The Police Polygraph Digest

January 2014

Table of Contents

Research & Information Chair: Mark D. Handler

Quality data acquisition begins with your instrumentation

contemporary Lemo® connectors ● medical grade compliance ● custom composite enclosure

The Paragon advantage

High resolution 24 bit data acquisition system.
Nickel plated brass medically approved Lemo connectors.
Lemo push-pull latching technology for a secure connection.
High-Retention USB requires 5 lbs force to disconnect.
Proven EDA technology that works when you need it.

Visit our video library to learn more
www.youtube.com/limestonetechinc

All-inclusive polygraph solutions for the professional examiner

Polygraph Professional Suite Silver Solution
Best instrument, best results, best value!

The Silver Solution is everything you need protected in a Pelican instrument case.

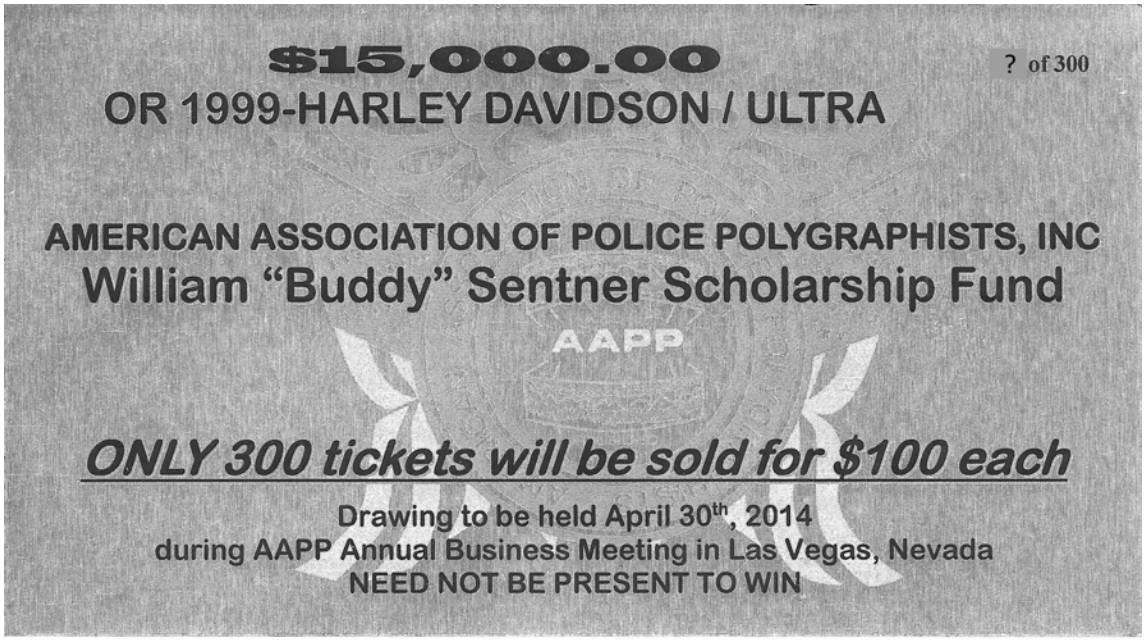
- ✓ Data acquisition system: 8 channel DataPac_USB or 9 channel Paragon
- ✓ Polygraph Professional Suite software license
- ✓ 2 pneumatic respiration transducers
- ✓ 1 EDA lead, 1 set of 24k gold plated electrodes, 1 set of snap ends, 1 package of 100 disposable Ag/AgCl wet-gel electrodes
- ✓ 1 adjustable blood pressure cuff, 1 FingerCuff, cardio tubing and Riester sphygmomanometer
- ✓ 1 StingRaySE Piezo electronic CM sensor
- ✓ OSS and Relative Response Magnitude (RRM) scoring algorithms included
- ✓ HARM psychometric pre-employment screening instrument included
- ✓ Printed and bound user manual
- ✓ Pelican 1450 instrument case
- ✓ Lifetime technical support
- ✓ 3 year total care warranty

Discounts available.

Contact us today
for a competitive quote.

Limestone
TECHNOLOGIES INC.

www.limestonetech.com 866.765.9770 sales@limestonetech.com



DO YOU HAVE YOURS???

Only 300 tickets will be sold this year for the AAPP Scholarship Raffle.

You could win your choice of a 1999 HD Ultra Motorcycle Or $15,000 cash.

But you can only get your tickets from your Regional Director, and *once there gone, there gone*.

Contact your Regional Director soon to secure your ticket!!!

*** 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.

THE IMPORTANCE OF ATTENDING THE AAPP'S 2014 SEMINAR IN HENDERSON, NV

The American Association of Police Polygraphists (AAPP) offers the latest in training and research, something which benefits every examiner. In the last five years, particularly in the last two (with the publication of validated techniques and recommended best practices), there has been enormous transformation in the polygraph profession. Because of research and science, polygraph is undergoing regular and improved change. This seminar addresses virtually every validated technique; the science behind each; and will assist examiners in choosing what will work best for them. There will be information on current research and where the field is headed in areas from screening to false confessions. Classroom presentations will cover topics that haven't been taught in a number of years and especially not under the same roof. Attendance at this seminar will help ensure that polygraphists and their agencies are utilizing the best practices and remain current with professional standards. If departments want to help protect against litigation, ensure that their members meet professional benchmarks, and keep their polygraph units current, attendance at the seminar becomes a simple necessity.

Reminder: Henderson is not in Las Vegas, it is more of a destination location for training with few distractions. The seminar will run five days with training (8) eight hours a day and a minimum of three sessions at all times.

The 2014 Course Outline/Schedule is posted on line at www.policepolygraph.org.

Hope to see you in Henderson in 2014!

WHY MEMBERSHIP IN THE AAPP IS IMPORTANT AND THE BENEFITS OF MEMBERSHIP

The AAPP is the largest law enforcement/criminal justice polygraph association in the world. Membership in the Association helps ensure professionalism and credibility. Membership, along with ever-important continuing education, allows the member to apply for certification. These things become crucial in court settings and can assist in protecting members and their agencies that might face litigation. Three times each year, members receive The Police Polygraphist journal, which is full of polygraph happenings (unique to each region) as well as research and news. Annually, the AAPP publishes The Police Polygraph Digest which contains the latest research and advice to assist members in making wise choices about their profession. The AAPP maintains a website which contains a forum and is a networking tool for those in the profession. Further, the AAPP assists members and agencies who have individual questions or concerns and seek information or advice. Our Director of Quality Control and our Research and Information Chair, assure individual members and their agencies access to assistance and information at critical times; a benefit that is not available to examiners not availing themselves of membership in a professional association. Thus, value of membership in the AAPP and attendance at the annual seminars far exceeds the cost.

AAPP 2014 Seminar (updated 12/3/13)

Co-Sponsored by the California Association of Polygraph Examiners

Hilton Lake Las Vegas
Henderson, Nevada
April 27 - May 2, 2014

CLASS/EVENT SCHEDULE

SUNDAY, APRIL 27, 2014

TIME	*EVENT*	*Room*
	Golf Tournament	*Details to be Announced*
5:15p – 5:45p	Worship Service—Barry Cushman AAPP Chaplain	Chapel La Capella di Amore
6:15p – 8:00p	President's Reception * Bring your registration drink tickets!*	Bridge

MONDAY, APRIL 28. 2014

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
6:15a – 7:30a (Board Members 6:00a)	Registration	Julie & Mark Gerspacher *National Office Manager* *Crew: AAPP BOD*	
8:00a – 9:00a	Opening Ceremony	Opening Prayer Honor Guard National Anthem Pledge of Allegiance President's Welcome Housekeeping Rules Special Announcements	Salons 1, 2, 3 & 4
9:00a – 10:00a	*Spouse Breakfast*	*Breakfast for Member Spouses and Guests*	*Tuscany or outside*
9:00a – 12:00p	TBA	TBA	Salons 1, 2, 3 & 4
12:00p – 1:00p	*Lunch- on your own*		
1:00p - 5:00p	Countermeasures	Raymond Nelson	Salon 1
1:00p - 5:00p	Future of Polygraph	Dr. Jennifer Vendemia	Salon 2

1:00p - 5:00p	**Question Formulation**	**Elmer Criswell**	**Salon 3&4**
1:00p-5:00p	**Federal CM (Meeting/Training)**	**Stu Senter**	**MonteLago 1&2**
2:00p – 5:00 p	***Spouse Wine Tasting***		**TBA**
Contact Elmer in the Seminar Area	**QC Your Charts One on one by appointment**	**Elmer Criswell**	**Village Boardroom**
5:00p – 6:00p	**Meet Your Regional Director**	**Regional Directors, I - V**	**Monte Vista, Deserto, Lago, Vineyard, Tuscany**
6:00 p – 9:00p	**Social Gathering Location TBA**	**General Membership**	

TUESDAY, APRIL 29, 2014

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00a	**Detecting Deception in Oral and Written Narratives Using Psychological Narrative Analysis (PNA)**	**John R. Schafer, Ph.D.**	**Salon 1**
8:00a – 12:00a	**Pre and Post Test Integration**	**Chip Morgan**	**Salon 2**
8:00a – 12:00p	**Test Data Analysis**	**Elmer Criswell Mark Handler**	**Salon 3**
8:00a – 12:00p	**Lafayette**	**TBA**	**Monte Lago 1&2**
8:00a – 12:00p	**Stoelting**	**TBA**	**Tuscany**
12:00p – 1:00p	***Lunch- on your own***		
12:00p – 1:30p	***State & National Leadership Luncheon*** ***Invitation Only***	**Jim Wardwell** ***AAPP- Vice President State Association Leadership***	**Tuscany Courtyard**
1:00p – 5:00p	**Detecting Deception in Oral and Written Narratives Using Psychological Narrative Analysis (PNA)**	**John R. Schafer, Ph.D.**	**Salon 1**
1:00p - 5:00p	**Making the Move to DLST "Directed Lie for Pre-Employment"**	**Chip Morgan**	**Salon 2**
1:00p – 5:00p	**Empirical Scoring System "ESS"**	**Pam Shaw**	**Salon 3**

1:00p-5:00p	DIA (Meeting)	Brett Stern	Salon 4
2:00p – 5:00p	QC Your Charts One on one by appointment	Elmer Criswell	Village Boardroom
1:00p – 5:00p	Lafayette	TBA	Monte Lago 1&2
1:00p – 5:00p	Stoelting	TBA	Tuscany
6:00p	Tuesday Night Function	General Membership	Village

WEDNESDAY, APRIL 30, 2014

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a-12:00p	Federal Techniques: ZCT, Bi-Zone, AFMGQT v1 & v2	Matt Hicks	Salon 1
8:00a-12:00p	Nonverbal Communication	Mike Liwicki	Salon 2
8:00a-12:00p	Utah Approach/Scoring System	Pam Shaw	Salon 3
8:00a-12:00p	DIA (Meeting)	Brett Stern	Salon 4
8:00a-12:00p	FBI (Meeting)	Kevin Day	Monte Lago 1&2
12:00p – 1:00p	*Lunch- On your own*		
1:00p - 5:00p	The Directed Lie Screening Technique and What it Has Taught Us	Walt Goodson	Salon 1
1:00p – 5:00p	The Psychology of Interrogation: State of the Science, Credibility, Veracity Judgements, Detecting deception & False Confessions	Maria Hartwig	Salon 2
1:00p – 5:00p	Examinee Suitability	Raymond Nelson	Salon 3
1:00p-5:00p	DIA (Meeting)	Brett Stern	Salon 4
1:00p-5:00p	FBI (Meeting)	Kevin Day	Monte Lago 1&2
3:00p	School Director's Meeting	Karen Clark	TBA
1:00p – 3:00p	QC Your Charts One on one by appointment	Elmer Criswell	Village Boardroom
5:00p-7:00p	Annual Business Meeting	General Membership	Salon 1

THURSDAY, MAY 1, 2014

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00p	Law Enforcement Employment Screening/Computer Evaluations	Charles Honts, Ph.D.	Salon 1
8:00a – 12:00p	Police Polygraph Interrogation	John W. Kaster	Salon 2
8:00a – 12:00p	Back to the Basics in Polygraph	Greg Adams	Salon 3&4
8:00a – 12:00p	Limestone	TBA	Monte Lago 1&2
8:00a – 1200p	Axciton	Bruce White	Tuscany
12:00p – 1:00p	*LUNCH – on your own*		
1:00p – 5:00p	Presentation by Mexican Delegation	TBA	Salon 1
1:00p – 5:00p	Polygraph Procedures for Diverse Hispanic/Latino and Islamic Cultures	Dr. Antonio v. Suarez-Barrio	Salon 2
1:00p – 5:00p	Concealed Information Test (CIT)	James McCloughan	Salon 3&4
1:00p – 5:00p	Limestone	TBA	Monte Lago 1&2
1:00p – 5:00p	Axciton	Bruce White	Tuscany
Contact Elmer in the Seminar Area	QC Your Charts One on One by Appointment	Elmer Criswell	Village Boardroom
1:00p – 5:00p	California Association of Polygraph Examiners Business Meeting	Ted Todd	Vineyard
6:00p – 7:00p	*Cocktail Hour/Awards*	*General Membership*	Courtyard / MonteLago 1&2
7:00p – 9:00p	*Annual Banquet*	*General Membership*	Courtyard / MonteLago 1&2

FRIDAY, MAY 2, 2014

TIME	*EVENT*	*INSTRUCTOR*	*ROOM*
8:00a – 12:00p	**Looking Behind the Mask During the Polygraph: Identifying and Interviewing the Psychopath-His behavior, affect and view of the crime**	**Mary Ellen O'toole**	**Salon 1**
8:00a – 12:00p	**Advanced PSCOT**	**Ray Nelson**	**Salon 2**
8:00a -12:00p	**Middle Eastern Cross Cultural Training**	**Nia Ackvan**	**Salon 3&4**
Contact Elmer in the Seminar Area	**QC Your Charts** One on One by Appointment	**Elmer Criswell**	**Village Boardroom**
12:00p – 1:00p	***LUNCH – on your own***		
1:00p – 5:00p	**Looking Behind the Mask During the Polygraph: Identifying and Interviewing the Psychopath-His behavior, affect and view of the crime**	**Mary Ellen O'toole**	**Salon 1**
1:00p – 5:00p	**Advanced PSCOT**	**Raymond Nelson**	**Salon 2**
1:00p – 5:00p	**Middle Eastern Cross Cultural Training**	**Nia Ackvan**	**Salon 3&4**

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 Henderson, NV in 2014

Executive Board: Piazza Boardroom

Agency meeting space: Tuscany, Estate Boardroom, Village Boardroom, Hospitality room

Revised 12/3/13

FORWARD TO THE JANUARY POLYGRAPH DIGEST

Mark Handler and Raymond Nelson

A unique characteristic of the American Association of Police Polygraphists (AAPP) is that its members often use the polygraph in a *forensic* setting. The term forensic refers to scientific processes and activities in legal settings and proceedings. We believe diagnostic polygraphs conducted in criminal investigations are forensic activities. Screening polygraphs are not likely to be forensic examinations because they are not normally conducted in response to a specific crime incident. Additionally, screening examination techniques have less scientific support than diagnostic techniques. Although diagnostic and screening examinations are conducted with different objectives, their differences are not always appreciated. Examiners and their clients may not grasp the importance of various decision theoretical priorities for test sensitivity and specificity or the effects of false-positive or false-negative errors.

Despite this difference, portions of diagnostic and screening polygraphs can be introduced as evidence in legal proceedings. Legal proceedings include criminal trials and hearings, civil trials or lawsuits, administrative hearings, arbitration hearings or any process of appearing before a court of law where a decision is made about an argument or claim. Legal debates may ensue over parts of *any* diagnostic or screening PDD examination.

Polygraph test results are usually not admissible in most criminal cases, but they can focus the investigation. Our referring professionals, agencies, legislators, and communities all expect scientific and professional integrity in both testing contexts. Most examiners work for someone, who usually works for (and answers to) someone else. That person to whom the examiner must answer, may ask him or her to explain and justify the examination, including field practices and results. Having knowledge of *evidenced-based practices* (EBP) can help examiners defend of their work.

EBP emerged around 1992 in the form of evidenced-based medicine, and its ideas quickly pervaded many other disciplines including nursing, dentistry, psychology, library sciences, information sciences, education, and forensics. The basis of EBP is that professional practices are: 1) based on research studies that measure a level of effectiveness and 2) that these research studies meet some particular standards. EBP uses data collected through scientific research for making decisions. It does not attempt to support the test results and professional conclusions using expert opinion without evidence. It does not rely on 'gut responses' alone, dogma, or anecdotal case studies without evidence of generalizability. And, EBP is not the mere reliance on research, but the reliance on research that informs professionals and the public of known and predictable level of effectiveness. A germane example of EBP is the 2012 American Polygraph Association decision that all examinations should be conducted using validated techniques meeting minimum standards.

Were a judge, prosecutor, treatment provider, defense attorney, supervision officer, administrator or supervisor to ask you to justify, explain or support your examination choice or practice how would you respond? Use of EBP would mean that the evidence has already been examined, published, and is reasonably well accepted and understood. Could you rely on EBP? Or would you be forced to provide a dogmatic response without evidence?

Fledgling professions rely on dogma initially until they build a body of supportive evidence. Once that body of evidence exists, however, continued reliance on dogmatic practices is unwise and untenable. The PDD profession cannot rely on dogma and expect to be taken seriously by courts and related fields of science. We leave you to consider what Joseph P. Bono, President of the American Academy of Forensic Sciences, wrote in his President's Message in: Academy News, American Academy of Forensic Sciences, 2010; 40(5):3.

> "The six most questionable words used to formulate the justification for a conclusion by any forensic analyst are 'BASED ON MY TRAINING AND EXPERIENCE…' Training and experience in the absence of demonstrative evidence mean little to me. A reputable examiner should be able to show the decision makers — the prosecutor, the defense attorney, the judge and the jury — the basis for a conclusion which is understandable and can be justified by data or images. If the examiner resorts [only] to the 'trust me, I know what I am doing logic,' a red flag should i mmediately go up: DON'T TRUST HIM!"

Of themselves, there is nothing inherently wrong with the words "based on my training and experience." The problem arises when these words are *the only* basis for

asking others to accept our work and results. An EBP explanation will trump a dogmatic defense every time. Fortunately, the PDD profession has attracted the attention and efforts of scientific thinkers throughout its existence. Our work is supported by a growing and formidable basis of scientific evidence at the theoretical and pragmatic level.

Many of those reading this forward know how lonely and uncomfortable it is being on the witness stand or on the "hot seat" in front of a superiors. Being inadequately prepared for the questions you encounter is stressful, and using dogma to support your answers is increasingly unlikely to make your evidence persuasive. Being knowledgeable and conversant with EBP will assuage some of that anxiety, and will better support your decisions. EBP will ultimately make each of us more valuable and effective witnesses and professionals.

Evidence is the core principle of EBP. It makes or breaks a case, persuades or influences a decision maker and can be the tangible tool that triers of the fact use to decide case outcomes. Evidence resulting from a diagnostic PDD examination should fall under the subcategory of *scientific evidence.* Scientific evidence is a type of evidence that either supports or refutes a scientific hypothesis and follows the scientific method.

The validity of scientific evidence is argued using research. The strength of the scientific evidence is that research is measured using statistical analysis, the language of empiricism. PDD examiners who have a basic understanding of the ideas and vocabulary of scientific testing and normative data will inevitably be perceived as more competent. The evidence they present will be regarded as more credible than those who have not obtained training and education in these areas.

Law-enforcement examiners are also aware of the problems associated with unreliable evidence. All evidence may be subject to criticism and attacked during a legal proceeding, whether reliable or not. Prevailing during those proceedings is often a matter of presenting a well-organized basis of information to support the evidence. In those rare circumstances when it becomes necessary to present polygraph evidence in a legal proceeding, using validated techniques will ease the task of locating and organizing the available evidence. The professional examiner and expert witness will be prepared to answer scientific questions.

We disagree with arguments that field examiners are not scientists or experts and are therefore not required to testify about the scientific basis of the test results. We hold the conservative view that field PDD examiners should be prepared to answer questions as experts. Examiners not prepared to answer scientific questions are at risk of losing credibility. And the evidence they present is at increased vulnerability to criticisms of unreliability. Although the primary concern in any legal proceeding are the rights of the individual accused and other members of the community, there is an important secondary concern to the polygraph profession. Successful criticisms may affect the usefulness of the test with future cases. We caution that any examination and any PDD examiner can be subjected to scrutiny that reflects on the professionalism of the individual examiner and the profession. We also caution against any impulse to suggest that results and accuracy of some types of polygraph tests are unimportant (e.g. screening examinations.) Such claims would be concerning to the communities and agencies we serve, and such a position would be of questionable usefulness to the profession.

The goal of this *Digest* is to better prepare you to defend your decisions. In this edition you will find a collection of articles for the "nuts and bolts" field examiners who ply their trade thousands of times a week in the search for the truth. The backdrop for the selection of this focus is the present activity in the United States Congress looking at the formation of an oversight agency or commission to better regulate the forensic sciences. PDD should be included in that effort and our exclusion will increase resistance to PDD testing. This can result in increased criticisms that PDD testing amounts to pseudoscience, not worthy of funding or administrative support. We oppose that view. We prefer to continue to educate ourselves, our legislators, the community, and professionals in related scientific and forensic disciplines about the validity and scientific basis for PDD testing. The most concerning alternative to EBP – professional practices not supported by scientific evidence – amount not only to reliance on dogma instead of evidence, but also the practice of experimenting on members of the public.

For those of you who consider your work *forensic*, this journal is for you. We have written and selected materials to help you understand the basic ideas that underlie an EBP approach to diagnostic and screening PDD testing. We refresh your knowledge of scientific terms and offer a primer on using normative data to describe error estimates in diagnostic settings. We know many forensic PDD examiners are experienced professionals who have learned the power of clear procedures to make order out of chaos. This helps protect us from uncertainty, doubt, and vulnerability during times of difficulty. Those experienced professionals are also sometimes a few years distant from their academic education and may have forgotten some mathematical and scientific concepts. This material will be new to some and a refresher to others. Coming to terms with EBP is a matter of learning and remembering a core set of ideas, vocabulary and procedures that will keep us anchored to the evidence.

Embracing EBP will help us in the "nuts and bolts" activities of interviewing, investigating, and reporting information and test results to the referring professionals, agencies and communities we serve. Some of the information here is procedural, and some is instructive information to be used as reference material. Each is highly useful, though in different ways. Whether the information is procedural or theoretical, we have attempted at every point to describe the relevance of the material to the PDD examination. All who conduct examinations in the field are practical-minded people who will continue to rely on tangible and procedural standards of practice. Those standards will be most useful when they are based on evidence. Our sincere belief is that becoming familiar with EBP will not only better prepare us to defend our work, but it will also make us better examiners.

A Terminology Primer for Scientific Testing, Research, and Evidence-Based Polygraph

Raymond Nelson and Mark Handler

Alpha: Greek letter alpha ( ), denoting the beginning. Referred to with either the upper case A, lower case a, or Greek . In science and statistics, alpha refers to the boundary between statistically significant and non-significant results. Alpha is used to describe the proportion or area under the curve that is separated from the whole by the alpha boundary. With the understanding that there is no such thing as a perfect test, and that testing errors may occur, alpha boundaries are used to express a tolerance for error or desired confidence level. Alpha is therefore directly related to the probability of error. Alpha can be set at any level but is often set at the .05 level, corresponding to a desired 95% confidence level. Alpha = .10 may increase the sensitivity or specificity rate and decrease the rate of inconclusive results, at a cost of increased errors. More conservative alpha levels are justified in some circumstances, in which case alpha is often set at .01. In practical PDD terms, alpha is a numerical cut-score that can be determined using normative parameters to calculate the associated probability of error. Known statistical phenomena occur in PDD and other testing context, and these affect the resulting probability of error when making multiple statistical decisions within a single test.

- **Inflated alpha:** a mathematical increase in the probability of error resulting from multiple statistical decisions within a single experiment or test. Every decision has an associated probability of error, and these probabilities are cumulative within a test or experiment when making multiple probabilistic decisions. Inflation of alpha results in an unintended increase in type I (i.e., false positive) errors and decreased accuracy. This becomes a problem in that the actual observed error rate can substantially exceed the desired alpha level or tolerance for error. In PDD testing this can occur when using subtotal scores in the context of an event-specific (i.e., single issue exam) for which the criterion states of individual test questions do not vary independently. Inflation of alpha can be managed using the Bonferroni correction. Unintended inflation of alpha can occur when using uncorrected subtotal scores of event-specific PDD examinations.
- **Deflated alpha:** a mathematical decrease in the probability of error resulting from the requirement for multiple statistically significant results in order to achieve a negative test result. Deflation of alpha is not associated with increased type I (i.e., false positive) errors, but is associated with increased inconclusive result and decreased test specificity (i.e., increased type II errors). Deflation of alpha can be managed by correcting the desired alpha level with the inverse of the Šidák correction. Unintended deflation of alpha occur when using uncorrected subtotal scores of multiple issue PDD examination.
- **Corrected alpha:** statistical methods for correcting the potential for distortion of alpha when making multiple decisions of statistical significance within a single test or experiment. Correction of alpha will ensure that observed error rates conform to the desired alpha level. Different types of corrections are used in event-specific and multiple issue PDD exams.
  - **Bonferroni correction:** a method for correcting the inflation of alpha, and corresponding increase in type 1 errors, that results from making multiple decisions of statistical significance. Bonferroni correction can be used to correct the alpha level, and reduce the potential for increased type I errors, when using the subtotal scores of event-specific PDD exams.
  - **Šidák correction:** another method for correcting the inflation of alpha, and corresponding increase in type 1 errors, that results from making

multiple decisions of statistical significance. Šidák correction is preferred over Bonferroni when multiple decisions are regarded as independent or not influenced by each other within a single test or experiment. The inverse of the Šidák correction can be used to correct the alpha level, and reduce the potential for type II errors, when using the subtotal scores of multiple issue PDD exams.

Analog study: any type of study that attempts to replicate or simulate real-world conditions under controlled circumstances. In PDD research the term has been used to describe quasi-experimental field studies that are not controlled experiments. They are usually employed in cases where the experimenter tries to gain greater experimental control over the independent variable.

Analysis of variance (ANOVA): refers to a number of statistical methods used to analyze differences between two or more groups. ANOVA methods can be used in PDD research.

Anecdote (case anecdote or anecdotal evidence): a short informal account of a case or circumstance, presented as based on real events. Unlike scientific evidence, anecdotal evidence cannot be derived or studied using the scientific method. Anecdotal evidence is generally regarded as unreliable and based on atypical circumstances because scientific evidence is required to make inferences about how common or typical an observation or event may be. Because anecdotes and anecdotal evidence may be novel or interesting it may be more memorable than a typical examples, and there is an associated risk of being misled by them. Anecdotal information is useful in the PDD context for teaching or illustrating existing knowledge, for hypothesis building, and to falsify a hypothesis. Anecdotes cannot be used as evidence or proof of the validity of a hypothesis and cannot be used to show causality. Anecdotal evidence is sometimes criticized as pseudoscientific when it is not falsifiable (e.g., alien encounters) or leads to flawed reasoning such as false or hasty generalization.

Apophenia: in psychology this refers to the human tendency to see or interpret patterns where none exist (e.g., the face of the "man in the moon" or the former "Old Man of the Mountain"). Apophenia can result in inaccurate decisions and incorrect conclusions, (i.e., deception based on incorrect subjective behavior patterns observed during a PDD test). Quantitative measurement and statistical analysis is intended to reduce the potential for error that may be related to errors of pattern attribution in PDD testing.

Area under the curve (AUC): describes the proportional area of a data plot or graphic that is covered by a defined region. AUC is a measure of discrimination, which is the ability of a test to make correct classifications. AUC is discussed in PDD research with regard to the normal distribution and also the receiver operating characteristic (ROC) curve.

Asymptotic (asymptote): a principle in mathematics in which parallel lines or curves proceeding into the distance will converge to a perceived distance of zero. The lines may not actually converge or cross, but a point on one line can be said to be the asymptote of the other. Asymptotes are a useful concept in PDD research because, together with the law of large numbers and the central limit theorem, they allow us to understand sample data as indicative of what to expect from the population.

Availability heuristic: an erroneous assumption that occurs when people attempt to reach conclusions about the probability of an event based on how easy it is to think of examples. It is based on the belief that if something can be recalled then it must be important (e.g., the easier it is to recall the consequences of an action, the greater the consequence may be perceived to be). The availability heuristic can lead to erroneous assumptions about PDD response features that may be related to certain population or referral groups. Examples include problems with data quality from groups for which a large proportion of people have chronic medical or psychiatric conditions. Questions may arise about the generalizability of those people's data to the sample data from which normative data were developed.

Axiom: generally accepted ideas that are regarded as self-evident but actually have no proof. Axioms are often mathematical, but also have logical implications (e.g., there is no such thing as a perfect test). Axioms are said to be consistent if they lack contradiction or contradictory evidence. Axioms are useful to the PDD testing context to establish a foundation of expectations that can lead to falsifiable hypothesis and quantitative analysis.

Base rate / prior probability: refers to our knowledge regarding the likelihood of guilt or involvement in a behavioral concern prior to conducting any testing. Ideally, our knowledge of prior probabilities will be informed by published epidemiological statistics. In practical terms, prior probabilities are based on data from past observation and experience of other similar circumstances (i.e., what proportion of people in a given circumstance are involved in the problem or concern under investigation). Base rates are important to the PDD context because they can be used with information on test sensitivity and specificity, to estimate the likelihood of a correct or incorrect test result. Test accuracy estimates involving base rates use Bayesian inference, or conditional probabilities, and are non-resistant to base rate differences (i.e., test accuracy is different or unknown when the base rate is different or unknown). In contrast, probability models

based on frequentist inference and are resistant to base rate differences (i.e., test accuracy statistics do not change when the base rate is different or unknown).

Bayesian inference: a method of statistical inference in which the probability of the hypothesis is determined based on available sample data which is regarded as a static representation of the population data or real world. Bayesian inference is used in prediction and forecasting models, epidemiology, economics, and other statistical paradigms. Bayesian inference depends on the use of Bayes´ Theorem and is non-resistant to differences or changes in base rates. For this reason, Bayesian inference can give unexpected or unreliable results when the population or real world are not represented by the base rates expressed by available sample data. Bayesian inference is useful in both PDD research and in algorithmic model that calculate statistical classifiers for individual exams.

Between-subjects design (between-group design): an experimental method in which two or more groups are tested simultaneously. A common between-subjects design involves an experimental group and a control group, for which the difference in results is evaluated for statistical significance. Between-group designs are used in epidemiological, psychological, medical, economical, and other types of sociological and physical research in addition to PDD research.

Bootstrapping: computer intensive statistical methods that are useful to calculate statistics that would be computationally difficult via other means. Bootstrapping can be used in PDD algorithm development, and to calculate the confidence ranges of test accuracy statistics.

Case study: a type of research strategy involving a single person, case, group, or event. Case studies may be retrospective or prospective, and are used to identify possible issues of causality and underlying principles. However, case studies cannot themselves support conclusions or inferences about causality. Case studies are often holistic studies, and can include both qualitative and quantitative information. Case selection for case study research often involves atypical cases such as outliers or localized observations that clarify issues of history or causation or illustrate interesting or unusual circumstances or outcomes. Case studies may be evaluative or exploratory, for the purpose of theory testing or theory building. Case studies may also be illustrative. Generalization from case studies is a difficult activity, and takes the form of falsification: if one observation or example can be shown to contract a theory or hypothesis then the theory or hypothesis must be revised or rejected (e.g., "all swans are white," has been shown to be false at least one time, they have found black colored swans in the world). PDD case studies can be used to illustrate, falsify or develop a hypothesis, but cannot be used as evidence to prove a hypothesis.

Central limit theorem (CLT): states that the means of a large number or random samples can be expected to differ slightly, and that the mean of the sampling means will take the form of a bell-shaped normal distribution around the population mean. The central limit theory, together with the law of large numbers, holds that as the number of sampling distribution grows larger the mean will converge with the expected population value that would be obtained if it were possible to test the entire population. This can be useful in PDD research because we can expect that the result of different studies, though imperfect representations of the population, can be combined – if randomly selected – to improve our knowledge of population parameters.

Chi-squared test (X^2 test): any statistical hypothesis test in which the sampling distribution of the test statistic conforms to the chi-squared distribution when the null-hypothesis is true, including asymptotic distributions. Chi-squared tests can be used to test hypotheses regarding the frequency distribution of mutually exclusive events such as categorical test results. Chi-squared tests can be used in PDD research.

Classification: a scientific activity involving statistical prediction of group membership based on measurable features, characteristics, or symptoms. Group membership in PDD testing usually refers to the group of deceptive or truthful people in the population.

Coefficient of determination: a statistic that describes the proportion of total variation in an observed dataset that is explained by a model. The coefficient of determination, also commonly known as R-squared, is used as a guideline to measure the accuracy of the model, by providing a measure of how well the outcome is predicted by the model. In PDD testing the coefficient of determination can be used to describe the sampling proportion of guilty and innocent persons that is explained by the data. This is a more useful statistic than the simple proportion of correct scores or decisions because it is possible that some correct scores or decisions occur due to chance alone.

Cognitive bias: a terms used to describe irrational ways of thinking that may affect the accuracy of conclusions and judgments. Cognitive biases may include social or attribution biases that affect social interactions or probability biases that affect decisions and judgments. The scientific method and the principles of statistical analysis are intended to minimize the influence that cognitive bias has on our conclusions. Cognitive bias can take many forms, including familiarity, recency or primacy, anchoring, confirmation, and other forms. Awareness of the potential for cognitive bias may be useful to both PDD researchers and examiners who wish to avoid these common errors.

Conditional probability: refers to the probability of an

event or condition to occur when another event or condition is true. Probabilities are said to be independent if one has no effect on the other (e.g., if a coin is flipped two times the first outcome has no effect on the second). However, probabilities are aggregated when multiple events are required or expected. Conditional probabilities are useful in PDD research and test data analysis, and are central to Bayesian inference which evaluates the first event as evidence for the second (conditioned) event (e.g., the probability of being deceptive, given a significant response outcome on a PDD examination).

Confidence interval: normally refers to the range of potential bias or inaccuracy of a sample statistic such as a mean (e.g., confidence interval for the mean) using the range within +/-2 standard errors to define the confidence interval. A confidence interval describes the interval or range in which we are 95% confident the actual population value exists. The term *confidence interval* is also sometimes used to refer to the 95% confidence range or normal range. Confidence intervals are important in the PDD context because the tell use the upper limit and lower limit of the range of expected sensitivity, specificity, error, accuracy and inconclusive results.

Confidence level: a simplistic way of describing probabilistic test results in informal contexts. The *confidence level* (CL) is calculated as the inverse of the error statistic (e.g., if $p = .05$ then CL = 95%). In contrast, scientific testing emphasizes discussion of the p-value or probability of error that is used to make a categorical inference when the probability of error is sufficiently low. This is because science is often concerned with the likelihood of erroneous or random test results. Probability values are commonly expressed in decimal values, while confidence levels are often expressed in percentages. PDD examination results can be described categorically, based on either the level of significance (probability of error) or the confidence level.

Confirmation bias: refers to the tendency to search for information that confirms, and ignore information that refutes or disputes, a prior belief or conclusion. Also refers to the tendency to interpret ambiguous evidence as confirmatory. A number of identifiable bias pattern phenomena are related to confirmation bias, including the tendency to gather and remember information selectively. Confirmation bias can contribute to both overconfidence and the tendency to remain fixed (anchored) to a conclusion even when faced with contradictory evidence. Confirmation bias can lead to problematic conclusions and decision-making and knowledge of the potential for confirmation bias may be useful to PDD researchers and examiners.

Correlation (correlation coefficient): a statistic that describes the strength and direction of association between two variables. Correlation coefficients vary continuously from -1 to 0 to +1. Correlation coefficients are used in PDD research to identify physiological response features to score that can be combined to achieve a high level of test accuracy.

Covariance: a statistic that describes how well two variables change together. It is similar to the correlation coefficient but not bounded by +/-1. In general variables that are perfectly covariant or highly covariant may not be desirable within a single model because they may be redundant. Covariance is important in PDD research in the study of physiological response features that are correlated with the criterion of interest (e.g., deceptive or truthful state) while also not redundant with other response features.

Criterion variance: differences in the criterion state among the cases in a research sample or among the test questions of multiple issue exams. Criterion states can be described using the value-neutral terms *positive* and *negative*, and these terms can also be used to describe the test result. PDD criterion states are usually described using the terms *deceptive* or *truthful*. Criterion variance (i.e., differences in criterion state) of the individual questions in PDD event-specific diagnostic exams is assumed to be non-independent, while the criterion variance of multi-issue PDD screening examinations has historically been assumed to be independent.

Decision theory: a scientific discipline involving the identification of uncertainty, risk, and benefit of outcome probabilities in any context in which there may be conflicting interests at play. Decision theory is used in game theory, economics, psychology, forensics, politics, community planning and safety. In PDD testing, decision theory relates to the selection of alpha boundaries and decision rules that optimize risk assessment, risk management and program objectives.

Dependent variable: refers to the output, effect or result of a scientific test, experiment or investigation. In PDD research contexts the *in*dependent variable is the structure of the test stimuli, administration and analysis methods, while the dependent variable is the observed classification accuracy rate (e.g., the observed test sensitivity, specificity, error and/or inconclusive rates). In the PDD testing context the independent variable is the test stimuli while the dependent variable is the observed test result.

Deterministic: refers to testing models in which randomness has no effect on results. If all variance can be explained or controlled the results of these models can be theoretically perfect, meaning that a test or experiment conducted repeatedly on the same issue will always produce the same result. However, differences in response variance may remain uncontrolled or incompletely controlled because of human presentation or human understanding of PDD test stimuli. Deterministic solutions are a hypothetical abstraction in the PDD setting, and it is said that there is no such thing as a

'perfect test' (i.e., there is always some small potential for error). PDD testing is therefore not a deterministic test.

Diagnostic test: any test conducted in the presence of a known problem or known symptom/s. The existence of a known problem means that some form of action will or will not occur depending on the test results. In PDD testing, resulting action may involve additional posttest interviewing, additional investigations, administrative action or other action. Diagnostic tests are commonly required to have a high level of criterion accuracy because resulting actions may involve individual rights and liberties. High accuracy is achieved by interpreting the results of event-specific diagnostic exams at the level of the test as a whole.

Discriminant Analysis (linear discriminant analysis, discriminant function): a statistical method used in pattern recognition and machine learning to find a linear combination of features that most effectively separates two or more classes or categories of objects or events. Discriminant analysis can be used to develop predictive models to calculate the probability of an unknown categorical outcome based on known response data. Discriminant analysis can be used in PDD research to develop and validate test data analysis models, and to calculate discriminant functions that provide information about optimal response item weighting.

Distribution: refers to the shape or pattern of the sample or population data when graphed or plotted. Several different distribution shapes can be observed in PDD phenomena.

- **Binomial distribution:** a bell shaped distribution similar to the normal curve but composed of discrete (i.e., non-continuous) stages that result from numerical frequency counting instead of a continuous measurement scale. Binomial distributions are useful in PDD research to calculate the level of significance of frequency counts related to correct or incorrect test results.
- **Exponential distribution:** is related to the lognormal distribution, and is useful in PDD research to understand the shape of RQ/CQ ratios in PDD testing.
- **F distribution:** a commonly used statistical probability distribution in PDD research and other types of research. The F-distribution is used to represent the null distribution of a null hypothesis that is true, and is used to calculate the level of significance of common statistical test such as the F-test and ANOVA methods.
- **Lognormal distribution (logarithmic distribution):** an asymmetrical pattern of data, that is commonly associated with sensory responses. The lognormal distribution can be useful to transform measured or observed PDD responses to a form that is more amenable to statistical analysis.
- **Normal distribution:** many naturally occurring phenomena such as height, weight, IQ, etc. will create a bell-shaped or normal distribution curve, also referred to as a Gaussian distribution after Gauss who first began to publish descriptions of this distribution shape. PDD scores generally conf form to a normal distribution.
- **Poisson distribution:** an asymmetrical probability distribution named after a French statistician, used to calculate probabilities associated with uncommon events (i.e., non-normally occurring) that occur independently during a period of observation. Uncommon events in PDD testing, which may not adhere to a normal distribution might include the occurrence of testing errors and inconclusive results among a group of cases.
- **Chi distribution:** a probability distribution of the sum of the squares of a set of normally distributed random variables. The sum of random values from any distribution can be made asymptotically equivalent to the normal distribution, meaning that the sum of squares of a set of random values from any distribution can be asymptotically equivalent to the chi distribution. The chi distribution is used with the Chi-squared test which can be used in PDD research.
- **T distribution (Student's *t* distribution):** a family of continuous probability distributions used to estimate the mean of a normally distributed population when the sample size is small and the population standard deviation is unknown. The t distribution is used with the t-test which can be used in PDD research.

Effect (effect size or statistical effect): a term used in science and research to describe the strength of an observed phenomenon such as the effectiveness of a treatment or test. In PDD research the effect of interest commonly pertains to unweighted test accuracy. Effect sizes are commonly described by comparing the observed effect to the expected level of effectiveness due to either chance alone or the prior base rate. Effect sizes are often calculated by subtracting the chance or prior probability from the observed research result. Effect sizes in PDD research are often described in simple proportions or percentages.

Empirical (empirical method): knowledge acquired through observation, experimentation, and evidence. Validated PDD techniques are those that have empirical support. Empirical methods are based on the observation, collection and analysis of data to support a conclusion, hypothesis or theory.

Evidence (scientific evidence): anything presented in support of an assertion, hypothesis or conclusion. Scientific evidence is anything intended to support a scientific theory or hypotheses, and is usually in the form of empirical evidence obtained in a manner consistent with the scientific method. The strength of scientific evidence is often a matter of statistical analysis and scientific controls intended to minimize the effects of variables other than the independent variable. Scientific evidence for PDD testing exists in the form of replicated empirical estimates of unweighted test accuracy, sensitivity, specificity, error and inconclusive rates. Also, scientific PDD evidence seeks to provide support pertaining to PDD testing procedures and the psychological and physiological basis for testing.

Evidence-based practice (EBP): an approach to professional practice premised on a requirement that professional decisions and practice standards are based on published research that conforms to some established requirements for quantitative outcomes (e.g., probability or likelihood of successful outcome, or the probability or likelihood of an accurate versus inaccurate test result). EBP in PDD testing is shown by using validated techniques that conform to standard practice requirements.

Experimental design (controlled experiments): refers to data and information gathering activities in which the experimenter is interested in some effect produced by a method. Experiments are conducted for the purpose of obtaining information to verify or refute the validity of a hypothesis. Controlled experiments are those that involve the use of a control group. An important characteristic of experimental research is random assignment to treatment and control groups, ensuring equality and representativeness that can lend support to assumptions about causality when analyzing data and observed effects. Experimental designs are important to all research contexts including PDD testing.

F-test: refers to any statistical hypothesis test for which data under the null hypothesis conform to an F-distribution, including the ANOVA F-test. F-tests and the F distribution can be used in PDD research.

False-negative index (FNI): the ratio of false negative results to all negative results or FN/(TN + FN). In PDD testing FNI represents the conditional probability that results are incorrect when the test results are significant for truth-telling (i.e., are not within the expected normal range of results from deceptive people in the normative sample data).

False-positive index (FPI): the ratio of false positive results to all positive results or FP/(TP + FP). In PDD testing FPI represents the conditional probability that results are incorrect when the test results are significant for deception (i.e., are not within the expected normal range of results from truthful people in the normative sample data).

Forecast: a concept in predictive modeling describing the likelihood that an event will occur. Because it is not possible to predict the future, forecasts do not attempt to predict exactly what will occur or when it will occur. Instead, forecasts are statistical estimates of the probability that an event will occur within a certain time. Forecasts are not made on individual PDD results, but may be made in PDD research contexts that attempt to understand and occurrence of correct or incorrect test results.

Forensic: refers to the interaction of scientific and legal processes. PDD diagnostic exams, conducted during criminal investigations of known events or known allegations, are forensic examinations.

Frequentist inference: a method of statistical inference in which the hypothesis is regarded as static (i.e., either true or false) while the available sample data are regarded as a random representation of the population or real world that makes use of the frequency or proportions of observations within the data. Frequentist statistics imply drawing conclusions about the probability of a true or false hypothesis through the evaluation of the frequency or proportion of the observed or measured characteristics within the sample data. Frequentist inference can be used in PDD research and with individual PDD exams. When applied to individual examination data, the goal of frequentist inference is to calculate the level of statistical significance (e.g., the probability of error or confidence level) of the hypothesis that the examinee belongs to the static category of either deceptive or truthful people.

Hypothesis (hypothesis testing, alternative hypothesis): a proposed explanation for a phenomenon or observation. A hypothesis should be static or deterministic, meaning it is either true or false (i.e., the probability of an effect is significantly different than what is expected due to chance or random/other causes). Hypothesis statements should be falsifiable via a null hypothesis (e.g., the probability of the observed outcome is not significantly different than chance, or something other than the inferred cause). The principles of hypothesis testing can be applied to both PDD research and to individual PDD exams. When applied to an individual examination data, the goal of hypothesis testing is to reject the null hypothesis that an observed deceptive or truthful test result was due to chance or error (i.e., the observed result differs significantly from the expected result if due to chance or error).

Inconclusive (no opinion): a description of examination results that are not statistically significant. Because PDD results involve the evaluation of statistical significance for both deception and truth-telling, results cannot be classified as ei-

ther deceptive or truthful. In practical terms the inconclusive/no-opinion is regarded as a third categorical test result. In a scientific sense, test results that are not statistically significant result in no classification. The term *no opinion* is used synonymously and interchangeably with the term inconclusive by PDD examiners, and provides the advantage of reminding us that inconclusive results do not constitute a categorical test result, and should therefore not be interpreted as either correct or incorrect (i.e., there is no classification).

Independence: refers to the assumption that a response variable is not influenced by other response variables. If the two are independent then one's response cannot be predicted by the other. There is no linear association so their covariance is 0. Independence, in PDD testing, relates to assumptions about both the criterion state (i.e., truthful or deceptive status) of the individual test questions, and with the physiological response to each test stimuli.

Independent variable: refers to the input condition or input variable for a scientific test, experiment or investigation. Independent variables are tested to determine if they are causally related to the output or dependent variable. In PDD research contexts the independent variable includes such things as the structure of the test stimuli, administration and analysis methods, while the dependent variable is the observed classification accuracy rate (e.g., the observed test sensitivity, specificity, error and/or inconclusive rates). In the PDD testing context the independent variable is the test stimuli while the dependent variable is the observed test result.

Information Gained Index (IGI): an epidemiological statistic that is a graphical representation of the likelihood ratio over the range of possible base-rates. The IGI statistic is useful in PDD research because it provides a measure of test effectiveness that is resistant to both sample imbalance and to base rate differences.

Interaction (interaction effect): a term used in research and statistics to describe the relationship between different variables. If two independent variables interact then the relationship of a third (e.g., the dependent) variable to each will be influenced by the other. For example: age group and gender may be said to interact with regard to their relationship to height – because the average height of males may be greater than that for females, but average height for female may be greater for some age groups. Interaction effects of interest to PDD researchers, examiners, and program administrators may involve the relationship between test accuracy and the criterion states of truth/deception along with the results obtained by different studies.

Interpretation: translation of numerical and statistical test results to categorical classifications. It is achieved through structured rules and predetermined alpha boundaries based on the testing objectives and different needs for precision. Interpretation in PDD testing is made with categorical classifications based on the level of statistical significance indicative of deception or truth-telling. In simple terms, PDD examination results are sometimes interpreted as indicative that an examinee has "failed" or "passed" the test, or has lied or told the truth about the issue of interest. However, PDD examiners usually limit their professional conclusions to terms that describe the evidence, such as *deception indicated* (DI) and *no deception indicated* (NDI), and the more conservative term *significant reaction* (SR) and *no significant reactions* (NSR).

Laboratory study: any type of study conducted in a fixed location such as a research environment, as opposed to real-world locations and field circumstances. Laboratory studies in PDD research are important because they provide the opportunity to use random assignment and better control potential confounds. This leads to increased support for inferences about causality.

Law (law of science): statements that summarize a large body of information explaining, often mathematically, an observation or phenomenon in a manner that is supported by large volume of evidence. Laws of science are fundamental ideas that do not change when new theories are developed, but the scope of application may change. Laws of science that pertain to PDD testing involve both measurement and probability.

Law of large numbers (LLN): in probability theory, the law of large numbers describes the expected result if an experiment were conducted a large number of times, and holds that as the number of experiments grows larger the mean of results will converge and become asymptotic with (approach) the expected population value. The law of large numbers is useful in PDD research because it allows us to expect that estimates of test accuracy will, with time and additional research information, become more accurate representations of test accuracy with the more broad testing population.

Likelihood ratio: the Likelihood Ratio (LR) provides a convenient and easily understandable index of how much a test result will change the probability or odds of having a condition with a known or assumed prior incidence rate (i.e., base rate). In the case of PDD testing, the condition of interest is involvement in the issue under investigation. The LR+ tells us how much more likely it is that a person is lying than not, after failing a PDD, compared with the likelihood before he or she sat in the chair and completed the test. If a person produces a truthful test result, the LR- tells us how much more likely a person is to be telling the truth than before the test. Likelihood ratios can be used to describe the benefit or value of conducting a specific PDD test instead of simply guessing the base rate. Additionally, they allow you to compare the ef-

ficacy of test metrics between testing techniques.

Linear (linearity): refers to data that produce a straight line when graphed or plotted. Multiplication of linear data by a scaling factor will retain a linear shape. Linearity is an important consideration in PDD test development because some physiological response phenomena may not conform to linear expectations. Nonparametric methods are sometimes used when linear or parametric assumptions cannot be satisfied.

Lower limit: the lower boundary of observations or measurements that are within or outside the normal range, typically two standard deviations below the mean. In PDD research the lower limit can be used to provide information about the lowest possible estimate (i.e., worst-case scenario) of test accuracy or test effectiveness.

Mean (average): a measure of central tendency. The mean is a statistic that describes how members of a group or population are similar. Statistical analysis of PDD results involves two means, the mean scores or reactions for deceptive people and those for truthful people.

Median: a measure of central tendency. A statistic to describe the middle value of a group of values. The median will be close to the mean with normally distributed data. However, the median is more resistant to outliers than the mean which can differ from the median when data are not normal. Medians are useful in PDD research to help understand the potential of outlier result to distort our knowledge or conclusions.

Mixed design (mixed model): a multivariate research model in which the items or levels of at least one independent variable is non-independent. Mixed designs may be useful in PDD research to evaluate questions about test reliability, and to study the complexities of multiple issue testing.

Mode: another measure of central tendency. The mode or modal response value is the most commonly occurring value in a group. Modes are easily appreciated when looking at histograms. The mode is similar to the mean and median with normally distributed data. The mode statistic may be useful in PDD research when evaluating nominal data.

Model: a defined structured system for measuring and explaining an individual case or group of cases. Models are used to explain how the world works, in circumstances in which the eventual outcome is unknown. Because PDD testing involves the investigation of past events that cannot themselves be tested or observed, PDD testing is a process of modeling the statistical likelihood that a person's physiological response to test stimuli fit the available normative data that describe our present knowledge of how truthful or deceptive people are expected to respond.

Monte Carlo: statistical methods that use existing knowledge and computer simulation and randomization to extend our knowledge. Monte Carlo methods are used in PDD research in ways similar to other fields: to study problems that would be unfeasible to study in other ways, and test hypothesis or develop models that are computationally difficult to achieve through other methods.

Negative: refers, in scientific testing, to test results that indicate a categorical classification of a statistically significant, high likelihood of the absence of the problem, issue or concern under investigation. Negative test results, in **PDD** testing, indicate that a person has "passed" the test. The term negative is preferred because scientific testing is intended to be objective and unbiased regarding the personal values or feelings of the involved professionals. Similarly, the term negative is preferred to "told-the-truth" because it recognizes that test results are probabilistic (i.e., not deterministic) and the interpretation of statistical significance is a function of a predetermined alpha boundary for statistical significance (i.e., use of a different alpha boundary might produce in a result that does differ significantly from those of deceptive people).

- **True-negative (TN):** negative test results known or confirmed to be correct. In PDD testing the true-negative rate is the proportion of truthful people that produce truthful (negative) test results.
- **False-negative (FN):** negative test results known or confirmed to be incorrect. In PDD testing the false-negative rate is the proportion of deceptive people that produce truthful (negative) test results.

Negative-predictive-value (NPV): the ratio of true negative results to all negative results or TN/(TN + FN) in a group of cases. In PDD testing NPV represents the conditional probability that results are correct when the test results are significant for truth-telling (i.e., differ significantly from the results of deceptive people in the normative sample data).

Nonparametric: refers to statistical methods in which data are not expected or required to fit a normal distribution or linear response pattern. Nonparametric methods and statistics are highly useful when data are not normal due to their robust simplicity. Rank ordering is an example of nonparametric approach, in which the actual distance between items is lost, with no subsequent assumptions about linearity. Another example of a nonparametric approach is the *bigger-is-better* rule in PDD test data analysis.

Normal range: also referred to as a reference range or reference interval. The normal range, associated with the normal distribution, is defined as the range in which 95% of observed results will occur. This range is calculated as the range within approximately two standard deviations above or below the mean value. Scores not within the normal range

are referred to as outside the normal range. In PDD testing, the normal range of interest is the scores normally observed among deceptive or truthful people. The normal range is sometimes referred to as a 95% confidence interval.

Normative data (norms): are data that inform us of the normally occurring characteristics in a population. Norms are statistical parameters in the form of means, standard deviations and descriptions of the shape of distribution of data. Norms are based on sample data because it is often impossible to test an entire population, and a single sample is insufficient. Normative data are expected to be representative of the population. Random sampling is a generally accepted method of ensuring that data are collected without bias, though it is axiomatic that all samples are biased (i.e., sample statistics are assumed to always differ slightly from the population if it were possible to know the population statistics). Statistical norms are central to scientific activities, including: forecasting, predication, model fitting, the calculation of the level of significance or probability of error for test results, and other activities. Normative data are also referred to as normative reference data. Like other professionals, PDD examiners in field settings are not expected to make statistical calculations. Instead, normative data are often provided in the form of *normative reference tables* (aka, look-up tables).

Null hypothesis: refers to the opposite of a hypothesis or the default position that there is no effect or relationship between the measured phenomena. The null hypothesis can be formulated by taking the position that the hypothesis has no effect or that observed results will be no different than what would have been observed by random chance. The null hypothesis is rejected when it is statistically unlikely. Rejection of the null hypothesis is required before a hypothesis can be accepted. Null-hypothesis testing is central to scientific study of PDD testing, and also to statistical methods for test data analysis.

Observer effect (Hawthorne effect): a known phenomenon in observational research in which participants modify their performance from normal as a result of the presence of an observer. These effects are thought by different people to be either conscious or unconscious, and possibly due to either novelty or to the potential for positive or negative personal consequences that have not been considered by the experimenter. PDD examiners attempt to avoid observer effects by restricting the presence of people during testing.

Odds ratio: a ratio of the odds of truthful or deceptive calculated using frequency counts. PDD test results involve two odds, including both the odds if deceptive and the odds if truthful. Odds if deceptive is calculated as number of true-positives/number of false-positives. Odds if truthful is calculated as number of true-negatives/number of false-negatives. The odds ratio in PDD testing is the ratio odds if deceptive/odds if truthful. The odds ratio provides a measure of association between the test results and the criterion state. Odds ratios are non-resistant to base-rate inequalities and can be misleading if the sample is not balanced, because they are based on observed frequencies.

Opinion poll: a survey of public or group opinion, designed to evaluate attitudes and perspectives through a series of questions for which data can be subjected to quantitative or qualitative analysis. Opinion polls have been used to study the attitudes of the psychological and scientific community regarding PDD testing, and have also been used within the PDD profession.

Outlier (statistical outlier): an observed value or score that is numerically distant from the remainder of the distribution of values. Outliers are problematic because they can distort test results or research results by causing unrealistic perceptions or expectations that are not replicable. There is no fixed standard for the identification or definition of outliers, but outliers are sometimes defined as those values more than 2.5 or 3 standard deviations above or below the mean value. Another method is to define outliers using quartiles and the interquartile range. In PDD testing, the term can be used to describe individual scores, numerical test results, or research results that differ widely from others.

Overfitting: refers to phenomenon in statists and model development in which a model is unintentionally designed to include random noise instead of the relationship between the diagnostic signal and result or outcome. Overfitting is more common to models that are overly complex, such as when attempting to interpret too many response parameters relative to the sample size or number of observations. Models that are overfit may perform well with the development sample but will generalize poorly to validation data and real-world circumstances. Cross-validation and other techniques can be used to avoid overfitting, though it is generally expected that models perform less well with validation data than with the development sample. Overfitting is an important concern in the development and validation of PDD test data analysis models.

Parameter: a mathematical value, similar to a statistic that describes a characteristic of a population. In PDD research, as in other types of research, population parameters are unknown. Sample statistics can be used as proxies for population information, and the effectiveness of this will depend in part on the size and representativeness of the sample data.

Parametric: refers to statistical methods that make inferences or assumptions about the parametric shape (e.g., normal distribution) or linearity of the measurable data. Parametric methods may be used in both PDD research and PDD test data analysis when necessary assumptions are satisfied.

Positive: refers, in scientific testing, to test results that indicate a categorical classification of a statistically significant or high likelihood of the presence of the problem, issue or concern under investigation. Positive test results, in PDD testing, indicate that a person has "failed" the test (i.e., the test results differ significantly from those of truthful people). The term positive is preferred over the term failed because scientific testing is intended to be objective and unbiased regarding the personal values or feelings of the involved professionals. Similarly, the term positive is preferred to "lied" because it recognizes that test results are probabilistic (not deterministic) and the interpretation of statistical significance is a function of a predetermined alpha boundary for statistical significance (i.e., use of a different alpha boundary might produce a result that does not differ significantly from those of truthful people).

- **True-positive (TP):** positive test results known or confirmed to be correct. In PDD testing the true-positive rate is the proportion of deceptive people that produce deceptive (positive) test results.
- **False-positive (FP):** positive test results known or confirmed to be incorrect. In PDD testing the false-positive rate is the proportion of truthful people that produce deceptive (positive) test results.

Positive-predictive-value (PPV): the ratio of true positive results to all positive results or TP/(TP + FP) in a group of cases. In PDD testing PPV represents the conditional probability that results are correct when the test results are significant for deception (i.e., differ significantly from the results of truthful people in the normative sample data).

Power (statistical power, power analysis): refers to the ability of a scientific experiment to reject the null hypothesis if the alternative hypothesis is true. Power is therefore related to the ability to detect the alternative hypothesis. Power is calculated as function sample size, and can be used to calculate the required sample size to reject the null hypothesis, if it is false. This is equivalent to the probability of not committing a Type II (i.e., false negative). Power of an experiment is allegorical to the sensitivity rate of a test of a single individual.

Prediction: an attempt at categorical assignment into two or more possibilities in advance of an event occurring. Because predictions focus on actual events, they are more localized in time and space than forecasts (e.g., who will win the World Series). Predictions are made in the PDD testing context about group membership (e.g., truthful or deceptive groups), before outcome confirmation evidence exists. PDD test accuracy is therefore a function of predictive accuracy and subject to the laws of probability.

Probabilistic: refers to testing models in which some degree of noise, random, uncontrolled or unexplained variance is expected to exist. Results are probability statements that describe likelihood of correct or incorrect classification. Randomness in probabilistic PDD testing models means that a test conducted repeatedly on the same issue can be expected to produce slightly different results that may have some predictably small effect on either numerical scores or categorical test results. PDD test results, like many other types of scientific test results, are probabilistic results. Results may differ based on the degree of random or uncontrolled variance and will be similar based on the degree of controlled or explained variance. Field practitioners are responsible for using evidence-based methods that allow them to account for the statistical likelihood of a correct or incorrect test result. PDD testing is a probabilistic and not deterministic test and it is the responsibility of PDD professionals to use testing methods that permit one account for the probability of a testing error.

Probability-value (p-value or probability of error): a statistic (p) that describes the probability of that the observed result was caused by random or erroneous response variance. A test result is statistically significant when the probability of error is less than alpha or tolerance for error (i.e., $p <$). In scientific terms p is the probability of the observed result under the null hypothesis – the probability of obtaining a result equal to or more extreme than the observed result. It is the probability of a type I error. Probability values in PDD testing describe the likelihood that a truthful result was produced by a deceptive person or that a deceptive result was produced by a truthful person. The inverse of the p-value can be thought of as a confidence level.

- **Level of significance:** another, more general, term for p-value, probability value, or probability of error. PDD test results can be described in terms of the level of significance. PDD research is concerned with the level of significance of an observed effect.
- **Error statistic:** another, even more general, way of describing the probability of error in PDD testing, but more commonly used to describe the standard error of a statistic.

Proxy data: data that are correlated with the phenomenon or issue of interest, and therefore permit indirect testing and probabilistic modeling that cannot be tested directly (e.g., pregnancy tests that work as a function of the presence of a hormone instead of attempting to observe/interfere with a developing fetus, or HIV tests that evaluate the presence of antibodies in lieu of the actual virus). In PDD testing it is sometimes said that there is no physiological response characteristic or response that is uniquely correlated with deception. This can be said of virtually every physiological

response characteristic and all human phenomena – all physiology is multi-purposed. PDD tests, like other tests, work as a function of the combination of physiological response characteristics that have been shown to be individually correlated with deception and structurally predictive of truthful/deceptive group membership when combined and evaluated probabilistically with respect for available knowledge regarding response characteristics among confirmed truthful and deceptive normative samples.

Psychology: the scientific study of mental functions and behaviors. Some theories and principles of psychology are used to conduct the PDD examination. Psychological theories are also to help understand and explain the reasons the test works, suitability for testing, and reasons for potential vulnerabilities to testing errors.

Psychophysiology: the scientific study of the interaction of physiological responses and psychological processes. PDD testing relies directly on the science of psychophysiology for information and knowledge regarding physiological recording and physiological response mechanisms.

Physiology: the scientific study of living systems. All of the recorded signals in PDD testing are physiological, including respiratory, cardiovascular, electrodermal and somatic activity.

Problem of multiple comparisons (multiplicity, multiple testing effect): refers to a known probability phenomena in statistical hypothesis testing, involving the inflation or deflation of alpha and the increase in the probability of error when making multiple probabilistic decisions within a single test or experiment. In PDD testing this phenomena is related to the use of subtotal scores to make categorical classifications of the test results, and can be managed through the use of common statistical solutions such as the Bonferroni or Šidák corrections.

Psychophysiological detection of deception (PDD): also known as polygraph testing or lie detection testing, is the use of instrumental recording and testing procedures to increase the accuracy of categorical classifications regarding deception and truth-telling.

Qualitative: refers to analysis methods involving subjective or impressionistic description of the observed data as opposed to quantitative measurement. Qualitative methods are use in PDD research and other social science research.

Quantitative: refers to analysis methods involving measurement. Quantitative methods are central to all types of research and testing, including PDD testing.

Quartile (interquartile range): a nonparametric method to describe the dispersion of a group of data values. Quartiles are defined by rank ordering the data and dividing the values in to four equal-sized groups. The interquartile range is the middle two (2^{nd} and 3^{rd}) of these groups. Quartiles and the interquartile range can be used with PDD data to determine that normal range of response, or to identify possible outlier responses.

Quasi-experimental design: refers to research methodologies similar to traditional experimental designs but lacking random assignment to treatment or control conditions. Group assignment may be subject to researcher control according to specified criterion, or may be beyond the control of the researcher. For this reason, quasi-experimental designs may have deficiencies in internal validity when treatment and control groups are not characteristically or statistically similar, as would be expected with random assignment. The result is that assumptions about causality are more difficult and may not be supported by quasi-experimental research without additional support. Quasi-experimental designs are often used in PDD field studies that lack random assignment.

Receiver operating characteristic (ROC): a statistic developed in signal detection theory and used in epidemiological research to describe the effectiveness of testing and classification models. The ROC statistic is calculated and described as the area under the curve, and is useful in PDD research as a measure of discrimination (i.e., the ability of a test to make accurate classifications), and is calculated as the ratio or true positive results and false positive results across the range of possible alpha boundaries and decision cutscores. ROC statistics are therefore resistant to differences in alpha boundaries and decision cutscores.

Regression analysis (including multiple regression and logistic regression): a family of statistical methods for estimating the relationships between variables. Simple regressing involves two variables and the effect that one has on the other. Multiple regression can be used to estimate the contribution of a group of variable to an outcome variable. Logistic regression is useful when the outcome is a binary or categorical variable such as the correct or incorrect classification of PDD test results. Regression analysis can be used to develop predictive models that allow for the calculation of the probability of observing an unknown outcome based on known response data.

Regression toward the mean (mean regression): a known phenomenon in probability theory in which if a score or value is extreme on its first measurement it will tend to be closer to the average on its subsequent measurement. Conversely, if a score or value is extreme on its second measurement it will tend to have been closer to the average on its first measurement. Regression toward the mean is not a causal phenomenon, and can help explain differences between extreme and non-extreme reactions only inasmuch as

variance in reaction is partially determined by random chance. This phenomenon assumes no actual change in state between measurements. Regression toward the mean illustrates an important caution in PDD research in that performance of the test or professional skills will tend to regress toward the mean as more data becomes available. For this reason, the selection of extremely qualified experts who produce extraordinary high rates of test accuracy may result in data that does not correctly or realistically model test accuracy. Generalizable results may be more easily obtained through random selection of both test participants and examiner participants.

Reliability: refers to the repeatability of an observed result. Two general forms of reliability are commonly discussed in PDD and other research: test-retest reliability, and inter-scorer reliability.

- **Test re-test reliability:** refers to the degree to which an observed result is likely to be observed again if the test is repeated. Re-test reliability refers, in PDD testing, to the likelihood of observing the same test result if the test is repeated at a later time.
- **Inter-scorer reliability:** refers to the degree to which the observed scores and result are likely to be repeated if the data are re-scored by another evaluator. Inter-scorer reliability, also known as interrater reliability, refers, in PDD testing, to the likelihood of observing the same result if two different evaluators score the recorded test data.

Research (scientific research): relies on the application of the scientific method to increase our knowledge regarding PDD testing or any domain of knowledge.

Risk ratio: similar to odds ratio but calculated using proportions instead of frequencies. The risk ratio if deceptive is calculated as the sensitivity rate/false-positive rate. The risk ratio if truthful is calculated as specificity /false-negative rate. The overall risk ratio is the ratio of the risk if deceptive/ risk if truthful. Because risk ratios are calculated with proportions, whereas odds ratios are calculated with frequency counts, these are more resistant to imbalance between the truthful and deceptive groups. The ratio of the two risk ratios provides a measure of association between the test result and the criterion state.

Sample: a group of cases selected to represent the population. Random sampling is more representative. Field samples are nonrandom and laboratory samples are usually random. PDD sample data exists for deceptive and truthful people, and test results are compared statistically with both.
Scientific control: a method intended to reduce the effects of variables other than the intended independent variable on the outcome of interest. Scientific controls are intended to reduce the occurrence of type I errors. Samples of truthful people serve as scientific controls in PDD research. In the context of the PDD test, comparison questions may serve as a form of scientific control, intended to reduce the likelihood that the outcome (i.e., positive or negative test results) are the result of factors not attributable to the dependent variable (i.e., test target stimuli).

Scientific method: a collection of techniques used to investigate observed phenomena, acquire new knowledge, synthesize new and old knowledge, or refute existing knowledge. The scientific method is one of following the data; if a prediction or model is confirmed then the hypothesis or theory is supported. If the prediction is not confirmed then the hypothesis or theory is not supported. Results that are not solely attributable to the independent variable are said to be c*onfounded*. Scientific methods are expected to be objective and repeatable by others. PDD research is expected to conform to the scientific method.

Screening test (exploratory test): refers to testing activities conducted in the absence of a known problem or known symptoms. PDD screening tests are defined by the absence of a known incident or known allegation. These can be conducted as either single issue or multiple issue examinations, depending on the screening circumstances and needs.

Sensitivity (test sensitivity): refers to the ability of a test to detect or notice the issue under investigation. The sensitivity rate is the same as the true-positive rate. In PDD testing, test sensitivity describes the proportion of deceptive people expected to be detected (e.g., produce positive test results).

Signal detection: A scientific activity involved in the determination of the presence or absence of a signal, feature, symptom, or characteristic of interest. Signal detection theory involves measurement, signal processing, and statistical decision theory. The goal of signal detection theory is to quantify our ability to detect information (i.e., signal) from noise. Signal detection theory is useful in PDD testing to determine physiological responses of interest and to aid in the development and validation of decision models based on those signals.

Signal discrimination: refers to activities in signal detection with an emphasis on classification. Whereas signal detection models attempt to answer a question about the presence or absence of a signal – for which the result will be *yes* or *no*, signal discrimination models attempt to predict group membership – with the result taking the form of either *A* or *B*. The signal being discriminated during PDD testing is truth or deception (involvement or noninvolvement) about a past behavioral event.

Significant (statistically significant): refers to probability of error (p) which less than a predefined alpha boundary ( ) or tolerance for error ($p <$). Results that are statistically significant are not likely to have occurred due to error or chance alone. The observed result can therefore be reasonably regarded as probably real, and therefore probably repeatable. The determination of statistical significance is a mechanism for the reliable interpretation of categorical classifications or decision based on statistical or probabilistic test results. In PDD testing this means that results can be reasonably attributed to the hypothetical model or assumption (i.e., deception or truth-telling).

Specificity (test specificity): refers to the ability of a test to determine the absence of the specific concern or concerns under investigation. Test specificity is the same as the true-negative rate. In PDD testing, test specificity describes the proportion of truthful people expected to produce negative test results.

Standard deviation: a statistic that describes how members of a group or population are different. Standard deviation is calculated as the square root of the variance. Statistical analysis of PDD results involves two standard deviation statistics, the standard deviation for deceptive group and the standard deviation for truthful group.

Standard error: a statistic that describes the degree of potential error or bias of a sample statistic (e.g., standard error of the mean or SEM). Whereas the standard deviation is a measurement of dispersion for the values in a sample or population, the *standard error* is the standard deviation of a statistic. Standard errors, in PDD research, can help us to better understand the level of precision or potential bias with which a research result attempts to estimate the means of the populations of truthful or deceptive people (i.e., the range within which we are 95% sure the population mean exists).

Standardization: mathematical transformation of different types of numerical data or different numerical scales to a common metric involving a mean of 0 and standard deviation of 1. Standardization of numerical values allows simple and easy comparison of different types of data (e.g., pneumograph, electrodermal, cardio). Standardized values permit the calculation of statistical classifiers to describe PDD test results.

Statistic: a mathematical value to describe some characteristic of a sample intended to represent a population. Statistics are used to describe normative data for PDD test scores of truthful and deceptive groups, and our knowledge of PDD test accuracy.

Statistics: the study of data through quantitative and mathematical analysis. Statistics are related to mathematics and science. Statistics and statistical analysis are useful to help improve and extend our foundation of knowledge regarding PDD testing.

Survey (survey research, statistical survey, or survey methodology): statistical and research methods that involve the sampling of individuals from a population. Survey methodologies include activities intended to improve the formulation of questionnaire and test item construction. Surveys are important to help understand trends within the PDD profession, and also to help understand opinions, attitudes and perspectives among professionals and members of the public.

T-test (Student's *t*-test): refers to any statistical hypothesis test for which the statistic conforms to the Student's t distribution. The t-test can be used to compare to sets of data and can be used when data are expected to be normally distributed though population parameter is unknown. The t-test can be used in PDD research.

Theorem: mathematical statements that are proven on the basis of previously established statements. Theorems are deductive statements (i.e., proceed from general statements to specific conclusions), whereas scientific knowledge is generally empirical. Theorems differ from scientific theories in that scientific theories must be falsifiable (i.e., are testable). PDD research depends on mathematics and statistics which are grounded in theorems such as the central limit theorem and the law of large numbers.

Theory (scientific theory): an explanation of observed real-world phenomena, based on knowledge that has been repeatedly confirmed through observation and empirical evidence. Although theories rest on a basis of evidence, it is sometimes said that all theories are wrong in that they are incomplete (i.e., awaiting falsification) and therefore strictly unprovable. In other words, there is always more to learn. A goal of all theories is to continue to explain additional observed phenomena until a theory fails to adequately explain the observed evidence or can be modified into a new theory to account for both old and new information. Theories require replicated empirical evidence before they can be accepted. In contrast, hypotheses are statements pertaining to outcomes of individual studies. Theories in PDD research may pertain to both fundamental psychological and physiological constructs, or to pragmatic constructs such as test operation.

Thought experiment: a type of scientific experiment that is conducted via verbal discussion or abstract thought, including paper and pencil analysis, for the purpose of evaluating or illustrating the potential consequences of an idea or action. It may or may not be possible or necessary to conduct these experiments in reality. Thought experiments can be applied to some PDD research questions (e.g., conditional probabilities or base rate estimates) and other scientific contexts.

Transformation: conversion of recorded data to numerical scores or measurements. Also refers to the reduction of multiple measurements of multiple response features and multiple presentations of the test stimuli to a smaller set of numerical values to represent the testing target issue or issues. Transformations in PDD testing involve converting physiological responses to either linear measurements or nonparametric integer scores, and converting those values to grand total and subtotal scores.

Treatment (condition): a term in PDD and other research context that refers to the experimental or testing conditions (i.e., independent variables) that are applied to, or manipulated for, the participants in a study.

Type I error: the incorrect rejection of the null hypothesis or the incorrect acceptance of a hypothesis. This can also be thought of as a statistically significant finding that is incorrect. False positive errors in PDD testing are typically classified as Type I errors. Of note however, in PDD field testing these can actually be either false positive or false negative errors because both positive and negative test results can be determined via statistical significance.

Type II error: failure to reject a null hypothesis that is false. Also the failure to observe a statistically significant result with the hypothesis is correct. False negative errors in PDD testing are typically classified as Type II errors. If inconclusive results are regarded as errors, these will also include inconclusive results. However, if statistically significant results are required for either a deceptive and truthful classifications, then inconclusive results should not be used to make categorical classifications. Non-significant test scores result in no classification. Inconclusive PDD results are an example of type II errors, but are not synonymous with false negative errors.

Unweighted accuracy rate: the arithmetic mean (e.g., average) of the proportion of correct deceptive and correct truthful sample cases, excluding inconclusive results. This statistic differs from the weighted accuracy in that it is calculated as the average of the proportions of correct classifications, whereas the weighted accuracy is calculated using frequency counts. Unweighted accuracy is thought to be more generalizable estimate of PDD test accuracy than the weighted arithmetic mean accuracy rate because it is resistant to sample imbalance.

Unweighted inconclusive rate: the arithmetic mean (e.g., average) of the proportion of inconclusive cases for the deceptive and truthful sample groups. The unweighted inconclusive rate, calculated using proportions, is thought to be more generalizable estimate of the rate of inconclusive PDD test result than the weighted arithmetic mean inconclusive rate, calculated with frequency counts, because it is resistant to sample imbalance.

Upper limit: the upper boundary of observations or measurements that are within or outside the normal range, typically two standard deviations above the mean. In PDD research the higher limit can be used to provide information about the highest possible estimate (i.e., worst-case scenario) of errors or inconclusive results.

Validity: refers to the degree to which a hypotheses, theory, measurement, or conclusion is supported by evidence. Also refers to how well a hypothesis or theory corresponds to, or explains, the real world. In simplistic practical terms validity refers to how well a test measures what it claims to measure. Validity is a complex set of ideas and includes several different aspects. Several different types of validity of interest to PDD research.

- **Content validity:** refers to whether a test or model measures all facets of a phenomenon being investigated. Questions about content validity in PDD research will address whether there are other aspects of psychophysiological response that should be recorded or measured.
- **Construct validity:** refers to the strength of inferences that a measurement or model actually represent or measure the phenomenon being investigated. Questions about construct validity, in PDD research, will involve correct assumptions about underlying psychological and physiological mechanisms related to recorded responses to test stimuli.
- **Criterion validity:** refers to how well a test or model predicts the outcome of interest. In PDD research, criterion validity is allegorical to test accuracy.
- **Ecological validity:** refers to whether the methods and context of a study adequately approximate real-world circumstances. This is related to, but distinct from, external validity. Ecological validity is related to generalizability, and studies with greater ecological validity may be more generalizable. However, ecological validity is not necessarily a requirement for generalizability of cause-effect relationships.
  External validity is a necessary requirement for generalizability. It is important, when evaluating the results of PDD and other research, to avoid confusing ecological validity with external validity.
- **External validity:** refers to the extent to which research results can be generalized to other situations and other people. Inferences or conclusions about cause-effect relationships are said to have external validity if they can be generalized from idiosyncratic research contexts to other groups, including the general testing population. External validity can be compromised

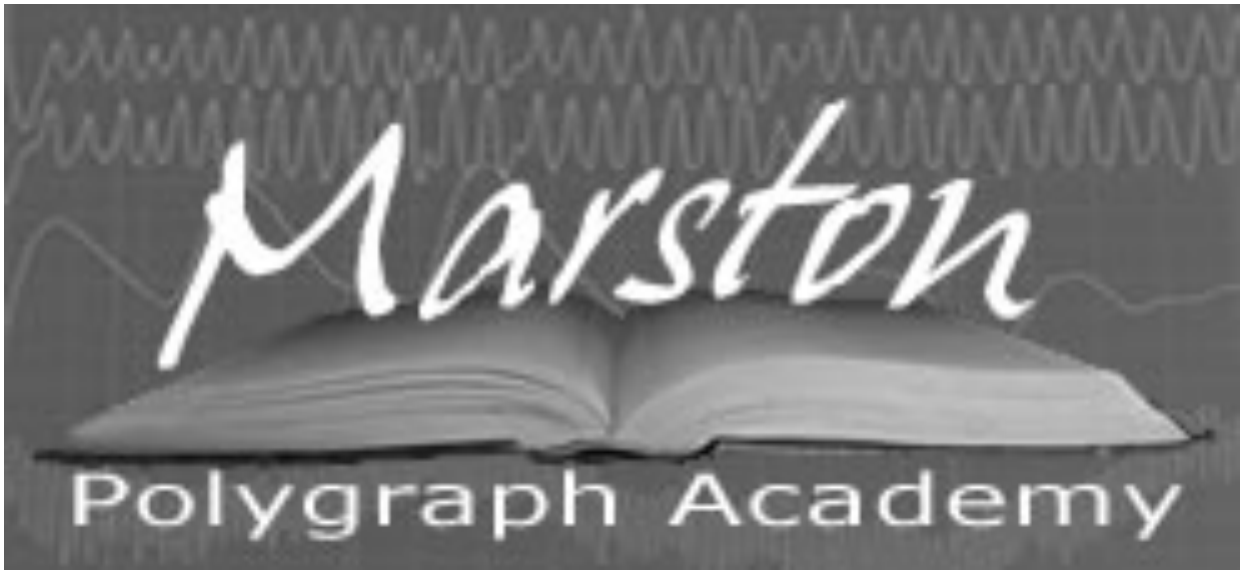
by small samples and non-random sampling. External validity is an important concern to PDD and other fields of scientific research that attempt to identify issues of causality.

- **Face validity:** refers to expert opinion regarding a hypothesis or basic idea. Face validity is a subjective opinion about whether a test appears to measure what it purports to measure. Face validity is an important aspect of PDD research at the onset of a course of investigation and whenever empirical evidence is not available. Face validity is generally viewed as an early solution that is inadequate to answer scientific questions or to form the basis of a theory.
- **Incremental validity:** refers to whether information from a test or model will increase the accuracy of judgments or predictions made without the new information. Incremental validity, in the PDD context, relates to the contribution that information and results from PDD testing offers to risk assessment and risk management activities.
- **Internal validity:** refers to a characteristic of scientific studies in which systematic error (e.g., bias) is minimized so that study results can support assumptions about causality. Inferences involving causality can be said to possess internal validity when three conditions are satisfied: 1) the cause precedes the effect in time, 2) the cause and effect are correlated, and 3) there is no plausible alternative explanation for the observed relationship. Internal validity is an important concern in PDD research and all research that attempts to convey conclusions about causality. Research that lacks internal validity cannot support causal conclusions.

Variance: refers to the dispersion or how far a set of data values is spread out. In a more general way variance refers to changes in the observed data. Variance is useful to PDD research to calculate and determine whether response differences between truthful and deceptive groups are significant (i.e., real differences). Variance is also useful in field PDD testing to calculate the level of statistical significance (e.g., probability of error or p-value) of the test results.

- **Controlled variance:** changes in the data attributable to or caused by the test model or test stimuli. In the PDD research context, controlled variance is related to the effect size that describes the effectiveness of a model, hypothesis or experiment. In the PDD testing context controlled variance can refer to measurements or observations within the data that are attributable to the test stimuli.
    - **Explained variance:** another way of describing changes in the data that are attributable to the test stimuli.
    - **Diagnostic variance:** also another way of describing changes in the data that can be assumed to be caused by the test stimuli or questions.
    - **Signal:** a more general way of thinking about and describing controlled variance in the observed data.
- **Uncontrolled variance:** changes in the data that are *not* attributable or not caused by the test model or test stimuli. In the PDD research context un controlled variance is related to the degree of error or imperfection of a model, hypothesis or experiment. In the PDD testing context uncontrolled variance can refer to the measurements or observations in the data that are not attributable to the test stimuli.
    - **Unexplained variance:** another way of describing changes in the data that are not attributable to the test stimuli.
    - **Error variance:** also another way of describing changes in the data that can be assumed to be caused by *something other than* the test stimulus questions.
    - **Noise:** a more general way of thinking about and describing *uncontrolled* variance in the observed data.

Within-subjects design (repeat measures design): a research or testing model in which participants are tested more than once. All participants are exposed to every treatment or condition of the independent variable. Within-subjects designs can increase statistical power, and can reduce between-subject variance that may distort some types of research data. Within subject designs are ideal for longitudinal studies, in which each participant serves as his or her own control. Within subjects designs may include the potential that responses are affected by boredom, fatigue, and learning/practice. Responses may also be affected by order effects, though these may be reduced through the use of counterbalanced data collection in which different subgroups are exposed to treatments/conditions in different order. Multiple issue PDD exams are allegorical to within-subjects research designs and may be subject to similar potential confounds and other carryover effects in which response to individual test target items may affect response to other test target items. The result of this is that response variance to target questions in multiple issue PDD exams may not be independent, even though these examinations are interpreted with the assumption that the criterion variance of test items is independent.



MARSTON POLYGRAPH ACADEMY, LLC
390 Orange Show Lane - San Bernardino, California
(877) 627-2223

Accredited by the American Polygraph Association
Recognized by the American Association of Police Polygraphists

www.marstonpolygraphacademy.com
mail@marstonpolygraphacademy.com

Using Normative Reference Data with Diagnostic Exams and the Empirical Scoring System

Raymond Nelson and Mark Handler

What is the level of statistical significance or probability of error that was calculated for the scores of this examination? What are the alpha boundaries at which a results will be considered significant? Are the examination data scored using an evidence-based method or an experimental method without published and replicated evidence? Is the level of significance calculated through normative or ipsative methods? Are error and accuracy estimates calculated using inferential or Bayesian statistical principles? What is the impact on the error estimate of this examination if the base-rate or incidence rate of the target behaviors is unknown or different from what was assumed while scoring and calculating the confidence level for this examination? What statistical methods are used to account for the impact on the statistical significance differences that result from the use of multiple subtotal scores?

As the emphasis on evidence-based practices increases in all fields of social, medical and forensic science, professionals faced with the need to defend a polygraph test result in a court or evidentiary hearing will increasingly be presented with an array of scientific questions. These, and other, scientific questions are most likely to be asked of examiners or experts attempting to defend the merits of polygraph examination results, either individually in courtroom or evidentiary proceedings, or in the context of legislative discussions pertaining to decisions about polygraph programs.

The least opportune moment to first encounter scientific questions will be during an evidentiary proceeding, while being questioned by the opposing counsel. It will be prudent instead to become familiar with scientific questions and the application of scientific testing principles to polygraph examination data in advance of any legal proceeding. Perhaps more importantly, incorporation of the conceptual language of scientific testing into the common vernacular of polygraph professionals will help to improve the stature of the profession and the esteem regarded the polygraph profession by professionals in related scientific disciplines.

The Empirical Scoring System (ESS) (Nelson, Handler, Shaw, Gougler, Blalock, Russell, Cushman and Oelrich, 2011) is an evidence-based model for test data analysis of comparison question test (CQT) methods, for which normative data have been published for both event-specific diagnostic and multi-issue screening polygraphs. The goal of the ESS is to empower field examiners, experts, and program administrators to answer scientific questions. Like evidence-based methods in other fields of science, the ESS includes only those assumptions and procedures for which there is published scientific evidence of support. ESS results are statistical estimates of the probability of error that can be easily calculated for any numerical score, in the form of a p-value that describes the level of statistical significance. Because it is not common - in medicine, psychology, educational testing, or field polygraph testing - for field practitioners to manually calculate the level of statistical significance for a test result, the ESS includes normative reference tables (also known as lookup tables, normative lookup tables, or simply normative tables) that can be used to quickly determine the probability of error or level of confidence for an individual test result.

Step 1: Locate the normative reference table.

The first step to determining statistical significance will be to locate the normative reference table for the type of examination that was conducted. Different examination techniques will have different distributions of scores (i.e., means and standard deviations of deceptive and truthful scores among the population and sample cases). Normative data tables in Nelson *et al.,* (2011) for ESS scores of ZCT examinations with two and three relevant questions, and for multiple issue screening examinations with two, three and four relevant questions. Table 1 shows the normative reference data for deceptive and truthful results of diagnostic polygraphs with three relevant questions.

Table 1, Normative reference table for ZCT exams

Truthful (NSR) Cutscores		Deceptive (SR) Cutscores	
Column 1	**Column 2**	**Column 3**	**Column 4**
Total NSR Cutscore	**p-value (alpha)**	**Total SR Cutscore**	**p-value (alpha)**
-1	0.159	1	0.159
0	0.130	0	0.127
1	0.106	-1	0.099
2	0.085	-2	0.077
3	0.067	-3	0.058
4	0.052	-4	0.043
5	0.040	-5	0.032
6	0.030	-6	0.023
7	0.02	-7	0.016
8	0.017	-8	0.011
9	0.012	-9	0.008
10	0.008	-10	0.005
11	0.006	-11	0.003
12	0.004	-12	0.002
13	0.003	-13	0.001
14	0.002	-14	<.001
15	0.001		
16	<.001		

Step 2: Determine the required alpha boundaries and cutscores.

The second step will be to determine the cutscores that will be required to classify the test result and examinee's answers as either deceptive or truthful. This is accomplished by locating the a priori alpha boundaries (determined prior to testing) within the normative reference table, and selecting the corresponding cutscore in the nearby column. *Alpha*, sometimes denoted with the Greek letter , is the statistical term for *cutscore*. Scores, in scientific and statistical analysis, are commonly expressed in terms of a *p*-value (*p*) which refers to a probability of error. In this way scientific thinkers have a common way of describing and understanding the strength and meaning of test results, without having to know the exact details of how the scores are determined in every different form of scientific testing.

Alpha boundaries are commonly set at .05 in the social and behavioral sciences, and this corresponds to a tolerance of 5%, or a maximum permissible error rate of 5%. Alpha boundaries are also set at .01, indicating a tolerance for error at 1%, and .10, indicating a 10% maximum tolerance for error. Polygraph testing with CQT formats will require the use of two alpha, one for deceptive and one truthful classifications. Recommended alpha levels for the ESS are .10 for truthful classifications and .05 for deceptive classifications.

The cutscore for truthful classifications can be obtained by using the left two columns of Table 1, corresponding to truthful test scores. Locate the largest value in column 2 that is still smaller than .10; then look in the corresponding row in column 1 to see the corresponding cutscore. A grand total cutscore of +1 will provide an error rate of .106 which is larger than .10. A cutscores of +2 will provide an error rate of .085, which is less than .10. A cutscore of +3 will provide an error rate of .067 which is smaller than .10 but is also smaller than .085. Therefor +2 is the optimal cutscore to achieve less than 10% errors when making truthful classifications, while also reducing inconclusive results to a minimum level. If greater accuracy or fewer errors is required, simply select the cutscore corresponding to .05 (+5) or .01 (+10). Although errors may be reduced by using more conservative alpha boundaries, there will be a corresponding increase in the rate of inconclusive results for truthful persons when doing so.

The cutscore for deceptive classifications can be obtained by using the right two columns (columns 3 and 4) of Table 1, corresponding to deceptive test scores. Locate the largest value in column 4 that is smaller than .05; then look in column 3 to find the corresponding cutscore. A cutscore of -4 will provide an error rate of .043, which is less than .05. A cutscore of -3 provides an error rate of .058 (5.8%) which is more than .05. A cutscore of -5 would provide an error rate of .032, which is satisfactory but will produce more inconclusive results than a cutscore of -4.

Although test results for event-specific examinations are determined at the level of the test as a whole, subtotal scores can also be used to make deceptive classifications. Because the grand total score generally provides the most accurate results, subtotal scores should be used when the result is inconclusive using the grand total score. Use of the subtotal scores with event-specific diagnostic exams amounts to this question: does any of the individual stimulus questions producing a subtotal scores that is equivalent to scores normally produced by the grand total? Or, is any one of the individual questions doing the work normally done by the grand total?

The level of statistical significance for the individual subtotals is determined by comparing each of the subtotal scores to the normative distributions (reference tables) for grand total scores. Because a decision is made about each individual subtotal scores, and because the number of opportunities to achieve a statistically significant result is equal to the number of subtotal scores, error rates become compounded when doing this. The actual error rate can be determined by multiplying the desired alpha level by the number of relevant questions (3 relevant questions in a ZCT format). This phenomena is known as *inflation of alpha or inflated alpha*, and will result in an increase or excessive rate of false-positive errors (.05 x 3 = .15). The solution, named after the statistician Bonferonni (Abdi, 2007), is applied, in the form of a *Bonferonni correction* to the alpha or *Bonferonni corrected alpha*, by dividing the desired alpha level by the number of decisions to be made (.05 / 3 relevant questions = .017). The statistically optimal cutscore for decisions based on subtotal scores can then be located in columns 3 and 4 of Table 1. Find the largest value in column 4 that is smaller than the Bonferonni corrected alpha (.017) and locate the corresponding cutscore in column 3. A cutscore of -6 would provide a corrected alpha level of .023 that is larger than .017. Use of this alpha/cutscore will allow the use of subtotal scores to reduce inconclusive results and increase test sensitivity to deception while also constraining false false-positive errors to less than 5% (.017 x 3 relevant questions = .05).

Recommended default alpha levels for ESS scores of ZCT examinations are thus: grand total scores of +2 or greater are truthful (< .10), while grand total scores of -4 or lower (further from zero) are deceptive (< .05). If the grand total score is inconclusive, deceptive classifications can be made when any subtotal score is -7 or lower.

As a practical matter, many field examiners will find that alpha boundaries and cutscores are the same for many examinations. As a result, they may find it unnecessary to determine alpha boundaries and cutscores for each examination, and will instead simply remember and apply established cutscores corresponding to desired alpha levels as determined by their agency or customers.

Step 3: Determine the level of statistical significance for the test scores.

The third step, when using normative reference tables, is to determine the actual level of statistical significance for the grand total and subtotal scores. If the grand total score is greater than zero, simply locate the score in column 1 of Table 1 and identify the corresponding p-value (*p*) in column 2 to determine the probability of error. The result is statistically significant, allowing a truthful classification of the test result, if the p-value is less than alpha ($p <$), which will always be the case if the score is greater than or equal to the truthful cutscore.

If the grand total score is less than zero, simply locate the score in column 3 of Table 1 and identify the corresponding p-value (*p*) in column 4 to determine the probability of error. Again, the result is statistically significant, this time allowing a deceptive classification of the test result, if the p-value is less than alpha ($p <$) and this will always be the case if the score is less than or equal to the deceptive cutscore.

When the grand total score is inconclusive, each subtotal score should be compared to the deceptive cutscore for subtotal scores, using the subtotal determined with the Bonferonni corrected alpha. The result is statistically significant, allowing a deceptive classification of the entire examination, if the p-value is less than alpha ($p <$) for any of the subtotal scores, and this will always be the case if any of the subtotal scores are less than or equal to the required cutscores for deceptive classifications based on subtotal scores.

Examples 1 through 4 show different patterns of numerical results that can occur using a ZCT format, along with the level of significance for each subtotal and grand total score. Example 1 shows scores for which the grand total is statistically significant for truth-telling. Example 2 shows scores for which the grand total is statistically significant for deception. Example 3 shows scores for which the grand total is inconclusive while a subtotal score is statistically significant for deception. Example 4 the scores for an inconclusive test result. Note that p-values are not provided for truthful subtotal scores because truthful determinations are not made using subtotals scores of ZCT examinations.

Example 1.		
Question 1	**Question 2**	**Question 3**
+1	+1	+1
+3 (p = .067)		

Example 2.		
Question 1	**Question 2**	**Question 3**
-2	-2	-1
-5 (p = .032)		

Example 3.		
Question 1	**Question 2**	**Question 3**
+2	+3	-8 (p = .011)
-3 (p = .058)		

Example 4.		
Question 1	**Question 2**	**Question 3**
+1	+1	-1 (p = .099)
+1 (p = .106)		

Step 4: Interpret the test scores.

The fourth step when using normative reference data is to interpret the test result for the referring professionals. Interpretation of the test result is intended to answer this question: what can be reasonably said in human language regarding the numerical scores obtained during this exam? This interpretation will often be written into the language of the examination report, and tends to become part of the standard lexicon and vocabulary of both polygraph professionals and consumers of polygraph examination results.

Test results are most effectively reported in terms of both a categorical determination of passing or failing (e.g., deception indicated, no deception indicated, significant reactions, or no significant reactions) along with sufficient numerical and statistical information so that another professional or scientific thinker could understand the scientific meaning of the test data and result. It is therefore helpful to include minimal information about the numerical subtotal and grand total scores, along with the level of statistical significance for those scores.

Referring to example 1, the following interpretation can be provided:

Using the ESS, an evidence-based, normed, and standardized protocol for test data analysis, the grand total score of +3 equals or exceeds the required cutscore of +2 for truthful classifications. The level of statistical significance, is calculated at p = .067, which is equal to or less than the required alpha boundary (= .10), and indicates that only small proportion of deceptive persons (6.7%) will produce an equal or greater test score. These results support the conclusion that there is no deception indicated by the physiological responses to the test stimulus questions during this examination.
Referring to example 2, the following interpretation is offered:

Using the ESS, an evidence-based, normed, and standardized protocol for test data analysis, the grand total score of -5 equals or exceeds the required cutscore of 0-4 for deceptive classifications. The level of statistical significance, is calculated at p = .032, which is equal to or less than the required alpha boundary (= .05), and indicates that only small proportion of truthful persons (3.2%) can be expected to produce an equal or lower test score. These results support the conclusion that there is deception indicated by the physiological responses to the test stimulus questions during this examination.

Referring to example 3, the following interpretation is offered:

Using the ESS, an evidence-based, normed, and standardized protocol for test data analysis, the strongest subtotal score of -8 equals or exceeds the required subtotal cutscore of -7 for deceptive classifications. The level of statistical significance, is calculated at p = .011, which is equal to or less than the required Bonferonni alpha boundary (= .017), and indicates that only small proportion of truthful persons (3.3%) can be expected to produce an equal or lower test score. These results support the conclusion that there is deception indicated by the physiological responses to the test stimulus questions during this examination.

(Note that the alpha boundary is described with Bonferonni correction, and that the percentage (3.3%) is expressed after multiplying the p-value by the number of subtotal scores.)

Referring to example 4, the following interpretation is offered:

Using the ESS, an evidence-based, normed, and standardized protocol for test data analysis, the grand total score of +1 does not equal or exceed the required cutscore of +2 for truthful classifications, nor does the strongest subtotal score of -1 equal or exceed the required subtotal cutscore of -7 for deceptive classifications. The level of statistical significance for the grand total score, is calculated at p = .106, which the required alpha boundary (= .10), and the level of statistical significance for the strongest subtotals score is calculated at p = .099, which exceeds the required alpha boundary (= .017). As a result of these non-significant numerical scores, the test result is inconclusive and no opinion can be offered regarding truth-telling or deception in response to the stimulus question during this examination.

Although numerical scores are of little actual use to referring professionals, many professionals may be familiar with the interpretation and reporting of scientific test results. Inclusion of both numerical scores and statistical results in an examination report can assist quality assurance reviewers to more expediently support examinations that are interpreted correct and more expediently identify examination that may not have been interpreted correctly. More importantly, withholding of statistical information from the interpretation and reporting of scientific test results encourages over-confidence and simplistic thinking regarding the test result. The impact and damage of lackadaisical and unscientific attitudes regarding scientific test results is often not felt until the need to defend the test result in a courtroom setting or scientific discussion, at which time an elevated level of discomfort and unfamiliarity will be experienced and demonstrated, leading to a loss of credibility and confidence for all involved. Though most examinations do not end up central to courtroom or legal discussions, every examination should be conducted in a manner for which the examiner will prevail in the event of any need to defend the test result in response to question about the evidence-based validity and statistical accuracy of the individual test result.

Scoring of polygraph examinations has traditionally been accomplished using ipsative procedures in which an individual's responses to relevant questions were compared only to responses to comparison stimuli. Differences in response were encoded numerically using integer scores and the results have been compared, in the past, using heuristic rules and traditional cutscores that were determined to work though not based on statistical analysis or normative data. The result has been that statistical classifiers for most manual scoring protocols have been determined through Bayesian models that are non-resistant to differences in base-rates or unknown base-rates. The ESS, using many of the procedural skills as other scoring methods, provides an evidence-based norm-referenced model for which a statistical classifier can be calculated (or determined) using normative data. The advantage of this is that statistical error estimates for the ESS are resistant to differences in or unknown base-rates.

As the trend towards evidence-based practices continues in all fields of behavioral, medical and forensic science, we can expect that the days in which subjective impressions and individual prowess or experience will continue to wane in importance as the basis of scientific confidence when interpreting polygraph test results. Instead practitioners of scientific polygraph testing will be increasingly obliged to answer questions about test accuracy using norm-referenced evidence-based methods.

References

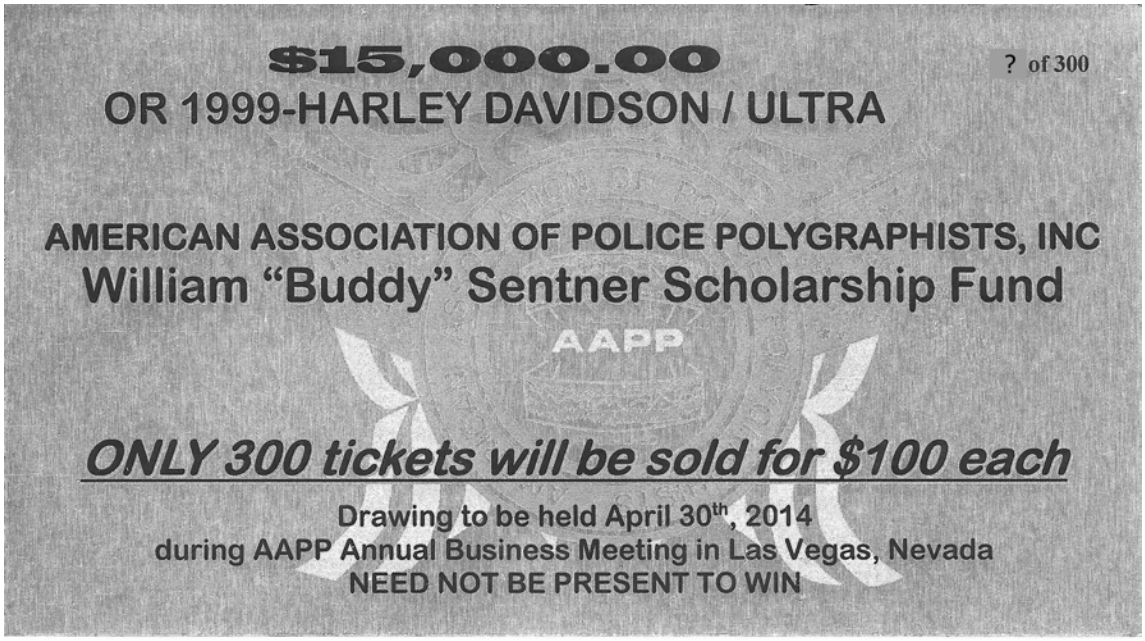
Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage.

Nelson, R. & Handler, M. (2010). Empirical Scoring System: NPC quick reference. Lafayette Instrument Company. Lafayette, IN.

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

This article was originally published by the American Polygraph Association, and is reprinted here with authorization.

A special thank you to the American Polygraph Association and the authors of this article for granting the American Association of Police Polygraphist permission to reprint this article



DO YOU HAVE YOURS???

Only 300 tickets will be sold this year for the AAPP Scholarship Raffle.

You could win your choice of a 1999 HD Ultra Motorcycle Or $15,000 cash.

But you can only get your tickets from your Regional Director, and *once there gone, there gone*.

Contact your Regional Director soon to secure your ticket!!!

*** 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.

Advanced. Powerful. Easy to Use.

The CPS*pro* combines the unparalleled accuracy of Stoelting's polygraph hardware with our all-new state-of-the art Fusion software. Designed from the ground up, CPS*pro* Fusion is loaded with innovative and powerful new features which will provide you with all the tools necessary to efficiently and reliably conduct, score, and report polygraph examinations.

When your reputation is on the line, and the truth is the only thing that matters, you can be confident that the CPS*pro* provides you with the tools to make the right call. Let CPS*pro* put science on your side…

Stoelting
SINCE 1886

Scan this QR code with your smart phone to go directly to our website

620 Wheat Lane • Wood Dale, IL • 60191 • Tel: 630-860-9700 • 800-860-9775 • Fax: 630-860-9775
www.StoeltingCo.com • Polygraph@StoeltingCo.com

QUALITY POLYGRAPH INSTRUMENTATION FROM A COMPANY YOU CAN TRUST.

At Lafayette Instrument we understand the importance of providing polygraph instruments that are trustworthy and dependable. Backed by a team of hardware and software engineers with decades of experience, our LX5000 provides reliable recording of physiological data and the most advanced and documented EDA solutions available.

Lafayette Instrument.

LX5000 HARDWARE FEATURES

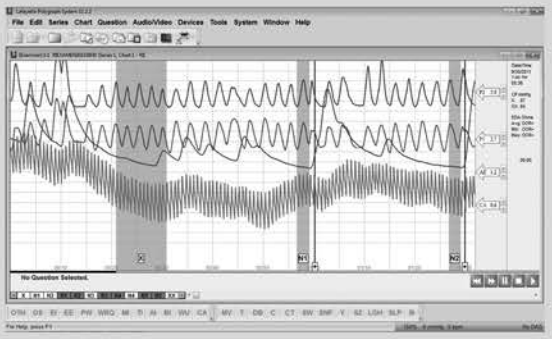
Designed as a robust system that is significantly smaller in size than the LX4000, our standard LX5000 System simultaneously records nine channels and includes many features:

- Data transfer rate up to 360 samples per second across all channels
- 24-bit analog to digital conversion
- Compact design allows for easy transport and storage
- Up to 9 additional channels (18 total)
- Dedicated channels for seat, hand, and feet activity
- Extended measurement ranges
- Selectable GSR or GSC channel
- Dedicated PPG channel
- Operates with proven, state-of-the-art, LXSoftware
- Enclosure is UL 94 V-0 rated, CE and ASTM tested for safety/durability
- 3 year warranty and lifetime technical support

LXSOFTWARE v11.2.2 FEATURES

Our software offers unparalleled ease-of-use, proven reliability, and Windows® 8 compatibility. LXSoftware comes with POLYSCORE® and OSS-3 Scoring Algorithms, as well as the following features:

- Six EDA choices (GSR or GSC manual, detrended, and automatic)
- Customizable reports and question templates
- Up to 16 configurable Audio/Video channels
- Integrated multi-language support for Spanish, Dari, and Russian
- Multi-camera support allows for multiple views of the subject
- Customizable Personal History and Exam/Series forms
- Heart Rate in beats per minute in the status bar
- Data recorded with Masseter Headset is split into jaw movement and audio-level sensor traces
- During Chart Review, charts may be edited and saved without changing original chart recording
- Question bars can be color coded and easily configured through the Preference menu
- No Software Maintenance Agreement (SMA) or fees for updates!



LXSOFTWARE 11.2.2 AVAILABLE NOW

Contact Us for a Quotation or More Information

Phone: (765) 423-1505 sales@lafayetteinstrument.com www.lafayettepolygraph.com Find us on Facebook: facebook.com/lafayettepolygraph