever contribution emotion makes to changes in our measurements, it is sufficient to allow us to effectively differentiate truthfulness from deception.

A review of relevant, probable lie comparison (PLC) and directed lie comparison (DLC) questions in PDD testing.

In investigative polygraph testing, relevant question targets are dictated by the circumstances of the investigation and are commonly formulated around the most salient or intense aspects of the allegation. In screening programs, relevant questions should describe the test subject's involvement in possible behavioral concerns to risk managers or adjudicators and should be designed to add incremental validity to their particular program. Effectively formulated relevant questions will directly assess the test subject's behavioral involvement in the issue of concern.

PLC questions are presented to the test subject as being necessary for further evaluating the test subject's character and the issue under investigation. PLC questions are based on transgressions whose subject matter is generally or conceptually related to the allegations of the examination and which virtually all persons may have committed, but will likely be denied in the context of the examination. PLC questions are broad in scope and the test subject is strongly, but indirectly, discouraged from making admissions to PLC questions. If the test subject makes an admission to a PLC question, the examiner typically addresses that admission with some dismay, minimizes the admission, or modifies the comparison question accordingly. The ultimate goal is to discourage admissions to PLC questions to ensure that the test subject perceives them as ambiguous and broad in nature. It is also important the examiner imply to the test subject that lying to any of the questions (relevant or PLC questions) will result in a failure of the polygraph test and resulting conclusion of deception to their involvement in the relevant issue under investigation.

DLC questions are those which the examiner instructs the examinee to answer falsely (Honts & Raskin, 1988; Raskin & Honts, 2002). DLC questions are presented to the examinee as necessary to ensure they maintain an ability to respond appropriately when lying. Examinees are told that if they fail to respond appropriately to the DLCs, the test result will be inconclusive. The rationale being that most truthful subjects will view an inconclusive result negatively.

DLC questions may offer some relief to potential problems identified in PLC versions of polygraph testing. Examiners may experience difficulty in standardizing comparison questions in the PLC version. Each test subject brings his or her own life experiences and idiosyncrasies that may sacrifice rapport while attempting to lay a foundation for the PLC questions. Additionally, test subjects who have prior polygraph experience or those who have researched polygraph techniques may not be naïve to the PLC principles. This sophistication could make laying the foundation for the comparison questions challenging. Non-naïve test subjects may

acquiesce to the procedure in order to not seem obstreperous, in which case the PLC questions become similar to DLC questions.

One theory behind the DLC approach is that it is similar to the PLC approach in that it is assumed the subject's cognitive and emotional attention will be focused more on the questions that pose the greatest concern of not passing the test, which should be a goal of most test subjects. Thus, the truthful subject will be more concerned with whether or not they are a suitable subject and whether or not they are producing appropriate responses to the DLC questions when they are lying. Though they have permission to lie on these questions, the questions still serve to draw the examinee's attention during testing to the questions that pose the greatest challenge to achieving their goal. The theory further proposes the guilty subject will remain primarily concerned about the relevant questions on the test and will consequently produce the greatest reactions to them. In this sense, the DLC questions operate as a distracter item for the truthful subjects, who are more capable of being distracted away from the relevant questions than are the deceptive subjects.

Conclusions

We have attempted to build on the work of Khan et al. (2009) for the reader interested in a deeper understanding of the concepts they offered on cognition in emotion. We chose to focus on emotion and sub-focus on a cognitive appraisal structure of emotion in hope of expanding our professions knowledge, vocabulary and appreciation of this complex concept. Our hope is that others will continue these investigations, focusing more on cognition and its inter-relationship with motivation and behavior. We proposed that emotional states, along with cognition and behavioral learning, form a constellation of response potential that produce measurable responses to stimulus questions during PDD testing which may be used to accurately and reliably categorize subjects as truthful or deceptive. We further suggested that the test questions in the PDD setting are perceived by the examinee and cognitively appraised with regard to goals, standards and attitudes, as well as coping potential, agency and future considerations. These appraisals serve a mediating function for salience of emotional states and their concomitant physiological response. Following the suggestions of Khan et al. (2009) we feel the appraisals can be generated in a schematic manner or via a reinstatement of a previous evaluated conflict and we offered a parallel between each mode of generation in a hypothetical examination setting.

We feel it is important to reinforce at this juncture the potential complications involved in PDD testing and how those complications may work to the detriment of specificity to deception. Many things can result in arousals that are not specific to deception. By considering the ways that reactions can be caused during PDD testing we may be in a position to strengthen specificity. PDD testing depends on the evaluation of the examinee's physiological reactions to test questions, and then works backwards to make inferences about whether the examinee's reactions correspond to expected responses from truthful or deceptive persons. The success of these methods depends, in part, on the assumption that the primary thing which differentiates

response magnitudes to various stimuli is the degree of salience which the examinee assigns the test question. The specificity of that salience is revealed through the degree of physiological reactions. Differential reactivity is presumed to occur in response to the examinee's cognitive appraisal, memory, behavioral experience, and emotional response regarding the test stimulus questions. The complexity of these physiological systems and assumptions will inevitably prompt discussion about the potential for error in the PDD context. Our ultimate confidence in the viability of PDD testing as a method for credibility assessment and the challenge of differentiating truthful from false statements is ultimately dependent on empirical evidence, and the substantial body of field and laboratory research that supports the validity of PDD testing (Honts, 2004; Raskin & Honts, 2002).

References

Averill, J. R. (1980). *A constructivist view of emotion*. In R. Plutchik and H. Kellerman (Eds.), Emotion: Theory, research and experience: Vol. I. Theories of emotion (pp. 305-339). New York: Academic Press. Reprinted in: Halberstadt, A. G., & Ellyson, S. L. (Eds.) (1990). Social psychology readings: A century of research (pp. 143-156). New York: McGraw-Hill.

Averill, J. R. (1994a). *Emotions Unbecoming and Becoming*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 265-269). New York: Oxford University Press.

Averill, J. R. (1994b). *In the eyes of the beholder*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 7-14). New York: Oxford University Press.

Barlow, D.H. (2002). *Anxiety and Its Disorders*, 2nd edition, New York, NY, Guilford Press.

Berntson, G. G., and & Cacioppo, J.T. (2007). *Integrative Physiology: Homeostasis, Allostasis, and the Orchestration of Systemic Physiology* In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), Handbook of Psychophysiology, 3rd edition (pp. 433-449). New York, NY: Cambridge University Press.

Bradley, M. M., and Lang, P.J. (2000). *Measuring Emotion: Behavior, Feeling and Physiology* . In R. D. Lane & Nadel L (Eds.), Cognitive Neuroscience of Emotion (pp. 242-276). New York, NY: Oxford University Press.

Buck, R.W. (2000). *The epistemology of reason and affect*, In J.C. Borod (Eds.) The Neuropsychology of Emotion, (pp. 31-55). New York, NY, Oxford University Press.

Cacioppo, J.T., Klein, D.J., Berntson, G.G., and Hatfield, E. (1993). The psychophysiology of emotion. In M. Lewis & J. Haviland (Eds.), Handbook of Emotions (pp. 119-142). New York: Guilford.

Cacioppo, J.T., Petty, R.E., Losch, M.L. and Kim, H.S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50, 260-268.

Cannon, W.B. (1927). *Bodily changes in pain, hunger, fear and rage*. New York: Appleton-Century-Crofts.

Clore, G. L., and Ortony, A. (2000). *Cognition in Emotion: Always, Sometimes, or Never* . In R.D. Lane and L. Nadel (Ed.), Cognitive Neuroscience of Emotion (pp. 24-61). New York, NY: Oxford University Press.

Coleman, A.M. (2001). *Oxford Dictionary of Psychology*. New York, NY, Oxford University Press.

Damasio, A. R. (1999). *The Feeling of What Happens*. New York, NY: Harcourt Brace and Co.

Damasio, A. R. (2000). *A Second Chance for Emotion*. In R D Lane & L Nadel (Eds.), Cognitive Neuroscience of Emotion (pp. 11-23). New York, NY: Oxford University Press.

Davidson, R. J. and Ekman, P. (1994). *Afterword: Is there Emotion-Specific Physiology?* In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 261-262). New York, NY: Oxford University Press.

Davidson, R.J. (1994). *Complexities in the Search for Emotion-Specific Physiology*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 237-242). New York: Oxford University Press.

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion." *Journal of General Psychology*, 25, 283-293.

Ekman, P. (1994). *All emotions are basic*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion, (pp. 45-47). New York: Oxford University Press.

Ekman, P. and Davidson, R.J. (1994a). *Afterword: Are there basic emotions?* In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion, (pp. 45-47). New York: Oxford University Press.

Ekman, P. and Davidson, R.J. (1994b). *Afterword: What is the function of emotions?* In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion, (pp. 137-139). New York: Oxford University Press.

Ellsworth, P.C. (1991). *Some implications of cognitive appraisal theories of emotion*. In K.T. Strongman (Ed.), International review of studies of emotion, Vol. 1 (pp. 143-161). New York, NY : Wiley.

Epstein, S. (1972). *The nature of anxiety with emphasis upon its relationship to expectancy*. In C.D. Spielberger (Ed.), Anxiety: Current trends in theory and research Vol 2. New York: Academic Press.

Frijda, N.H., (1986). *The emotions*. Cambridge: Cambridge University Press.

Gainotti, G. (2000). *Neuropsychological theories of emotion*. In J.C. Borod (Eds.) The Neuropsychology of Emotion, (pp. 214-236). New York, NY, Oxford University Press.

Gray, J. A. (1988). *The psychology of fear and stress*. 2nd edition. Cambridge, MA: Cambridge Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion,* 4, 269-288.

Gray, J. A. (1994). *Three Fundemental Emotional Systems*. In Ekman, P & Davidson, R J (Ed.), The Nature of Emotion (pp. 243-247). New York: Oxford University Press.

Gray, J. A., and McNaughton, N (2003). *The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system*. 2nd edition. Oxford: Clarendon Press.

Handler, M, Rovner, L. and Nelson, R. (2008). The concept of allostasis in polygraph testing. *Polygraph* 37 (3), 228-233.

Harre, R. (1986). *The Social Construction of Emotions*. Oxford: Basil Blackwell.

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

Honts, C. and Raskin, D. (1988). A Field Study of the Validity of the Directed Lie Control Question. *Journal of Police Science and Administration*, *16*, 56-61.

James, W. (1890). *Principles of psychology*. New York: Dover.

Kehoe, E.J. and Macrae, M. (2002). *Fundamental behavioral Methods and findings in classical conditioning*. , In J.W. Moore (Eds.), A Neuroscientist's Guide to Classical Conditioning, (pp. 171-231), New York, N.Y., Springer.

Khan, J., Nelson, R., and Handler, M. (2009). An exploration of emotion and cognition during polygraph testing. *Polygraph*, 38 (3) , 184-197.

Lane, R. D., Nadel, L., Allen, J. J. B. and Kaszniak, A.W. (2000). *The study of emotion from a perspective of cognitive neuroscience*. In Lane R D & Nadel L (Eds.), Cognitive Neuroscience of Emotion (pp. 407-410). New York, NY: Oxford University Press.

Lazarus, R. (1994). *Appraisal: The Long and the Short of It*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 208-215). New York: Oxford University Press.

Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.

Le Doux, J. E. (1994). *Cognitive-Emotional Interactions in the Brain*. In P. Ekman & Davidson R J (Eds.), The Nature of Emotion (pp. 216-223). New York: Oxford University Press.

Le Doux, J. E. (1996). *The Emotional Brain*. New York: Simon and Schuster.

LeDoux, J.E. (2000). *Cognitive-Emotional Interactions: Listen to the Brain*. In Lane R D & Nadel L (Eds.), Cognitive Neuroscience of Emotion (pp. 129-155). New York, NY: Oxford University Press.

Levenson, R. W. (1994a). *Human emotion, a functional view.* In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 123-126). New York: Oxford University Press.

Levenson, R. W. (1994b). *The Search for Autonomic Specificity*. In P. Ekman & R. J. Davidson (Eds.), The Nature of Emotion (pp. 252-257). New York: Oxford University Press.
(pp. 252-257). New York: Oxford University Press.

Levenson, R.W., Ekman, P., and Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27, 363-384.

National Research Council (2003). *The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph.* Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Ohman, A. (1993). *Fear and anxiety as emotional phenomena: Clinical Phenomenology, evolutionary perspectives, and information processing mechanisms*. In M. Lewis & J.M. Haviland (Eds.), The handbook of emotions. New York: Guilford.

Ortony, A., Clore, G. L. and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Panskepp, J. (1982). Toward a general psychological theory of emotions. *Behavioral & Brain Sciences,* 5, 407-467.

Plutchick R (1980). *Emotion: A Psychoevolutionary synthesis*. New York: Harper & Row.

Plutchik, R. (1962). *The emotions: Facts, theories, and a new model*. New York: Random House.

Power, M. and Dalgleish, T. (2008). *Cognition and Emotion, from Order to Disorder*, 2nd Edition. New York, NY: Psychology Press.

Raskin, D.C. and Honts, C.R. (2002). *The Comparison Question Test*, In M. Kleiner (ed.) Handbook of Polygraph Testing (p. 1-48), San Diego, CA, Academic Press.

Reber, A.S. (1995). *The Penguin Dictionary of Psychology*, 2nd edition, New York, NY, Penguin Books.

Roseman, I. J. (1984). *Cognitive Determinants of Emotion: A structural theory*. In Shaver, P. (Ed.), Review of Personality and Social Psychology: Emotions, relationships, and health, Vol. 5. Sage, Beverley Hills, CA, pp. 11-36.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.

Schacter, S. and Singer, J.E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 69, 379-399.

Scherer, K.R. (1982). Emotion as a process: Function, origin, and regulation. *Social Science Information*, 21, 555-570.

Scherer, K. R. (1984b).*On the nature and function of emotion: A component process approach.* In K. R. Scherer & P. Ekman (Eds.), Approaches to emotion (pp. 293–317). Hillsdale: NJ: Erlbaum.

Scherer, K.R. (2000). *Psychological models of emotion.*, In J.C. Borod (Eds.) The Neuropsychology of Emotion, (pp. 137-162). New York, NY, Oxford University Press.

Scherer, K. R. (2001).*The nature and study of appraisal: A review of the issues*. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), Appraisal processes in emotion: Theory, Methods, Research (pp. 369–391). New York and Oxford: Oxford University Press.

Scherer, K. R., Schorr, A and Johnstone, T. (2001). *Appraisal processes in emotion: Theory, Methods, Research*. New York and Oxford: Oxford University Press.

Schulkin, J. (2003). *Rethinking Homeostasis, Allostatic Regulation in Physiology*. Cambridge, MA: MIT Press.

Shweder, R.A. (1993). *The cultural psychology of emotions*. In M. Lewis & J.M. Haviland (Eds.), Handbook of Emotions (pp. 417-434). New York: Guilford Press.

Sinha, R., Lovallo, W.R., and Parsons, O.A. (1992). Cardiovascular differentiation of emotions. Psychosomatic Medicine, 54, 422-435.

Smith, C. A., and Kirby, L. D. (2001). *Toward delivering on the promise of appraisal theory.* In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), Appraisal processes in emotion: Theory, Methods, Research (pp. 121-138). New York: Oxford University Press.

Sterling, P. and Eyer, J. (1988). *Allostasis: a new paradigm to explain arousal pathology*. In S. Fisher & J. Reason (Eds.), Handbook of Life Stress, Cognition and Health (pp. 629-649). New York: Wiley and Sons.

Sterling, P. (2004). *Principles of allostasis: optimal design, predictive regulation, pathophysiology and rational therapeutics*. In J. Schulkin (Ed.), Allostasis, Homeostasis, and the Costs of Adaptation (pp. 17-64). Cambridge, MA: Cambridge University Press.

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

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: of brief measures of Positive and Negative Affect: The PANAS Scale. *Journal of Personality and Social Psychology*, 54, 1063-1070.

Acknowledgements

The authors thank Mr. Donald Krapohl, Mr. Chris Fausett, Dr. Charles Honts and Dr. Tim Weber for their thoughtful reviews and comments to earlier drafts of this paper. The authors would grant unlimited use and duplication rights to any polygraph school accredited by the American Polygraph Association or the American Association of Police Polygraphists. Questions and comments are welcome at polygraphmark@gmail.com.

"ABSTRACT THESIS" by Bruce Robertson

(For publication purposes all footnotes are listed at the end of this paper just prior to the List of References)

This thesis studies the polygraph Empirical Scoring System (ESS) to determine its potential use in homeland security and the war on terror. The research based its analysis on raw data previously collected by other researchers, who removed identifications from the data and subsequently provided it for study here. The results are described in regards to criterion accuracy; diagnostic capability; proportions of correct, errors, and inconclusive results; and the difference in scoring accuracy based upon participant employment and experience. Twelve scorers in three cohorts scored 22 You-Phase examinations taken from the Department of Defense–confirmed archives. One cohort used the three-position test data analysis (TDA) system, another cohort used the seven- position TDA system, and the final cohort used the ESS TDA system. All TDA systems proved equally capable of diagnostic ability. ANOVAs showed no significant differences between the distributions of ESS and transformed scores. No significant differences were found in decision accuracy with correct, inconclusives, errors rates for ESS scores, and those from the other two TDA systems. That ESS can complement other current hand- score TDA systems is suggested. However, that it could supplant other TDA systems is not confirmable by this study. Further study is recommended.

ACKNOWLEDGMENTS

My heartfelt thanks to the faculty and staff of the Center for Homeland Defense and Security at the Naval Postgraduate School for broadening my mind, helping me see the world through a different set of eyes, and making me and my classmates part of their family. Your dedication and devotion are unsurpassed.

Special thanks to Dr. Lauren Wollman and Mr. Don Krapohl for encouraging me through this process and keeping me on track. The final product would not have been possible without your support and hard work.

Mr. Raymond Nelson and Mr. Mark Handler are geniuses of the polygraph profession. Without their professional guidance, analytical skills, and insight, this thesis never would have gotten off the ground. And working with them as their research assistant, Julie Gerspacher showed an extraordinary understanding and patience of things for which I knew little and never let me be discouraged at the end of the day. All of your encouragement propelled me to the light at the end of the tunnel and for that I am ever in your debt.

Most of all, I send special love to Beckie, Amanda, Lynn, David, Kenny and Evan. Your understanding, patience, and encouragement during the missed days, hours, weeks, and months while I chased yet another dream cannot be repaid—but I will do my best!

I. INTRODUCTION

Physiology has been used in the United States in the detection of deception since World War I, when the government commissioned Dr. William Marston[1] to devise a technique to question prisoners of war (Alder, 2007). Intelligence officials from the National Research Council sponsored Marston's research. In his first real-world case, Marston used his techniques to attempt to identify the culprit in the theft of a military codebook from the United States Surgeon General's office. Although he narrowed the field of suspects to one, there is no record that the identified man was ever charged or in fact had committed the theft (Adler, 2007). Method more than instrumentation was Marston's contribution to lie detection. He believed that, by monitoring changes in systolic blood pressure, verbal deception could be detected.

As described by Ball and Gillespie on their website: He used a standard blood pressure cuff, or sphygmomanometer, to take measurements of systolic blood pressure during interrogation. This was the first time anyone used any kind of an instrument to detect truthfulness or deception. His method was simple. Take and record the subject's blood pressure, release the cuff. Ask the subject a question. Take and record the subject's blood pressure once again to identify any changes. He called this the "discontinuous method" of detecting deception. (Ball & Gillespie Polygraph, n.d.)[2]

The polygraph, with this long and controversial history, has been used at the federal, state, and local levels for a variety of purposes ever since. These uses include criminal cases, pre-employment screening, informant and witness testing, and counterintelligence purposes (Warner, 2005). There are 26 federal polygraph programs spread across nine federal agencies (see Table 1 for a listing of the polygraph programs), as well as numerous state and local law enforcement agencies.

Table 1. Federal Agencies That Utilize the Polygraph[3]

Department of Defense	Non-Department of Defense
Air Force Office of Special Investigations	Alcohol, Tobacco and Firearms
Army Intelligence Polygraph Program[4]	Bureau of Prisons/Office of Internal Affairs
Army Intelligence Polygraph Program[5]	Customs and Border Protection/Internal Affairs
Defense Criminal Investigative Service	Coast Guard Investigative Service
Defense Intelligence Agency	Central Intelligence Agency
Naval Criminal Investigative Service	Drug Enforcement Administration
National Geospatial-Intelligence Agency	U.S. Department of Energy
National Reconnaissance Office	Federal Bureau of Investigation
National Security Agency	Food and Drug Administration
U.S. Army Criminal Investigation Command	Homeland Security Investigations[6]
Defense Intelligence Agency	Internal Revenue Service-Criminal Investigation
	Transportation Security Administration
	U.S. Postal Inspection Service
	U.S. Postal Inspection Service, Office of Inspector
	General
	United States Secret Service
	Veteran Affairs Office, Office of Inspector General

The controversy over polygraph validity and reliability is ongoing, but the utility of the polygraph to obtain information is widely acknowledged (Warner, 2005). In homeland security and the war on terror, the polygraph has many applications. Specifically, it has been used by intelligence and other federal agencies for counterintelligence and espionage purposes. Many agencies use it as part of ongoing security screening programs for current employees. The Central Intelligence Agency's (C.I.A.) Aldrich Ames case and the Department of Energy's Wen Ho Lee[7] case are just two controversial examples of polygraph use in espionage investigations at the federal level. These two cases exemplify why scoring techniques are so important to the field and why poor technique or diagnostics or lack of interrater reliability can be detrimental to national security. The common public perception is that Ames passed his polygraphs (Alder, 2007; Pentagon's intelligence arm, 2008) while the polygraph was partially responsible for the bungling of the Lee investigation (Hoffman & Stober, 2001; Alder, 2007; Wen Ho Lee's Problematic Polygraph, 2000). It is myth that the polygraph's alleged failure allowed the two men to continue their deception.[8] These cases raised questions about the very foundation on which the government bases its use of the polygraph for national security purposes. A brief look at each case will demonstrate some of the issues of test data analysis, the component of the polygraph process that this research studies.

Ames, a Central Intelligence Agency Directorate of Operations officer, was arrested in 1994 for selling information to the Soviet Union. According to publications, Ames had spied for the KBG for nine years, and his duplicity had resulted in the death of at least ten agents who had spied for the C.I.A. in the Soviet Union (Earley, 1997; C.I.A., n.d.). In 1994, Dan Glickman, the House Intelligence Committee chairman, noted that the Federal Bureau of Investigation had concluded that Ames did not pass either of two tests (Kleiner, 2002). Then-C.I.A. Director James Woosley in 1994 revealed that the F.B.I. had not properly investigated Ames's two failed polygraphs (Kleiner, 2002).

Wen Ho Lee, a naturalized United States citizen, became suspect as a Chinese spy in 1995, after his employer, the Department of Energy (DOE), deduced that China had stolen classified nuclear weapons designs that allowed the country to develop a miniaturized nuclear warhead (Wen Ho Lee Case Study, 2008). Lee had been employed at the DOE's Los Alamos National Laboratory in New Mexico since 1978 and later became a nuclear weapons scientist at the laboratory. During the 1980s and 1990s Lee had numerous contacts with Chinese officials and scientists, some on official business and others while attending parties or conferences (Hoffman & Stober, 2001). As part of his employment, Lee was subject to periodic polygraph examinations: one in 1984, one in 1998, and another in 1999 (*Polygraph and lie detection*, 2003). However, the results of these polygraphs are in dispute (*Polygraph and lie detection,* 2003; Wen Ho Lee's Problematic Polygraph, 2000). More specifically, it is the disagreement between the opinions rendered by the original polygraphist and other polygraphists who later reviewed the polygraph charts as to whether or not Lee was truthful that creates the issues of concern.[9] The interpretation and scoring of polygraph charts is the focal point of this thesis.

The polygraph is used by federal, state, and local governments in determining the credibility and suitability of prospective employees who potentially will have a role in homeland security and/or the war on terror. The author has primary-source information that polygraph pre-employment screening in a major city law enforcement agency uncovered two attempted infiltrations, one by a Chinese operative and the second by a member of Al-Qaeda.[10] In the case of the Chinese operative, the agent was to gain employment at a law enforcement agency and work there long enough to establish a record of credibility in order to later become an employee of a federal law enforcement agency. In the case of the Al-Qaeda affiliated applicant, the effort was just an attempt to infiltrate law enforcement in a major city, one with a large Muslim population.

The polygraph has been used in Guantanamo Bay, Kandahar, Bagram, and other front-line combat theatres. In September 2003, the Air Force Office of Special Investigations (AFOSI) deployed its first full-time polygraphist to Baghdad (Collins, 2004, p. 1). Prior to this, the Air Force had deployed polygraphists on temporary duty (TDY). The (then-) polygraph program manager, Special Agent Pat Muller, was quoted as saying, "The polygraph exams we have administered over there have been some of the most critical and important work we have ever done in this program" (Collins, 2004, p. 1). The scope of examinations in

the theatre of war includes vetting coalition force members, determining the veracity of prisoners and informants on whose information tactical operations are initiated, and assisting in the conduct of criminal investigations (Collins, 2004, p. 1).

A. PROBLEM STATEMENT

Information provided to decision makers should be as accurate, trustworthy, and robust as possible, and it is clear that the polygraph plays an important role in achieving these requirements. Each day decision makers in federal, state, and local governments rely on the results of polygraph examinations to make their decisions. In its 2002 Polygraph Program Annual Report to Congress (Department of Defense, 2003), the Department of Defense (DoD) reported that it had conducted 11,566 polygraph examinations.[11]

The possible results of a polygraph examination are: "no deception indicated" (passed), "deception indicated" (failed) or "inconclusive" (the tracings were such that no opinion can be rendered). These judgments are rendered using one of several scoring mechanisms. None of these manual scoring systems in common use will deliver error estimates except ESS. That is to say, there is no current scoring mechanism that allows the polygraphist or the consumer to compare a calculated probability of error to a stated tolerance for error (Handler et al., 2010). The p-value maps the scores over a probability distribution such that the consumer can estimate the error likelihood of a decision based on the scores. These error estimates allow the consumer to take a more informed value judgment about tolerance for risk or error. Other current scoring systems in use do not have the same empirical level of decision accuracy as ESS (Handler et al., 2010). ESS provides accuracy profiles to include the total proportion of correct, inconclusive, deceptive, truthful, sensitivity, specificity, false negative errors (liars called non- deceptive), and false positive errors (truthful called deceptive).

This has been the state of the profession since the early use of the polygraph, much to the derision of its critics. The employment of an empirically based scoring mechanism would allow polygraphists to render an opinion based upon confidence in a scientifically derived result. The questions therefore become: does the scoring mechanism that provides that p-value have at least the same or better accuracy profiles as current scoring mechanisms; how can it be applied, and would it be accepted?

B. RESEARCH QUESTION

The broad question under consideration is whether the accuracy profiles associated with various scoring techniques should have an impact on the technique chosen in the homeland security arena. Additional questions to support this analysis will include: 1) Are there differences in the effectiveness of the three-position, seven-position, and ESS test data analysis (TDA, chart interpretation) models at extracting diagnostic information from the raw data, as reflected by the distributions of numerical scores? 2) Are there significant differences in criterion accuracy for the three-position, seven- position, and ESS TDA models? 3) What is the effect on accuracy of transforming three- position and seven-position scores to ESS scores? 4) How accurate are the combined three-position, seven-position, and ESS results? How accurate are the combined results when all scores are transformed to ESS scores? Is the difference significant? 5) Are there differences in accuracy that can be attributed to experience? Does more experience result in increased accuracy? 6) Does accuracy vary with the examiner's type of employment? Are there differences between private examiners and those who work for government (law enforcement/federal government) agencies?

LITERATURE REVIEW

The polygraph has demonstrated an important role in homeland security and the war on terror. This role has included the screening of personnel within many federal, state, and local agencies across the United States to assist in ensuring that prospective hires do not have an illicit motive for joining the ranks. It is not only important to understand that criminals and terrorists alike have attempted and been successful in acts that threaten homeland security, but enemies of the nation have the intent to spy and/or recruit potential agents for the purposes of espionage within our intelligence agencies and throughout other levels of government. The polygraph was used in World War I in counter-intelligence operations. It gained greater and more specific use in Korea (Alder, 2007). Since then, it has been used to assist decision makers in taking strategic and tactical decisions that directly protected American assets and lives as well as those of our allies. This review will identify literature about polygraph scoring techniques that are currently relevant to the topic, as well as those that will enhance the reader's understanding of the field of the polygraph.

This literature review will address three areas related to the polygraph and hand- scoring techniques. The first section will give a brief overview of polygraph research. The second section will provide an overview of types of testing techniques. Finally, the third section will discuss research related to scoring techniques.

A. POLYGRAPH RESEARCH

Polygraph expert and researcher Stuart M. Senter claims that polygraph examination is an inimitable field: "Polygraph examiners are trained to accomplish a task that, in the mind of the public, should only be made possible through rapid advances in seemingly futuristic technological equipment or through the weaving of mystical powers thought to be proffered by wizards and magicians." In other words, Senter is implying that many consider polygraph nothing but a magic trick, subject to ridicule and derision (Senter, 2008). Senter goes on to note that

providing a more pragmatic view of the polygraph will be accomplished through increasing the body of knowledge about the field. To date, the research has focused on applied research. That is, it focuses on real- world problems and tends to ignore theoretical knowledge. However, the basic foundations of polygraph principles have been ignored. There is little work on the understanding of factors that look at the diagnostic value of the polygraph (Senter, 2008, p. 278).

The National Research Council points out that there must be a solid theoretical base to have confidence in polygraph tests, lest erroneous results in populations such as "spies and terrorists" fail national security *(Polygraph and lie detection*, 2003, p. 92). However, the field has not made proper use of theoretical systems about the processes that underlie the measurements taken by the polygraph (*Polygraph and lie detection*, 2003, p. 93). Further, the research on the concept of decision thresholds (which are part of scoring techniques) has largely been ignored in polygraph research.

The consensus is that, although improving, in order to bring the polygraph into the realm of a recognized science, robust research must continue to be pursued.

B. TESTING TECHNIQUES

This section is not an exhaustive overview of all testing techniques in use in the field of polygraph examinations. It is a literature review of sources pertaining only to the most common techniques currently being utilized. Donald Krapohl and Shirley Sturm, in their 2002 article in Polygraph identify a number of testing techniques. The Air Force Modified General Question Test is a single-issue, multiple-issue, or multi-facet technique (Krapohl & Sturm, 2002b). The Comparison Question Technique is a term applied to a number of test formats that use probable- or directed-lie test questions. A Concealed Information Test is a type of test that involves a series of tests in which one critical item is used in each series. The intent of the test is to determine the person's knowledge of the particular item. A Counterintelligence-Scope Polygraph (CSP) is a type of test given to federal government employees who have access to sensitive security information. The CSP is designed to "detect and deter espionage, security breaches, sabotage, or other acts against the government" (Krapohl & Sturm, 2002a, p. 172). A test format that is widely used in the field is known as the Modified General Question Test (MGQT). The MGQT consists of more relevant questions than comparison questions. It does not use what is known as a "symptomatic question."[12] A Modified Relevant/Irrelevant Technique is a specific-issue test that uses situational comparison questions, which are then compared to the relevant questions. The relevant/irrelevant technique is a family of test formats that forgo the use of a traditional comparison question. They are most widely used in screening tests. U.S. government agencies use a test known as the Test for Espionage and Sabotage, which is a multi-issue screening test typically used with government employees who have access to sensitive information or programs/projects. The Utah Technique is a technique that uses modules of questions that consist of a comparison, relevant, and irrelevant question. The You Phase is a single-issue test in which the relevant question is slightly varied throughout the test. It is a highly focused test. The Zone Comparison Test (ZCT) uses three zones that refer to categories of questions (relevant, comparison, and symptomatic) that then compare two of the zones (relevant and comparison) to determine whether the examinee was truthful or deceptive. It is designed to "focus their attention to specific zone question(s). It is the first modern polygraph technique to which numerical analysis was widely applied" (Krapohl & Sturm, 2002b).

C. SCORING TECHNIQUES

The global evaluation technique is one in which the polygraphist visually inspects the charts to determine whether there is a stronger response to the relevant questions. It is most commonly used to score the Relevant/Irrelevant Technique (RI Technique). The NRC, as well as Krapohl and Dollins, notes that there is a lack of standardization to the scoring technique and that it has numerous idiosyncrasies (*Polygraph and lie detection*,2003; Krapohl & Dollins, 2003). Literature on this technique is scant, and its general use has declined.

The technique favored by most current polygraphists is numerical scoring in its several variations. The introduction of numerical scoring for the Comparison Question Technique is attributed to Cleve Backster, a well-known school director and instructor in modern polygraph techniques (Weaver, 1980). He introduced the seven-position scoring TDA system. The scale assigns scores ranging between +3 and -3 to the respective questions and their "comparison" questions. Weaver notes that the scoring technique was first developed by Backster to assist students in chart analysis in classroom settings.

In later research conducted by the University of Utah, it was concluded that numerical scoring had higher rates of accuracy and reliability than other scoring techniques (Raskin, Barland, & Podlesney, 1978), and it became the benchmark for the profession. The scoring system has evolved to include a three-position TDA system. This scoring system is now in wide use by polygraphists.

Krapohl and Dollins undertook what they described as a rudimentary investigation of the three primary scoring rule systems that can be applied to these numerical scoring techniques (Krapohl & Dollins, 2003). The three scoring systems are known as the Utah, the Backster, and the federal scoring systems. These scoring systems have three common components: scoring rules, computation rules, and decision rules (cut scores) (Krapohl & Dollins, 2003, p. 150). It is important to understand these three terms as used in the literature as they will be explored further as part of this research. Scoring rules are those that relate to the choice of tracing features in the charts, rejection of artifacts, and the

choice of how question pairs are compared and numbers assigned to the scheme. The weight and how the numbers are combined describe the computation rules. Decision rules, otherwise known as cut scores, govern the relationship between the computation rules and the polygraphist's choice of a decision (opinion), which will either be Deception Indicated (DI), No Deception Indicated (NDI) or inconclusive (INC) (Krapohl & Dollins, 2003).

Decision rules predominated in conclusions reached by the NRC and Krapohl, Stern and Bronkema (*Polygraph and lie detection*, 2003; Krapohl, Stern, & Bronkema, 2009). Specifically, each came to the conclusion that risk tolerance, and the corresponding decision rules, should be set by the consumer of polygraph results. That is, this decision should not be left to the polygraphists, but to the consumer of the results, who ultimately decides what risk can be accepted in the decision making process. In short, the determination of decision rules is a policy decision and will come into play later in the discussions of this research.

Two things become apparent in the literature: Those who speak to the topic agree on the paucity of research into the polygraph, and some note that the research concerning scoring techniques is even rarer. Secondly, the research into hand-scoring techniques looks into many things. Prior research includes accuracy and reliability of the scoring technique and the relative simplicity or lack of it within the respective technique and interrater reliability. What prior research lacks is the incorporation of the study of normative data (Handler et al., 2010).

Another scoring technique—the topic of this research—is the Empirical Scoring System (ESS). This scoring system was first described by Krapohl, Nelson, and Handler in 2008 (Krapohl, Nelson, & Handler, 2008). The development and research conducted on ESS allowed for the first time in the development of a polygraph hand-scored technique the application of p-values and normative data. It is profound in its simplicity, and based on associated p-value tables in regard to specificity, sensitivity, and inconclusive rate, the decision maker or policy setter can compare the probability of error and choose the error rate that best fits into his schema for risk aversion. It is because of this unique ability, in conjunction with the simplicity of its use, that ESS may prove to be the most robust scoring technique and capable of protecting American lives and assets at home and in the field of combat.

II. HYPOTHESES OR TENTATIVE SOLUTIONS

The polygraph is used in many circumstances for the purposes of national security, as well as law enforcement and security issues at the state and local levels. Its use in combat zones as well as the rear areas in theatres of war is documented. It has proven to be an extremely useful tool by assisting decision makers in the field to make both strategic and tactical decisions. The claim is that, by providing polygraph experts with a simpler hand-scoring technique, based on empirical data to which probability values have been determined, they in turn can provide these decision makers with a more informative answer to the questions at hand. In the combat arena, those questions can revolve around whether or not to undertake a tactical operation based on the word of an informant, collaborator, or captured enemy combatant. Such decisions involve great risks to life and limb, and the decision makers must be given the best tools available to make them. In other homeland security concerns, they can involve the credibility of informants, witnesses, accused or suspected criminals, spies, and other ne'er-do-wells.

Evidence to support this claim can be found in the review conducted by the National Research Council (NRC). The NRC notes that decision scientists and policy advisers have worked to develop systematic methods for resolving hard decision problems that arise in business, medicine and public policy. These methods are used explicitly in many scientific articles, and they are used implicitly in practical advice, where the goal is to get decision makers to think systematically before acting. (*Polygraph and lie detection*, 2003, p. 358).

The history of the polygraph is such that the lack of a sound scientific basis, in the minds of some, has led to the dismantling of various polygraph programs,[13] caused decision makers to be reluctant to rely on it—even in the absence of alternatives—and caused much conversation in the halls of Congress, state houses, and local government buildings as to its usefulness. It is a proven tool in the war on terror and national security. The Empirical Scoring System is the simplest hand-scoring technique to have empirical and scientific support as its foundation.

IV. SIGNIFICANCE OF RESEARCH

A. LITERATURE

There is a dearth of literature on scientific and empirically based hand-scoring techniques in the field of polygraph examination, particularly the impact of the techniques used on the robustness of decisions taken by those who rely on the polygraph to assist them in their decision making process. The Empirical Scoring System is one of the first and simplest hand-scoring techniques with intent to anchor TDA on empirical evidence and scientific study (Handler et al., 2010). This research should not only impact the use of the polygraph as it relates to national security, homeland security, and the war on terror, but it should further the scientific advancement in the polygraph community as a whole.

B. FUTURE RESEARCH EFFORTS

This research will reinforce the concept that a solid scientific basis for the polygraph will enhance its use and make it more readily defensible. The National Research Council (NRC) has stated that no lie detection technique has

been shown to outperform the polygraph and none shows any promise in the near term (*Polygraph and lie detection*, 2003, p. 173). However, it also notes that past efforts at polygraph research have not laid a sound foundation of scientific knowledge in the field (*Polygraph and lie detection*, 2003, p. 213). On page 221 of its review (*Polygraph and lie detection*, 2003), the NRC goes on to say that the detection of deception and information withholding is important to national security and that "government agencies will continue to seek accurate ways to detect deception by criminals, spies, terrorists, and others who threaten public safety and security interests." This thesis is just one small part of this effort, and it is hoped that it encourages others in the field and those who are consumers of its product to engage in further scientific study, particularly as it relates to security on the national, state, and local levels.

C. CONSUMERS

The immediate consumers of this research are the Department of Defense and its various military branches, as well as all federal agencies that have polygraph programs in place as part of their national and internal security interests. Further, all state and local law enforcement and criminal justice agencies who rely on polygraph results as part of their decision-making process should find this research useful. It is anticipated that the three national polygraph associations—the American Association of Police Polygraphists, the American Polygraph Association, and the National Polygraph Association—will utilize this research in the training and education of their respective members.

D. HOMELAND SECURITY PRACTITIONERS AND LEADERS NATIONALLY

This research should be of interest to many federal program managers within DHS and various federal agencies outside DHS, both those who use the polygraph and others who may not for various reasons. As this is just one small step in an effort to roll a component of lie detection onto a sound scientific basis, it can be anticipated that those who have been reluctant to utilize the polygraph, or perhaps even those who have been detractors of the field, might be encouraged and convinced to reconsider their positions.

V. METHOD

The present research based its analysis on raw data previously collected by other researchers instrumental in the development of the Empirical Scoring System (ESS), who removed identifiers from the data and subsequently provided it for study here.

Data was obtained from three groups (Cohorts 1, 2, and 3) of four scorers each. These participants were randomly grouped volunteers from a group of 300 students trained in the Empirical Scoring System as part of a training seminar hosted by the American Association of Police Polygraphists[14] in Cambridge, Massachusetts, on March 28, 2011. Cohort #1 scored the sample examinations using the Empirical Scoring System Cohort #2 scored the examinations with the three-position Test Data Analysis (TDA) system (DACA, 2006) and Cohort #3 scored them using the seven-position TDA system (DACA, 2006).

The Empirical Scoring System is an evidence-based numerical hand-scoring technique used for test data analysis of polygraph charts obtained from comparison question tests (Nelson et al., 2012). The ESS system utilizes a three-position scale of +, 0, or—and relies on the *bigger-is-better* rule;[15] scores are assigned when the scorer visually observes a difference in reaction strength between relevant and comparison questions (Nelson et al., 2012). A positive score (+) is assigned when there is a larger response to a comparison question, and a negative (-) score is assigned when there is a larger response to a relevant question. In typical comparison-question test formats, relative questions are normally compared to comparison questions (Nelson et al., 2012).

In "Terminology Reference for the Science of Psychophysiological Detection of Deception"[16] (Krapohl & Sturm, 2002b), the seven- and three-position TDA systems are defined as follows:

7-position scale

System of assigning values to individual physiological responses in PDD, based on differential responding to relevant and comparison questions. The values in 7-position scoring are whole numbers between -3 and +3. By convention, negative values represent greater responding to relevant questions, while positive values indicate greater responses to comparison questions. A zero usually indicates equal or no reactions to the relevant and comparison questions, or that the spot does not meet minimum standards for interpretation. The assigned numbers are summed across all three PDD parameters for each question for all spots and all charts. There are thresholds for determinations of truthfulness or deception, with an inconclusive region separating them. In the PDD literature, the 7-position scale is sometimes referred to as a semi-objective scoring system. There are three major versions of the 7-position scoring system: Backster, Utah, and DoDPI. See: Bell, Raskin, Honts, & Kircher (1999); Swinford (1999); Weaver (1985)

3-position scale

Abbreviated form of the 7-position scale for PDD test data analysis. The major difference is that the range of values for each comparison is from -1 to +1, rather than the range of -3 to +3 in the 7-position scoring system. See: Capps & Ansley (1992); Krapohl (1998); Van Herk (1990)

The analysis method applied to the research questions was an analysis of variance (ANOVA), which will be further described for each research question.[17]

VI. RESULTS AND ANALYSIS

Each study participant provided a demographic data form (see Appendix D). This demographic data included age and experience as a polygraphist, as well as gender.

The average age was 54, with a standard deviation of three. The maximum age was 65, and the minimum age was 37. The median age was 58. Ages do not appear normally distributed.

There were ten males and two females. Females n=2 is too small for analysis. Compared to groups of equal size, differences in the group size is significant. Z=9.334 (p<.001) Test of Proportions. Gender was not evaluated as an independent variable in the remainder of the analysis.

The average years of experience were 15, with a standard deviation of three. Median experience was 14 years, and the mode was also 14 years. The maximum years of experience were 33, and the minimum was three. Proximity of the mean, median, and mode indicated no increased concerns regarding the normality of the distribution of participant ages.

The participants in the study included four private examiners, seven law enforcement examiners, and one federal examiner. Additional data collected on the hand-score sheets were the individual scores assigned by the participant to the two relevant questions on the three charts of each examination in the study sample. A score was assigned, according to the structured rubric for each scoring system, for the tracings of each of these sensors: pneumograph,[18] electrodermal (EDA),[19] and cardiograph.[20] Subtotal scores across all three PDD parameters for each question for all spots and all charts. There are thresholds for determinations of truthfulness or deception, with an inconclusive region separating them. In the PDD literature, the 7-position scale is sometimes referred to as a semi-objective scoring system. There are three major versions of the 7-position scoring system: Backster, Utah, and DoDPI. See: Bell, Raskin, Honts, & Kircher (1999); Swinford (1999); Weaver (1985) were calculated for each of the relevant questions, and a grand total was calculated for the test as a whole. Scores were then interpreted using structured decision rules, according to the requirements of each scoring method, to make categorical determinations as to no deception indicated (NDI),[21] deception indicated (DI),[22] or inconclusive (INC).[23] Inconclusive is sometimes referred to as "no opinion" or "indefinite." Each participant then rendered his personal confidence level in the opinion rendered.

A. RESEARCH QUESTION #1

1. Results

Do differences exist in the effectiveness of the three-position, seven-position, and ESS test data analysis systems at extracting diagnostic information from the raw data?[24]

It is important to understand that the end result of any polygraph examination, whether for event-specific criminal investigations, security screening, law enforcement pre-employment, or post conviction supervision of convicted offenders, is a set of tracings (charts) that can be systematically analyzed to make determinations of truthfulness or deception at rates that are greater than can be obtained by other methods. Other professions, such as medicine and education, use both diagnostic and screening methods in their respective fields. The scientific work that has been applied to these methods can also be applied to polygraph examination (*Polygraph and lie detection*, 2003). Among consumers of the information in both the medical and educational testing methods, there is a general implicit understanding that test results are helpful to professional decision results and have been shown to be significantly greater than chance, even if imperfect. This assumption is based on several predicated assumptions: that those administering and analyzing the tests have acquired advanced training and education; (and) that these practitioners are qualified in their respective fields to select, administer, and interpret tests that will provide information that will assist the referring professionals to make better decisions.

Although signal detection theory[25] is not an integral part of this thesis, it is important to understand that the diagnostic analysis of polygraph tracings involves signal detection, particularly as an underpinning in the scientific work necessary to advance the field. Signal detection involves the diagnostician's being able to distinguish between signals and noise. McNicol called it, "a theory about the way in which choices are made" (McNicol, 2005). Signal information is diagnostic information that the observer wants to see, and noise is any non-signal information or background noise (Keating, 2005) that can make the identification of diagnostic information difficult. Clearly, extracting diagnostic information from the "raw" data of polygraph tracings involves the diagnostician—in this case a polygraphist or blind reviewer—making observations about two states and assigning an assessment of which state he observes. Test sensitivity (*Polygraph and lie detection*, 2003) involves the effectiveness with which signal information can be extracted and used to identify the issue of concern. Test specificity also involves the effectiveness with which the absence of signal information is determined and affects the ability of a test to determine when the issue of concern is not present.

Harvey further describes this phenomenon in "Detection Sensitivity and Response Bias." He explains that the "detection performance" (diagnostics) is based on both a sensory process and a decision process. A simple yes or no can be the response as to whether or not a signal was present, or there can be a "rating of the confidence" that a signal was present. In the case of most polygraph TDA systems, the response is a yes or no, with the value of yes described in a positive or negative number. This involves a sensory process (sensitivity), as well as a decision process

with a defined criteria parameter (in this case, the instructions contained on the hand-score sheet) (Harvey, 2003).

In signal detection theory, this sensory and response criteria process involves "hits" and "misses." That is, there is a hit when the diagnostician says yes to the signal that is present (hit rate), and a miss (false-alarm rate) occurs when the diagnostician says yes to a signal that is not present, meaning that noise was wrongly identified as a signal. Table 2 graphically displays this theory.

	"Yes"	"No"
Signal Present	Hit Rate (HR)	Miss Rate (MR)
Signal Absent	False Alarm Rate (FAR)	Correct Rejection Rate (CRR)

Table 2. Conditional Probabilities, Signal Detection Theory

In polygraph, the FAR and MR are respectively known as false positive[26] and false negative.[27]

2. Analysis Method

Three-position and seven-position TDA numerical scores were transformed to ESS scores and subjected to a 2 x 3 ANOVA (criterion state x TDA system) for absolute magnitude of mean numerical scores. Transformation to a common numerical scale ensures that differences are not attributable to scale differences and are a reflection of differences in the effectiveness with which examiners extract diagnostic (signal) information using the three TDA systems.

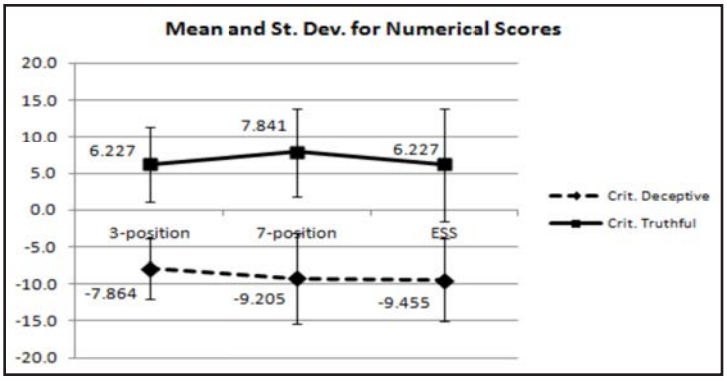


Figure 1. Mean and Standard Deviations for Numerical Scores

Table 3. 2 x 3 ANOVA Summary (Criterion State x TDA model) for Mean Scores

Source	SS[a]	df[b]	MS[c]	F[d]	p[e]	F crit .05[f]
TDA System	96.212	2	1.093	0.032	.969	3.031
Criterion state	284.379	1	2.154	0.063	.803	3.878
Interaction	44.576	1	44.576	1.294	.256	3.878
Error	8890.591	258	34.460			
Total	425.167	262				

[a] SS Sum of Squares
[b] df Degrees of Freedom
[c] MS Mean Square
[d] F The F Value
[e] p Probability Value
[f] F Critical Value of F with α= .05

The ANOVA analysis produced no significant differences—an indication that each of the TDA systems is capable of extracting similar signal (diagnostic) information from the raw data. That is, using any one of the three TDA systems, the polygraphist should be able to observe the criterion for truthfulness or deception, with no one system being more or less diagnostic.

B. RESEARCH QUESTION #2

1. Results

Are there significant differences in criterion accuracy for the three-position, seven-position, and ESS TDA systems?

Criterion accuracy (validity) refers to how effectively the testing system places individual cases in the correct criterion category. In polygraph, the signals intended to be captured are the test results of deception indicated or no deception indicated. In the case of a single issue examination, such as a criminal investigation or event-specific incident, this measure (criterion) is the polygraphist's opinion about the examinee's deception or truthfulness corresponding to actual truthfulness (ground truth).

Within signal detection theory, one measure of stimulus is sensitivity, discussed below. Another measure within signal detection is response bias. This thesis does not research response bias, and it is left for future research; however, it is important to understand that the phenomenon exists. Response bias is the tendency of the diagnostician to choose one response over another. In other words, the tendency of a diagnostician to favor, that is, to be biased toward, the selection of one response over another. The more features available, the more opportunities for a diagnostician to become biased. Detection theory allows for determining or delimiting the distributions consistent with bias or sensitivity and specificity of a test measure. Sensitivity and bias taken together all lead to a decision system in which the stimulus classes reach equal-variance normal distributions for the decision variable, making them more meaningful. This decision system can then be tested using receiver operating characteristic curves, which then leads us graphically to the proportion of hits (signal) to the proportion of false alarms (noise). This becomes important in determining how to manipulate response bias—either through instruction or by use of a confidence rating (p value) (Macmillan, & Creelman, 1996). More specifically, as response bias relates to polygraph scoring, the development of the ESS-TDA method is designed to reduce the response bias of polygraphists. Specifically, older TDA methods relied on more features and criteria to arrive at a final score. These attributes make the scoring methods difficult to learn (instruction) and more subjective (introducing response bias), with less interrater reliability (Blalock, Cushman, & Nelson, 2009). ESS utilizes the "bigger-is-better" rule, which means fewer features to score allows for ease in learning. Also, the ESS is the only hand-scoring method that has a p-value table. The use of the p-value addresses the second method of dealing with

response bias—the use of confidence rating. Again, response bias is a topic for future research, but it is mentioned here to demonstrate that ESS addresses it and that the p-values associated with ESS allow for criterion selection that addresses levels of sensitivity.

Sensitivity is but one aspect of accuracy (validity). If deception is perfectly indicated whenever a lie is present, then the signal proves positive (deceptive) whenever a lie is present; the measure is positive for deceptive in all positive cases and no false negatives are produced; in other words, perfect sensitivity (*Polygraph and lie detectio*n, 2003).

Specificity is the other aspect of accuracy. If deception is absent, then the signal always shows negative and is therefore perfectly specific to deception; it produces no false positives. A test is more specific the greater the proportion of persons who appear non-deceptive on the test; in other words, perfect specificity (*Polygraph and lie detection*, 2003).

2. Analysis Method

The analysis method used was multivariate ANOVAs (criterion state x TDA system) for decisions with inconclusives (i.e., test sensitivity to deception and test specificity to truthfulness), inconclusive rates, and error rates.

Table 4. Means, (Standard Deviations), and {95% Confidence Intervals} for Criterion Accuracy

	3-position	7-position	ESS
Sensitivity	.886 (.087) {.716 to >.999}	.841 (.136) {.574 to >.999}	.886 (.045) {.797 to .975}
Specificity	.591 (.091) {.413 to .769}	.614 (.087) {.443 to .784}	.727 (.129) {.475 to .979}
Inc D	.114 (.087) {<.001 to .284}	.159 (.136) {<.001 to .426}	.114 (.045) {.025 to .203}
Inc T	.341 (.155) {.037 to .645}	.341 (.114) {.117 to .565}	.182 (.148) {<.001 to .473}
FN Errors	<.001 (<.001) {<.001 to <.001}	<.001 (<.001) {<.001 to <.001}	<.001 (<.001) {<.001 to <.001}
FP Errors	.068 (.087) {<.001 to .239}	.045 (.052) {<.001 to .148}	.091 (.074) {<.001 to .236}

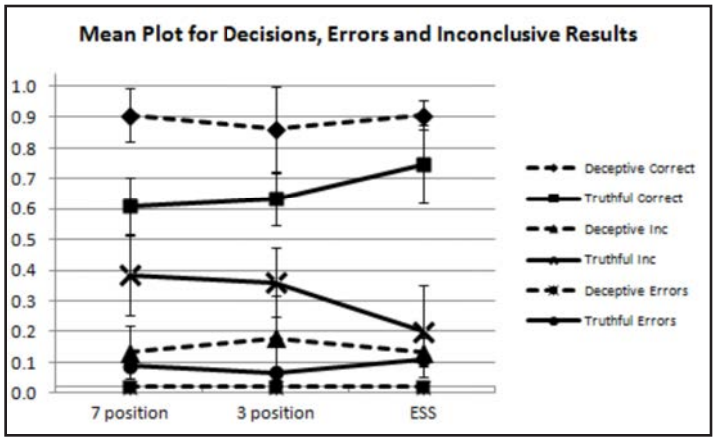


Figure 2. Mean Plot for Decisions, Errors, and Inconclusive Results

Table 5. Three-way (2 x 3 x 3) ANOVA Contrast for Test Accuracy (Criterion State x TDA System x Accuracy Dimension)

Source	SS	df	MS	F	p	F crit .05
Criterion dimension	6.814	2	3.407	392.260	.000	3.168
Status	0.000	1	0.000	0.018	.893	4.020
TDA system	0.000	2	0.000	0.013	.987	3.168
Criterion dimension x status	0.504	2	0.252	28.985	.000	3.168
Status x TDA system	0.000	2	0.000	0.013	.987	3.168
Criterion dimension x TDA system	0.082	4	0.020	2.352	.065	2.543
Criterion dimension x status x TDA system	0.054	4	0.014	1.559	.198	2.543
Error	0.469	54	0.009			
Total	7.923	71				

The value of this three-way contrast is that it encompasses the entire experimental question; it provides greater degrees of freedom; and it provides more power than a series of two-way analyses.

There was no significance in this three-way interaction, which suggests no statistically significant differences in the accuracy of the three compared TDA systems. It is noted that the two-way interaction was significant for criterion dimension (x case status). This suggests that the different TDA systems may perform differently with criterion truthful and criterion deceptive cases.

In this instance, differences in criterion dimension are expected, in that it is hoped that error and inconclusive rates are lower than decision accuracy rates. This main effect did not undergo additional analysis. The most significant interaction in the three-way analysis was the two-way interaction of criterion dimension x case status. Again, this interaction supports the expectation that correct, inconclusive, and erroneous will not result in similar proportions. Because of this, two-way post-hoc ANOVAs were completed for each of the three dimensions of test accuracy: decisions, errors, and inconclusive results.

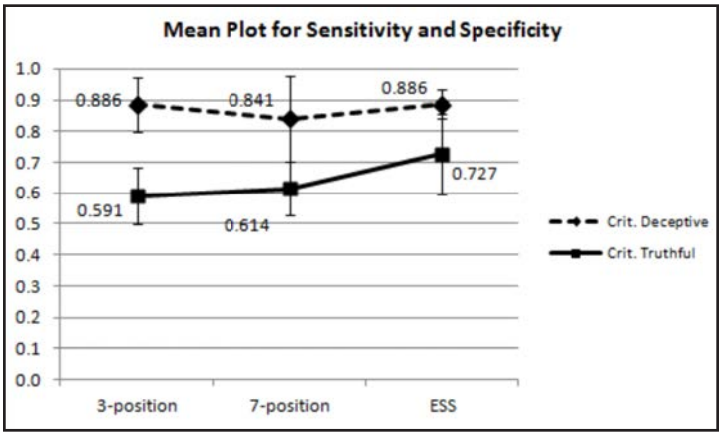


Figure 3. Mean Plot for Sensitivity and Specificity

Table 6. Two-way ANOVA Summary (Case Status x TDA System) for Decision Accuracy, Including Inconclusive Results (i.e., Sensitivity and Specificity)

Source	SS	Df	MS	F	p	F crit .05
TDA System	0.030	2	0.004	0.366	.698	3.555
Criterion state	0.310	1	0.026	2.557	.127	4.414
Interaction	0.019	1	0.019	1.841	.192	4.414
Error	0.182	18	0.010			
Total	0.358	22				

Neither the two-way interaction nor the main effects for case status or TDA system were significant for sensitivity and specificity.

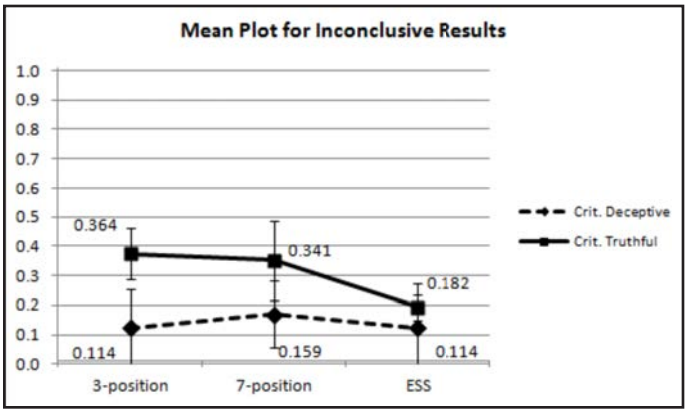


Figure 4. Mean Plot for Inconclusive Results

Table 7. Two-way ANOVA Summary (Case Status x TDA System) for Inconclusive Results

Source	SS	df	MS	F	p	F crit .05
TDA System	0.050	2	0.006	0.472	.631	3.555
Criterion state	0.167	1	0.014	1.043	.321	4.414
Interaction	0.034	1	0.034	2.534	.129	4.414
Error	0.240	18	0.013			
Total	0.251	22				

Neither the two-way interaction nor the main effects for case status or TDA system were significant for inclusive rates.

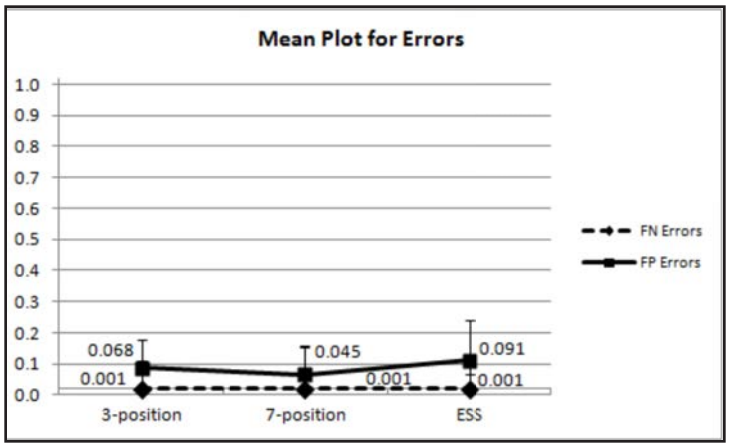


Figure 5. Mean Plot for Errors

Table 8. Two-way ANOVA Summary (Case Status x TDA System) for Inconclusive Results

Source	SS	df	MS	F	p	F crit .05
TDA System	0.002	2	0.000	0.098	.907	3.555
Criterion state	0.027	1	0.002	0.855	.367	4.414
Interaction	0.002	1	0.002	0.782	.388	4.414
Error	0.048	18	0.003			
Total	0.031	22				

Neither the interaction nor the main effects of case status and TDA system were significant for errors.

The results of these analyses indicate that the three-position, seven-position, and ESS TDA systems produce different rates of correct, erroneous, and inconclusive results. However, there was no significance in the differences in the three TDA systems. It is noted that this may be a result of sample size and the size of the cohorts. Larger sample sizes and larger cohorts may produce significant differences.

No statistical power analysis was completed. Confidence intervals can be found in the table of means (Table 4).

It is noted that there is an absence of false-negative errors in this study. In a 2006 study, Krapohl reported a field study with a false-negative rate at 2.7% (Krapohl, 2006). The current error rate should be taken as statistically meaningless. It is unrealistic to expect this in field settings or larger studies. The result should be used with caution.

C. RESEARCH QUESTION #3

1. Issue Posed

What is the effect on the accuracy of transforming three-position and seven- position scores to ESS scores?

2. Analysis Method

ESS scoring rules were applied to three-position and seven-position TDA systems, and a two-way ANOVA (TDA system x ESS transformation) was calculated.

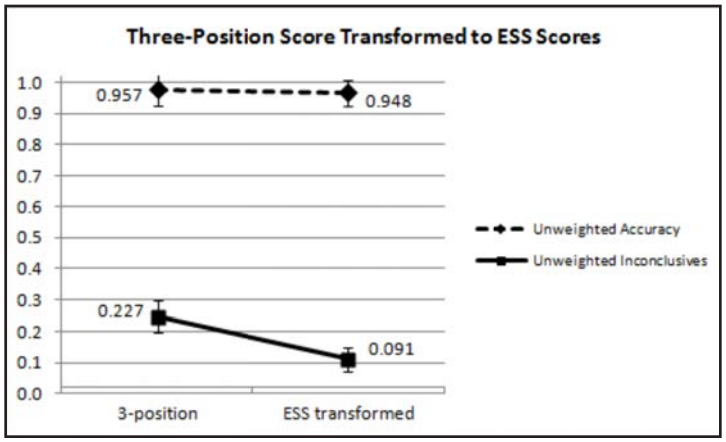


Figure 6. Three-Position Score Transformed to ESS Scores

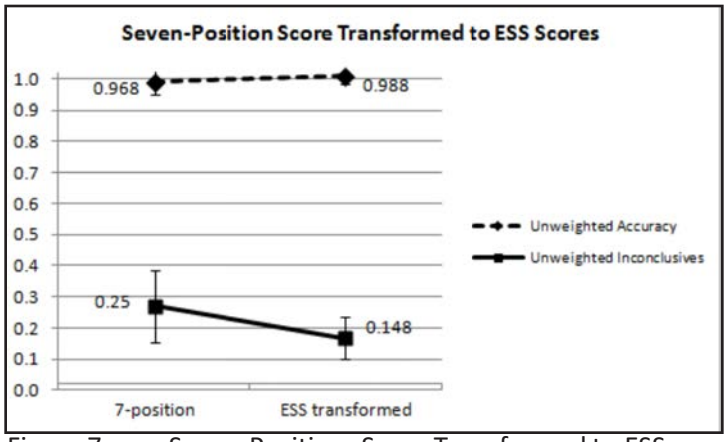


Figure 7. Seven-Positions Score Transformed to ESS Scores

Table 9. Unweighted Accuracy

	Raw	ESS transformed
3-position	.957 (.054) {.851 to >.999}	.948 (.041) {0.867 to >.999}
7-position	.968 (.037) {.895 to >.999}	0.988 (0.025) {.939 to >.999}

Table 10. Unweighted Inconclusives

	Raw	ESS transformed
3-position	.227 (0.052) {.124 to 0.33}	.091 (.037) {.018 to .164}
7-position	.250 (.114) {.026 to .474}	.148 (.068) {.014 to .281}

Table 11. Two-way ANOVA Summary (TDA System x ESS Transformation) for Accuracy

Source	SS	df	MS	F	p	F crit .05
Transformation	0.000	1	0.000	0.009	.928	4.747
TDA System	0.003	1	0.000	0.197	.665	4.747
Interaction	0.001	1	0.001	0.467	.507	4.747
Error	0.020	12	0.002			
Total	0.004	15				

No significant differences were found between the distributions of ESS scores and the transformed three-position and seven-position scores when a two-way ANOVA was conducted. Also, there were no significant differences in unweighted accuracy when transforming the scores of these TDA models to ESS scores.

Table 12. Two-way ANOVA Summary (TDA System x ESS Transformation) for Inconclusive Results

Source	SS	df	MS	F	p	F crit .05
Transformation	0.057	1	0.007	1.302	.276	4.747
TDA System	0.006	1	0.001	0.145	.710	4.747
Interaction	0.001	1	0.001	0.213	.653	4.747
Error	0.066	12	0.005			
Total	0.064	15				

There are also no significant differences between the three-position and seven- position TDA inclusive results when a two-way ANOVA was conducted for inconclusive results. A larger study may produce statistical power that could provide for expected improvement.

D. RESEARCH QUESTION #4

How accurate are the combined 3-position, 7-position, and ESS TDA results? How accurate are the combined results when all scores are transformed to ESS scores? Is the difference significant?

Table 13. Accuracy of ESS, Three-position and Seven-position Scores Combined

	Raw scores	All scores transformed to ESS Scores
Unweighted Accuracy	.957 (.043) {.874 to >.999}	.961 (.040) {.883 to >.999}
Unweighted Inconclusives	.208 (.090) {.032 to .384}	.129 (.064) {.004 to .254}

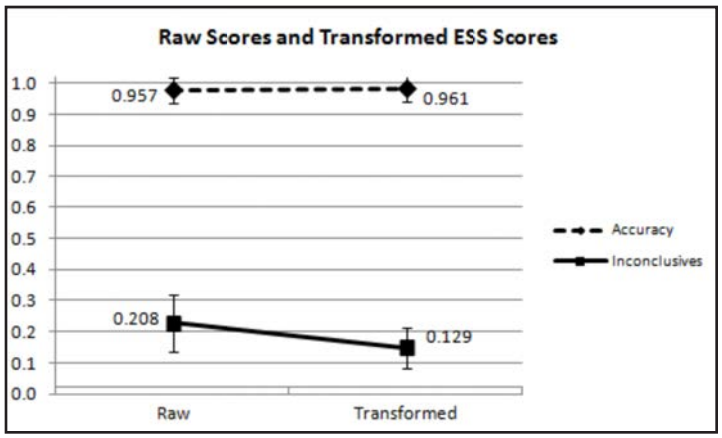


Figure 8. Raw Scores and Transformed ESS Scores

Table 14. Two-way ANOVA Contrast (transformation x accuracy dimension) for Test Accuracy

Source	SS	df	MS	F	p	F crit .05
Transformation	0.017	1	0.001	0.186	.669	4.062
Dimension	7.498	1	0.312	80.381	.000	4.062
Interaction	0.021	1	0.021	5.330	.026	4.062
Error	0.171	44	0.004			
Total	7.537	47				

Because it is known that the proportion of inconclusive differs from the proportion of correct, there is an expected significant main effect for accuracy dimension. ESS-transformed scores will produce different types of changes in decisions and inconclusive, as significant interaction for transformation and accuracy dimension suggests; decision accuracy increases and inconclusive results decrease.

One-way differences for decision accuracy were not significant [F(1,22) = 0.004, (p = 0.952)].

A larger sample may have found a significant difference in

these results: one-way differences for inconclusive results were also not significant [F(1,22) = 0.522, (p =0.478)].

E. RESEARCH QUESTION # 5

Are there differences in accuracy that can be attributed to experience? Does more experience result in increased accuracy?

The average years of experience are 15. The standard deviation is three. The maximum years of experience are 33. The minimum years of experience are three. The median years of experience are 14, and the mode is 14. None of the participants is considered inexperienced.

For the purpose of this research, fewer than ten years is considered low experience and more than ten years is considered high experience.

Table 15. Accuracy and Inconclusive Rates for Low-Experience and High- Experience Participants

	Low Experience	High Experience
Unweighted Accuracy	.958 (.043) {.873 to 1.042}	.963 (.041) {.883 to 1.044}
Unweighted Inconclusives	.118 (.069) {<.001 to .253}	.136 (.064) {.010 to .262}

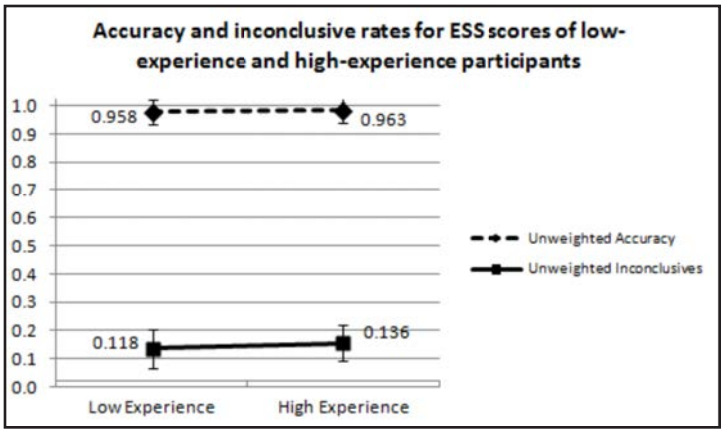


Figure 9. Accuracy and Inconclusive Rates for ESS Scores of Low-Experience and High-Experience Participants

Due to differences in sample size and an expected difference in decision and inconclusive rates, unbalanced one-way ANOVA were used.

Results between high- and low-experience participants were not significant for decision accuracy [F(1,10) = 0.009, (p = 0.925)]. Neither were results significant for inconclusive results [F(1,10) = 0.037, (p = 0.851)].

There was no effect for low or high experience in this sample data. That is, the low-experience participants scored polygraph charts using ESS with the same accuracy and inconclusive rates as high-experience participants. This outcome is consistent with that reported between inexperienced scorers and experienced scorers by Blalock, Cushman, and Nelson (2009) and Krapohl and Cushman (2006).

F. RESEARCH QUESTION #6

Does accuracy vary with the examiner's type of employment? Are there differences in accuracy between private examiners and those who work for law enforcement or government agencies?

One federal examiner was combined with the county/local law enforcement group for a combined group of government employees.

Table 16. Accuracy and Inconclusive Rates for ESS Scores of Private-Practice and Law Enforcement/Government Participants

	Private	LE/Gvt
Unweighted Accuracy	.977 (.026) {.926 to 1.029}	.953 (.045) {.865 to 1.04}
Unweighted Inconclusives	.114 (.079) {<.001 to .268}	.136 (.060) {.020 to .253}

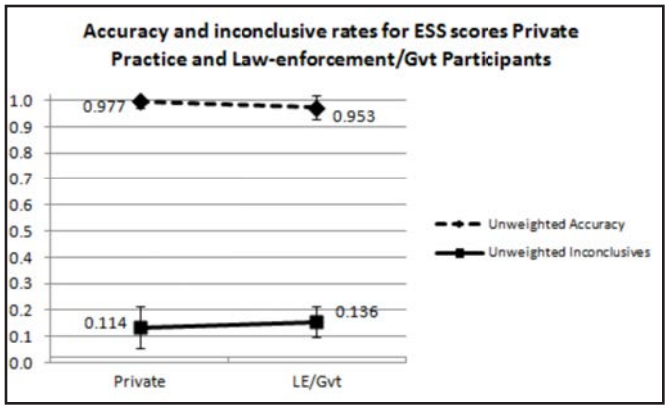


Figure 10. Accuracy and Inconclusive Rates for ESS Scores of Private-Practice and Law Enforcement/Government Participants

Due to differences in sample size and an expected difference in decision and inconclusive rates, unbalanced one-way ANOVA were used.

Results were not significant for decision accuracy [F(1,9) = 0.228, (p = 0.644)], nor were results significant for inconclusive results [F(1,9) = 0.053, (p = 0.823)].

There was no effect for type of employment in this sample data, although a larger sample size may be expected to produce different results.

VII. DISCUSSION

A. INTRODUCTION

Polygraph has been used as a tool by the federal gov-

ernment, the military, and state and local governments for several decades. It has been and continues to be used as a successful instrument in national-security issues, homeland security, and the war on terror. Nevertheless, its detractors and those unfamiliar with its utility and successes, as well as disagreements and lack of foresight within the profession itself, have caused some of the agencies and decision makers who do or could benefit from its use to be reluctant to rely on it. Some of this reluctance and even abandonment, in spite of the lack of alternatives, is due to outside pressure. The pressure can come in the form of political pressure, from uninformed law or policy, or from those who believe they have been wronged or harmed through the use of the polygraph. The challenge then has become multifold: Polygraph proponents must ensure that there is ongoing research to address the concerns and, in some instances, the valid arguments and criticisms of detractors (some of whom work in other scientific disciplines), and they must continue to move the profession onto a sound scientific foundation. Within the profession, infighting and lack of foresight and vision must be overcome for the sake of ensuring that this tool remains viable and valuable in its contribution to the homeland's safety and security, regardless of the form that proven instrumentation and technology takes. It is imperative that decision makers, policy makers, and other consumers who currently rely on the polygraph (as well as those who should) be educated by those within the profession who can and should undertake such a goal.[28]

Various studies and reviews have been undertaken in regard to hand-scored polygraph techniques. Two primary hand-scoring techniques in use today are the three- position and seven-position TDA systems. These two systems employ twelve scoring features for the purpose of assigning positive values (no deception indicated) or negative values (deception indicated) when the responses to relative questions are compared to the comparison questions. The rules (instructions) for assigning values are complex. The Empirical Scoring System uses observation of three scoring features for the purpose of assigning these negative and positive values. The instructions for assigning these values are simple and rely on the bigger-is-better rule.

The purpose of this study was to extend the research into the Empirical Scoring System to see whether it has additional value or is at least the equivalent of other hand-scoring techniques currently in use. Various research questions were posed, and through the use and analysis of raw data, comparisons were made between the three-position and seven-position scoring techniques, arguably two of the most highly utilized scoring techniques in the polygraph profession. These two techniques have been in use since the 1960s and are taught at the National Center for Credibility Assessment (NCCA),[29] as well as other private and government-funded polygraph schools across the United States and internationally.[30] Previous research on hand-scoring techniques was normative-based, while research on the ESS is empirically based and has allowed for the assignment of p-values to the technique. The intent of this study was to conduct additional research of the ESS, to further determine whether its design and method of use offer advantages over the compared techniques.

B. DISCUSSION

The research design, primarily through use of ANOVA, was intended to measure several facets of the ESS as compared to the three-position and seven-position TDA systems. The study used 22 archival matched random samples of You-Phase examinations from the confirmed case archive at the Department of Defense. Eleven of these cases were confirmed truthful examinations. "Confirmed truthful" in the instance of these 11 examinations means that an alternative person was identified as a suspect or the examinee was exonerated, as there was evidence or a confession outside the opinion rendered by the specific polygraphist. Eleven matching confirmed deceptive examinations were also provided. "Confirmed deceptive" in the instance of these 11 examinations, means that there was evidence or a confession outside the opinion rendered by the specific polygraphist. As per the You-Phase protocol, which is part of the examination technique, these are single-issue examinations that contain two relevant questions and three comparison questions, as well as other procedural questions.

The first question was to discover whether there were differences in the ability of each TDA system to extract diagnostic data from the provided examinations. This analysis was undertaken through use of ANOVA. The ANOVA analysis produced no significant differences between the three TDA systems; each was as capable as the others of extracting diagnostic information.

The next research question was whether there were differences in the three TDA systems for criterion accuracy (validity). This analysis was conducted through the use of participants were randomly selected and consisted of three groups (Cohorts 1, 2, and 3) of four scorers each. The first cohort utilized the ESS-TDA. The second cohort used the three-position TDA system. The third cohort used the seven-position TDA system. There were six research questions in the study multivariate ANOVAs and targeted inconclusives, inconclusive rates, and error rates. No significant differences were found in the three-way interaction. However, in the two-way interaction it was significant for criterion dimension. This suggests that the TDA systems may perform differently with criterion truthful and criterion deceptive cases. This interaction supports the expectation that correct, inconclusive, and error rates will not result in similar proportions. That is, the hit rate should be better than the miss rate and the indeterminate rate. The two-way interaction showed no significance for sensitivity or specificity, and this supports the expectation. Although the systems produced different rates of correct, errors, and inconclusive results, there are no significant differences in the three TDA systems. ESS seems to have a better specificity, lower inconclusive, and an equivalent error rate (which approached zero for all three TDA systems; again, an unrealistic result that should only be used

with caution). It is noted that both the sample size and the size of the cohorts may have had an effect on this lack of significance. A larger sample size and larger cohorts may produce significant differences. What can be said as a result of this research is that ESS appears to have at least the same criterion accuracy as the three-position and seven-position TDA systems.

Transforming the three-position and seven-position TDA scores to ESS scores was conducted to determine whether there was an effect on the accuracy of the three- and seven-position scoring systems. This was accomplished through application of a two-way ANOVA. No significant differences were found between the distributions of the three TDA systems when the two-way ANOVA was conducted, which means that there is a high correlation between the three when transformed. There was no significant difference in unweighted accuracy as the result of transformation. In terms of inconclusive rates for the transformed three-position and seven-position scores transformed, no significant differences were found. This was an unexpected result. The expectations were that the inconclusive rates would be higher for both the three-position and seven-position TDA systems transformed to ESS. This unexpected result is likely the result of a small sample size. A larger study should produce statistical power that may provide the expected results.

A fourth research question looked at the accuracy of the result of the combination of the raw scores of all three TDA systems to ESS. There was a significant main effect for the accuracy dimension, and this was expected, given that it is known that the proportion of inconclusive will differ from the proportion of correct. Decision accuracy increased and inconclusive results decreased. One-way differences in decision accuracy were not significant; however, neither were one-way differences for inconclusive results. It is hypothesized that a larger sample may find significant differences.

The fifth question under study was to discover whether there were differences in accuracy based on level of experience. A one-way unbalanced ANOVA was utilized for analysis, and there were no significant differences for decision accuracy or inconclusive results based upon experience. Although there were no inexperienced participants in any of the cohorts, there were participants with low experience and participants with high experience. The results seem to show that there is no effect in the application of ESS scoring techniques based on years of experience in the field of polygraph scoring.

A final analysis was conducted to determine whether type of employment, private or government, had any effect on accuracy. An unbalanced one-way ANOVA found that there were no significant differences for decision accuracy or for inconclusive results. There was no effect for type of employment based on this research. However, it is again hypothesized that a larger sample size might produce different results. The analysis seems to indicate that type of employment has no effect on TDA diagnostics.

C. LIMITATIONS

Sample size was the primary limitation of the current study, both in terms of confirmed case sample, as well as the number of participants. Larger sample sizes would produce more statistical power, and it is hypothesized that they would have an impact on the significance of some of the findings of this study. Additionally, the study participants were experienced polygraphists who had attended a continuing education seminar and classroom instruction on ESS. It cannot be concluded that these cohorts are representative of the wider population of polygraphists. Another limitation is that it is not known how the confirmed cases came to be selected into the archive, other than being confirmed cases. The researchers intentionally used cases confirmed by extra-polygraph means, but one must consider that the selection may potentially lead to criterion accuracy rates that could be overestimated.

D. RECOMMENDATIONS FOR FUTURE RESEARCH

Several recommendations for future study can be made as the result of this research. First, since the sample size was small, it is possible that the statistical power in a larger sample could reveal differences that escaped detection in this project. This larger sample size may be of interest in research by type of employment, decision accuracy, criterion accuracy, and other considerations. It is of note that government polygraphists, particularly federal government examiners, typically attend government-sponsored polygraph schools, while private examiners typically attend private schools, although many private examiners are retired government polygraphists. The results may reveal differences in instruction, expectations, types (quality) of exams conducted, or overall workload (number of tests conducted), among other possible variables that can then be studied.

Another aspect of test data analysis that may be of interest is the amount of time required to use the various types of hand-scoring techniques. These time studies can then be correlated to other demographic facets of the participants, again including age, experience, and type of employment. The p-value tables for ESS are well developed, although their significance to field polygraphists, decision makers, and other consumers is not well known.

Response bias is another issue that was not addressed by this study. It has a direct impact on sensitivity and should be further studied. Further research into the potential importance of this attribute of ESS and its potential contribution to policy decisions should be undertaken. This research suggests that ESS can at least complement, if not supplant, the two compared TDA systems, and perhaps others, to increase the value of polygraph to homeland security and the war on terror. Further research should be conducted into this potential.

E. CONCLUSION

The first conclusion that can be drawn from this study is that ESS has at least as much diagnostic ability as the three-position TDA system and the seven-position TDA system, even taking into consideration the newness of the ESS to the polygraph profession. Despite the limitations of the small sample size, the study produces partial evidence and suggests that ESS has consistently high criterion accuracy. The study hints that ESS seems to have a better specificity, less inconclusive rates, and at least equivalent error rates as the three-position and seven-position TDA systems. Lastly, the study seems to support past research into ESS that the length of experience has no impact on the ability of the polygraphist to apply ESS scoring rules. This offers a particular advantage over the more complex scoring systems, which include more features and scoring rules, since graduates of polygraph schools seldom have time to ease into their new jobs. That is, the new graduate of a polygraph school can typically expect that soon after his assignment, he will undertake a polygraph examination that has high impact and consequences. The impact and consequences can literally save or cost lives, determine the future course of major tactical plans and actions, or forever change the lives of individuals.

It is imperative that those who can impact the use of the polygraph in the United States continue to pursue the lofty goal of sound and scientifically based lie detection techniques, procedures, instrumentation, and technology. Consideration must be given to programs and projects that will get the information about best practices into the hands of practitioners, decision makers, and consumers, some of whom have little knowledge about the abilities or contributions of the polygraph. There are decision makers—such as military officers, police chiefs, judges and prosecutors, government program directors, government officials and others—who make decisions based upon polygraph examinations and subsequent rendered opinions, who have never been educated about the polygraph. They do not know that there are methods available, based on well-founded chosen policy decisions, that will better provide them with the information that they want to have, taking into the account sensitivity, specificity, error rates, and inconclusive rates that ESS offers. It could well be worthwhile for the profession to develop educational seminars to inform stakeholders about these considerations. This could be a particularly worthwhile endeavor for the American Polygraph Association and the American Association of Police Polygraphists. Given that these two associations already have networks with various stakeholders, such as the Department of Defense, the International Association of Chiefs of Police, the National Sheriffs Association, and their state counterparts, short programs could be developed to introduce these ideas at their respective conferences and meetings and provide follow-up through articles in their widely circulated professional publications. It is also important to reach rank-and-file personnel who are actually in the field and may be unaware of the current state of the polygraph and best practices. The author is aware of state law enforcement academies that address the polygraph in basic courses as well as continuing education courses. There are analogous educational undertakings in the military and other programs through which these short, informational classes could be offered.

In terms of the profession itself, there must be a major internal push to continue the research that has been undertaken in the last several years. We must keep our eye on the target, and that target cannot be misidentified. As the National Research Council suggested in its seminal report, the concern must be on national security and, by implication, homeland security and the war on terror. If research into lie detection and other social sciences identifies better methods, instrumentation, technologies, and techniques, then they must be further studied and embraced, if proven, even at the expense of letting go of what we know and what gives us comfort.

In terms of the present, careful consideration must be given as to how to keep current practitioners within the bounds of known best practices. Scientific and scholarly research and peer-reviewed articles are part of that equation; however, one must not lose sight of the polygraphist in the field whose primary concern is learning today what can be applied tomorrow. The science and research must be translated and presented in such a way that these individuals take an interest in it, understand it, and apply it.

Lastly, this study supports some of the findings of previous research into the Empirical Scoring System. It seems to support the position that ESS can complement the three-position TDA and the seven-position TDA systems and potentially others. This study did not find that ESS improved the scoring ability of the polygraphists. It does support the position that ESS offers the ability of polygraph consumers to choose their own tolerance for risk, something that is not readily available with other scoring systems.

This ability for the consumer to choose levels of risk when relying on the polygraph is important, but often not understood, and it can play a valuable role in homeland security and the war on terror.

No theory is going to be inviolate. Let me put it clearly. The only kind of theory that can be proposed and ever will be proposed that absolutely will remain inviolate for decades, certainly centuries, is a theory that is not testable. If a theory is at all testable, it will not remain unchanged. It has to change. All theories are wrong. One does not ask about theories, can I show that they are wrong or can I show that they are right, but rather one asks, how much of the empirical realm can it handle and how must it be modified and changed as it matures? (Chadee, 2011, quoting Leon Festinger, 1987).

FOOTNOTES

1. Dr. William Marston was a Harvard psychologist who is likely better known for the creation of the comic book character "Wonder Woman" under the nom de plume "Charles Moulton."

2. In a scholarly article on the history of the polygraph, Paul Tro-villo notes that Angelo Mosso, an Italian psychologist who studied under Cesare Lombroso, first experimented with a plethysmograph to study the effects of fear on human blood pressure. These experiments, as well as several that came later, were viewed as instrumental in the early study of the polygraph. In 1895, Lombroso, an Italian physician, psychiatrist, and criminologist, modified a medical instrument known as a hydrosphygmograph (similar to a modern cardiophymograph) to measure the blood pressure and pulse rate of a criminal suspect under police inte

3. As of February 3, 2012.

4. Formerly the United States Army Intelligence and Security Command (INSCOM).

5. Formerly the United States Army Intelligence and Security Command (INSCOM).

6. Formerly Immigration and Customs Enforcement (ICE).

7. It is of note that the National Academy of Sciences believed the Lee case so important to the government's reliance on the polygraph that it devoted an appendix to the case in its report (*The polygraph and lie detection*, 2003) and that eNotes, a popular research site for students and teachers, uses it as its case study for polygraph on its website (Lerner & Lerner, 2006).

8. This statement is based on personal interviews with primary sources who cannot be identified due to security concerns. These personal conversations have taken place during the 25-plus-year polygraph career of the author.

9. What does not seem to be at issue is that Lee illegally removed huge amounts of classified nuclear information from the laboratory, estimated at over 400,000 pages, and that once removed, its final destination(s) have never been learned (Shelby, 2001).

10. Due to the classification, sensitivity, and state civil-service rules, neither the name of the agency nor the minute details can be divulged.

11. This is the final report made to Congress, since Congress relieved the DoD of its reporting responsibilities after fiscal year 2002. No current figures are available as to the number of examinations currently conducted by the DoD. In 1991, Congress authorized the DoD to conduct no more than 5,000 CSP examinations annually. However, this quota was lifted in 2005, and there is currently no cap on CSP examinations. The figure of 8,512 includes those conducted by the DoD for non-DoD federal agencies. It is also noted that these numbers include only the DoD and not the National Security Agency or those conducted under the authority of the director of Central Intelligence.

12. A "symptomatic question" is a question used to identify whether or not an examinee is fearful that the polygraphists will ask an unreviewed question embracing an outside issue that is bothering the examinee. This mistrust of the examiner will putatively dampen the examinee's responses to other test questions. Symptomatic questions are widely used, though the trend in the research is that there is no meaningful effect (Krapohl & Sturm, 2002a)

13. No polygraph programs have been dismantled at the federal level, and new federal programs have, in fact, been added since the 2003 NRC report. However, legislative decisions and court rules have impacted or outlawed polygraph programs at the state and local levels.

14. The American Association of Police Polygraphists is the largest law enforcement polygraph association in the world. The author is both a past and current president.

15. The instructions for the rule are simple: if you can see it, point to it, and support that the reaction is bigger, then you score it. If you can't point to it and support it, then do not assign a score.

16. "Psychophysiological detection of deception" (PDD) is a term used primarily by the federal government and is interchangeable with the terms "polygraph" and "lie detector."

17. Special gratitude is expressed to Raymond Nelson for his computational assistance.

18. The pneumograph sensors, one tube placed around the abdomen and another around the thorax, record respiration data. Features included in the manual scoring model pertain primarily to suppression or reduction of respiration activity.

19. Changes in the electrical properties of the skin (exosomatic and endosomatic) typically measured by placement of electrodes on the central pad of skin of two fingers. This term superseded the term "galvanic skin response" (GSR), which can still occasionally be found in the older literature.

20. A term for recording heart activity, typically done by placement of a blood pressure cuff on an arm, which then measures pulse wave and changes in relative arterial blood pressure. In this context it is more correctly called sphygmograph or plethysmography (Krapohl & Sturm, 2002b).

21. No deception indicated, in layman's terms, means that it is the polygraphist's opinion that the person is truthful as to the matter at hand.

22. Deception indicated, in layman's terms, means that it is the polygraphist's opinion that the person is not truthful (lying) to the matter at hand.

23. Inconclusive, in layman's terms, means that the polygraphist has no opinion as to whether or not the person is truthful or lying to the matter at hand. It is typical that "no opinion" is rendered when the diagnostic quality of the tracings is such that they cannot be analyzed. It is the author's experience that those within the field of polygraph scoring do not consider "no opinion" as an error and that in many cases with subsequent testing (sometimes called a "reexamination") a definitive opinion can be rendered. However, it is duly noted that some outside the profession consider "no opinion" to be an error, and in research this dissenting opinion is sometimes taken into account.

24. Mr. Raymond Nelson assisted in the research question designs, as well as the computation of the tables and figures and the interpretation of the results and analysis.

25. As one might deduce, signal detection theory had its early beginnings with those researching radar. Its psychological roots began in the 1950s and were primarily led by John A. Swets (Herbert, 2010). See Herbert for an insightful article about Mr. Swets and signal detection theory.

26. The false detection of something that is not actually present. In polygraph it is the incorrect decision that deception was practiced by the examinee (Krapohl,& Sturm, 2002b).

27. The failure to detect the presence of a particular event or item. A false negative in polygraph refers to the incorrect decision that deception was not practiced by the examinee (Krapohl & Sturm, 2002b).

28. Polygraph has a long history of infighting within the profession, as the author can attest to from his 25-plus years in the field. This infighting tends to revolve around scientific research and its importance to the trade. There are those within the field who believe that the profession need not be concerned about what detractors say about the validity and reliability of the polygraph. This side of the house tends to argue that "we know it works" and its utility is incontrovertible. The other side of the house argues that, for the field to survive, the polygraph must continue to build a strong scientific foundation. That is, in order to continue to serve its important role in national security and law enforcement, it must prove its validity and reliability so that its worth can be proven to policy makers and legislators in contrast to the naysayers' claims. This thesis falls on this side of the argument

29. NCCA is the Department of Defense's polygraph school, which all federal polygraphists attend as part of their initial training. It was formerly known as the Department of Defense Polygraph Institute (DODPI) and later as the Defense Academy for Credibility Assessment (DACA).

30. Currently, the American Polygraph Association, the largest professional polygraph association in the world, accredits 16 schools in the United States and 13 international schools.

LIST OF REFERENCES

Alder, K. (2007). *The lie detectors: The history of an American obsession*. New York: Free Press.

Ball & Gillespie Polygraph. (n.d.). William Marston. Retrieved January 16, 2012, from http://www.lie2me.net/thepolygraphmuseum/id17.html

Blalock, B., Cushman, B., & Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph* 38(4), 281–88.

Chadee, D. (2011). *Theories in social psychology*. Malden, MA: Wiley-Blackwell. C.I.A..

(n.d.). *DHRA Web site | Home*. Retrieved October 11, 2011, from http://www.dhra.mil/perserec/espionagecases/cia.htm

Collins, C. (2004). OSI provides the truth in Baghdad. *Learn How to Pass (or Beat) a Polygraph Test* | AntiPolygraph.org. Retrieved January 17, 2012, from http://www.antipolygraph.org/articles/article-044.shtml

DACA. (2006). Test Data Analysis. *Psychological detection of deception analysis II— Course #503*. Lecture conducted from Defense Academy for Credibility Assessment, Ft. Jackson, South Carolina.

Department of Defense Polygraph Program: FY 2002 Report to Congress. (2003). *Federation of American Scientists*. Retrieved January 15, 2012, from http://www.fas.org//sgp/othergov/polygraph/dod-2002.html

Earley, P. (1997). *Confessions of a spy: The real story of Aldrich Ames*. New York: G.P. Putnam's Sons.

Handler, M., Nelson, R., Goodson, W., & Hicks, M. (2010). Empirical scoring system: A cross-cultural replication and extension study of manual scoring and decision policies. Pending publication.

Harvey, L. (2003). Detection sensitivity and response bias. Retrieved December 17, 2011, from psych.colorado.edu/~lharvey/p4165/p4165_2003_spring/2003_Spring_pdf/P4165_SDT.pdf

Herbert, W. (2010). John A. Swets: A signal idea, a singular life. *Observer* 23(7), 150–65. Retrieved January 15, 2012, from http://www.psychologicalscience.org/index.php/publications/observer/2010/septe mber-10/john-a-swets-a-signal-idea-a-singular-life.html

History. (n.d.). *Joe Klingensmith—Polygraph service*. Retrieved September 12, 2011, from http://polygraphservice.us/History.htm

Hoffman, D., & Stober, I. (2001). *A convenient spy: Wen Ho Lee and the politics of nuclear espionage*. New York: Simon & Schuster.

Keating, P. (2005). D-prime (signal detection) analysis. *UCLA Linguistics.* Retrieved January 17, 2012, from www.linguistics.ucla.edu/faciliti/facilities/statistics/dprime.htm

Kleiner, M. (2002). *Handbook of polygraph testing*. San Diego, CA: Academic Press. Krapohl, D. 2006). Validated Polygraph Techniques. Polygraph 35(3), 149–55.

Krapohl, D., & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph* 35(1), 55–63.

Krapohl, D., & Dollins, A. (2003). Relative efficacy of the Utah, Backster, and federal scoring rules: A preliminary investigation. *Polygraph* 32(3), 150–65.

Krapohl, D., Nelson, R., & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System Version 3 (OSS-3) and human scorers. *Polygraph* 37(3), 185–215.

Krapohl, D., Stern, A., & Bronkema, Y. (2009). Numerical evaluations and wise decisions. *Polygraph* 38(1), 57–69.

Krapohl, D., & Sturm, S. (2002a). Polygraph research: Value, shortcomings and perspectives. *Polygraph* 31(3), 157–230.

Krapohl, D., & Sturm, S. (2002b). Terminology reference for the science of psychophysiological detection of deception. *Polygraph* 31(3), 154–239.

Lerner, B., & Lerner, E. (2006). Polygraph, case histories— *eNotes.com. eNotes - Literature Study Guides, Lesson Plans, and More.* Retrieved October 10, 2011, from http://www.enotes.com/polygraph-case-histories-reference/polygraph-case- histories

Macmillan, N., & Creelman, C. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin & Review* 3(2), 164–70.

McNicol, D. (2005). *A primer of signal detection theory*. Mahwah, N.J.: L. Erlbaum Associates.

Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., et al. (2012). Using the Empirical Scoring System. *Police Polygraphist Digest* 4, 9–15.

Pentagon's intelligence arm steps up lie detecting efforts on employees. (2008, August 24). Retrieved February 22, 2012, from http://www.foxnews.com/story/0,2933,409502,00.html

Pincus, W. (1994, August 10). Ames says more polygraphs might have deterred him. *Washington Post*, p. 8.

The Polygraph and Lie Detection. (2003). Washington, D.C.: National Academies Press. Raskin, D. C., Barland, G., & Podlesny, J. A. (1978). *Validity and reliability of detection of deception*. Washington, D.C.:

National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Dept. of Justice.

Scientific validity of polygraph testing: A research review and evaluation. (1983). Washington, D.C.: Congress of the U.S., Office of Technology Assessment.

Senter, S. (2008). Polygraph research: Value, shortcomings and perspectives. *Polygraph* 37(4), 277–80.

Shelby, R. (2001). Intelligence and espionage in the 21st century. *Conservative Policy Research and Analysis*. Retrieved October 11, 2011, from http://www.heritage.org/research/reports/2001/05/intelligence-and-espionage-in- the-21st-century

Trovillo, Paul. (1972). The history of lie detection (part 1). *Polygraph* 1(2) 46–75.

Warner, W. (2005). Polygraph testing: a utilitarian tool. *F.B.I. Law Enforcement Bulletin* 74, 4.

Weaver, R. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph* 9(2), 94–108.

Weinberger, S. (2008). Handheld lie detector goes to war. *Wired.com*. Retrieved January 16, 2012, from http://www.wired.com/dangerroom/2008/04/handheld-lie-de/
Weiner, T. (1994, August 10). Polygraph detected lies by Ames, F.B.I. finds. *New York Times*, p. 6.

Wen Ho Lee Case Study. (2008). Retrieved August 22, 2011, from www.pherson.org/PDFFiles/WHLCaseStudy.pdf

Wen Ho Lee's Problematic polygraph. (2000, February 4). *CBS News*. Retrieved August 22, 2011, from http://www.cbsnews.com/stories/2000/02/04/national/main157220.shtml

1994-HARLEY DAVIDSON/Dyna Wide-glide

or choice of ***$10,000 Cash***

AMERICAN ASSOCIATION OF POLICE POLYGRAPHISTS, INC

William "Buddy" Sentner Scholarship Fund

Classic Motorcycle Raffle

Tickets are $20 each or 6 for $100

Drawing to be held June 5th, 2013
during AAPP Annual Business Meeting
NEED NOT BE PRESENT TO WIN

OR

OVER $30K INVESTED!

- CUSTOM "SHOW" BIKE RARELY DRIVEN 'LIKE NEW' ONLY 4,400 TOTAL MILES- EXCELLENT CONDITION
- CUSTOM "FLAME" PAINT W/ SILVER LEAF LOGOS
- CUSTOM LIGHTING BY ARLEN NESS AND OTHERS
- CUSTOM FRONT FORK SET UP BY ARLEN NESS
- CUSTOM WHEEL SPOKES BY LANDMARK
- CUSTOM MADE "FLAME STITCH" LOW-PROFILE LEATHER SEAT W/BACKREST
- CUSTOM CHROME HANDLEBARS
- CUSTOM 11.5" ROTORS AND CALIPERS BY ARLEN NESS

- HI-PERFORMANCE ENGINE:
  1. S & S 89" HI-PERFORMANCE STROKER KIT
  2. HI-PERFORMANCE H-D "MILLED" HEADS
  3. CRANE HI-PERFORMANCE CAM
  4. CRANE HI-PERFORMANCE IGNITION
  5. S & S SUPER "E" RACING CARBURETOR
  6. EAGLE ONE HI-PERFORMANCE VOLTAGE REGULATOR
  7. UNIQUE CUSTOM OIL COOLER KIT
  8. HI-PERFORMANCE EXHAUST
  9. PINGEL "HI-FLOW" CHROME FUEL COCK

...located in beautiful San Antonio Texas, near the airport, malls and historic locations.

We specialize in on site training for clients who have the need anywhere in the world!

Basic Examiner Course
Basic & Advanced PCSOT
Advanced Interview & Interrogation
Interdiction Officer Training

the Polygraph Institute

Texas TDLR, AAPP, and APA Approved Training

2013 Schedule

Basic Examiner

* Feb 4 - April 12
* Aug 5 - Oct 11

Interview Training

- **3 Day Advanced**
  - May 15 - 17
  - Nov 12 - 18
- **1 Day Interdiction**
  - January 11
  - April 19

JPCOT/PCSOT Course

For Polygraph Examiners, Therapists, & Supervision Officers

* April 8 - 12
* Nov. 12 - 18

Phone: 817- 290- 0033
Fax: 210- 481- 7639
E- mail:
info@thepolygraphinstitute.com
jpoburke@thepolygraphinstitute.com

- Our staff has participated in polygraph training for more than twenty five years. We retain experienced instructors that have all utilized and managed polygraph in law enforcement, government and private sectors. We delivered the only recognized polygraph curriculum in Iraq under war zone conditions for the Department of Defense and the Government of Iraq.
- Our PCSOT training incorporates multiple instructors, including experienced examiners, treatment provider therapists and supervision officers. Our advanced training will provide an understanding of Risk Assessment that will guide question selection and formulation. You will learn how to apply Risk Identification principles and not have to use or rely on some "*cookie cutter*" approach to every exam.
- Our 3 day advanced interview course includes cognitive interview techniques and behavioral profiling that facilitate individualized theme development with any subject. Our interview assessment tools can be performed quickly and simply to prepare an interview approach that offers higher degrees of success rates than other canned methods with resistant suspects.
- Our interdiction interview course teaches narrative interview style and collecting valid behavioral cues. We expose the myth of relying on outdated "*eye contact*" and "*nervous*" behavior cues typically being handed down in outdated training for patrol officers.

Register on line at: *www.thepolygraphinstitute.com*

We accept Visa, MasterCard, Discover and American Express.

Basic Examiner Course — $5500/Student (**Lodging included.** *Dbl occupancy*)

Interview/Interrogation Course - $425/Student—3 day / 1 day $125

PCSOT/JPCOT Training Course—$495/Student

AXCITON

Leading in applied research

- **Polygraph software easy to learn, Easy to use, robust & rugged.**
- **5 to 24 channel sensor box's available, all sensors included**
- **24 bit resolution**
- **Motion, Plathesmo, CAM, & other customizable sensors available**
- **Leading in on screen signal processing with such exclusive tools as white devagus, white standard, white balance, and many more.**
- **Free warranty, easy financing available, trade in credit available.**

Call Today Free at (800) 460-2645