-vs-

No. CS 2003-00026
(Supreme Court No. 27,915)

ELIZABET

HON. LOURDES MARTINEZ,

Respondents.

Findings of Fact and Conclusions of Law

Introduction:

Pursuant to Supreme Court Order issued in this matter, this Court is directed to enter findings of fact and conclusions of law. Given the tremendous volume of information presented by the parties as well as the testimony of several of the leading authorities on the issues decided, the Court has taken upon itself to provide an introductory section that includes an overview of the status of the law on polygraph examinations nationwide in both state and federal courts and a description of the polygraph examination process with the hope that it will assist the reviewing court. The findings of fact and conclusions of law follow these sections.

While many of the materials presented by both sides are worthy of note, a recent publication, The Polygraph and Lie Detection (PALD), a 2003 publication of the National Academy of Sciences (NAS), is particularly helpful. PALD focuses on the use of the polygraph in relation to employee screening. But since most of the research is in the area of event-specific investigations, its analysis of that research is highly useful in this context as well.

Another highly useful source is Faigman, The Law and Science of Expert Testimony (2002), or "Faigman". In Volume 2, § 19-2.0 is an article titled, The Scientific Status of Research on Polygraph Techniques: The Case for Polygraph Tests, by Honts, Raskin, and Kircher. Later, §19-3.0. is an article titled, The Scientific Status of Research on Polygraph Techniques: The Case Against Polygraph Tests, by Iacono and Lykken.

The Court recommends the two sources listed above for excellent overviews of some of the issues. In addition to the above, the parties to this action provided many exhibits, articles on nearly every aspect of polygraph examinations, studies relating to polygraph examinations, transcripts of testimony, and caselaw.

Without trying to oversimplify the issues presented, in evaluating the standards adopted in State v. Alberico, 116 NM 156, 861 P.2d 192 (1993), and restated in State v. Anderson, 118 NM 284, 881 P.2d 29 (1994), the testimony and arguments tended to gravitate to a number of key issues:

> First, whether there is a theory and whether it can be and has been tested. This includes the effect of base rates in determining reliability of test results in assisting the trier of fact and determining the balance between the probative value and prejudicial effect of the testimony;
>
> Second, whether the theory or technique has been subjected to peer review and publication;
>
> Third, whether there is a known potential rate of error in using polygraph techniques as well as whether there are standards that exist and are maintained that control the technique's operations;
>
> Fourth, acceptance of the test in the relevant scientific community; and,
>
> Fifth, whether the technique is based upon well-recognized scientific principles and whether it is capable of supporting opinions based upon reasonable probability rather than conjecture.

To the extent possible, the findings of fact will be set out in sections that will address each of these factors.

POLYGRAPH EXAMINATION PROCEDURES

A polygraph examination combines interrogation with physiological measurements made by the instrument, or polygraph. The instrument typically measures and records an examinee's heart rate, blood pressure, rate and depth of respiration and flow of electrical current at the skin surface as an examiner poses questions that require yes or no answers. Blood pressure is measured by a cuff over the biceps. Electrodermal activity (activity of the eccrine sweat glands) is measured by electrodes on the palm or on two fingers. Rate and depth of breathing are measured by pneumographs located on the chest and abdomen. Fluctuations in the heart and blood are recorded by a cardiosphygmograph, while a galvanometer records the body's electrical activity. [1]

The sensors attached to the examinee are connected to the instrument by wires. The data is recorded by analog or digital technology. Because the first analog instruments recorded the data with several pens writing lines on a piece of moving paper, the record of the examinee's physiological responses is known as the polygraph chart. [2]

The instrument does not measure or detect lies directly. Instead, proponents believe it measures physiological responses that are stronger when an examinee lies than at other times. A lie in response to a question may cause a reaction such as fear of detection or psychological arousal that changes heart rate, blood pressure, breathing rate, or skin conductance relative to what they were before the question was asked and relative to what they are after control questions are asked. [3]

Polygraph testing is used for three main purposes: 1. Screening of job applicants by law enforcement or other government agencies (preemployment screening); 2.

---

[1] NAS. The Polygraph and Lie Detection 12-13, 81 (2003)
[2] Id. at 13.
[3] Id.

Screening by agencies involved in national security of current employees; and 3. Investigating specific incidents, as in criminal cases.[4] When police conduct a polygraph test of a suspect, it is considered to be under adversarial conditions. In contrast, when defense counsel asks a client to take a privately administered test, it is called a "friendly" test. If the client passes the friendly test, defense counsel will often attempt to enter the results into evidence, and this is the more typical background for an evidentiary hearing like the present one.[5]

There are three major questioning techniques used in polygraph examinations: the relevant-irrelevant test (RIT), the guilty knowledge test (GKT), and the control question or comparison question test (CQT). The CQT's "are the most widely used techniques in criminal investigations and judicial proceedings."[6] Because the CQT is the most used test in criminal cases and because the tests in the instant cases were apparently CQT's, this Court's analysis will focus on that technique. Under Rule 11-707 NMRA 2003, tests using any of the three techniques would be admissible if that Rule's criteria were met.

The CQT tries to determine if the examinee is lying in response to a specific question or questions about the incident at issue (relevant questions). This involves comparing physiological responses to the relevant questions with physiological responses to control questions. Because the cuff on the arm begins to hurt after several minutes, a

---

[4] Id. at 11-12.

[5] William G. Iacono and David T. Lykken, *The Scientific Status of Research on Polygraph Techniques: The Case Against Polygraph Tests*, § 19-3.3.4 [5], in 2 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman, David H. Kaye, Michael J. Saks & Joseph Sanders eds., 2002)

[6] Charles R. Honts, David C. Raskin, & John C. Kircher, *The Scientific Status of Research on Polygraph Techniques: The Case for Polygraph Tests*, § 19-2.2.3 [1], in 2 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman, David H. Kaye, Michael J. Saks & Joseph Sanders eds., 2002)

limited number of questions, about ten, are asked to complete one chart.[7] Rule 11-707 requires that an examination include at least three charts.

Prior to the actual CQT, there is a pre-test interview. The examiner and examinee discuss the test, test procedure, examinee's medical history, and details of the test issues. Both relevant and control questions are reviewed, to minimize surprise and to ensure the examinee understands the questions. This portion of the examination may last from 30 minutes to 2 hours or more. [8] The expectation is that innocent examinees will react more strongly to control questions than to relevant questions, and guilty examinees will react more strongly to relevant questions. For example, a relevant question might be, "Did you rob the First City Bank?" Control questions are vague, cover a long period of time, and describe acts that most people have committed but are reluctant or embarrassed to admit during a polygraph exam. That is, if the examinee were suspected of theft, a control question could be, "During the first 22 years of your life did you ever take something that did not belong to you?"

Innocent people answer the relevant questions truthfully, but are expected to lie or be uncertain about their truthfulness when answering the control questions. That is, in these "probable-lie" control question tests, the instructions are designed to induce innocent people to answer "no" to control questions, even though most would then be lying. In contrast, guilty people are expected to be more concerned about failing the test because their answers to the relevant questions are lies, and they are likely to be more disturbed by the relevant questions, or so the reasoning behind CQT goes. Thus, the "art of the polygrapher lies in composing control and relevant questions that elicit the appropriate relative responses from truthful and deceitful parties." See State v. Porter,

---

[7] Iacono at § 19-3.1.1 [1].
[8] PALD at 253.

698 A.2d 739, 762 (Conn. 1997)(assuming without deciding that polygraph evidence met Daubert criteria but upholding per se rule barring its admissibility because prejudice outweighed probative value).

In another version of the CQT, the "directed-lie" test, examinees are instructed to lie to control questions such as, "Before 2002, did you ever make even one mistake?" The examiner tells the examinee that these questions will ensure that the examinee will be correctly classified as truthful or deceptive on the polygraph test to follow. Where the polygrapher in the probable-lie test chooses control questions during the pre-test interview to suit each examinee, the directed-lie control questions are a small set of simple questions that are "much easier to standardize." [9]

After the test, the charts are scored by a polygrapher or by a computer. Each relevant question response is measured against an adjacent control question response. Scores for each comparison range from +3 to -3. When the response to the control question is much stronger than to the relevant question, it is scored +3, indicating truthfulness. A score of -3 indicates a much stronger response to the relevant question relative to the response to the control question, indicating deception. If the two responses are about the same, the score is 0, with scores of $\pm$ 1 and $\pm$ 2 for intermediate values. The scores for all three charts are totaled. Examinees with scores of +6 or greater are considered truthful; those with scores of -6 or lower are deemed to be lying. Scores between +5 and -5 are inconclusive. The total score may range from approximately +30 to -30." [10] But see United States v. Galbreth, 908 F.Supp. 877, 894 (D.N.M. 1995), where the leading proponent of polygraph evidence, Dr. David Raskin, scored the defendant's

---

[9] Honts at § 19-2.1.2 [3]

[10] Iacono at § 19-3.1.1[2][*b*].

charts as +32. Charts may also be scored by computers using standardized algorithms, a relatively recent development.

ADMISSIBILITY OF POLYGRAPH EVIDENCE IN OTHER STATE COURTS

Eighty years ago, polygraph evidence was held inadmissible because it was not "sufficiently established to have gained general acceptance in the particular field in which it belongs." See Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). The standards for the admission of scientific evidence were changed by Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S.579 (1993), and many states, including New Mexico, adopted those standards. See State v. Alberico, 116 N.M. 156, 861 P.2d 192 (1993). Consequently, supporters of polygraph evidence sought its admission under the new standards. They have had little success before courts that have maintained pre-Daubert standards or courts that have adopted Daubert.

Twenty-seven (27) states and the District of Columbia apply a *per se* rule of exclusion of polygraph evidence for all purposes. See Pulakis v. State, 476 P.2d 474 (Alaska 1970); People v. Anderson, 637 P.2d 354 (Colo. 1981) (applying Frye, which Colorado abandoned in People v. Shreck, 22 P.3d 68 (Colo. 2001)); State v. Porter, 698 A.2d 739 (Conn. 1997); State v. Okumura, 894 P.2d 80 (Haw. 1995); People v. Sanchez, 662 N.E.2d 1199 (Ill.1996); Morton v. Commonwealth, 817 S.W.218 (Ky. 1991); State v. Harnish, 560 A.2d 5 (Me. 1989); State v. Hawkins, 604 A.2d 489 (Md. 1992); Commonwealth v. Mendes, 547 N.E. 2d 35 (Mass. 1989); State v. Anderson, 379 N.W.2d 70 (Minn. 1985); Weatherspoon v. State, 732 So.2d 158 (Miss. 1999); State v. Hall, 955 S.W.2d (Mo. 1997); State v. Staat, 811 P.2d 1261 (Mont. 1991); State v. Steinmark, 239 N.W.2d 495 (Neb. 1976); State v. Ober, 493 A.2d 493 (N.H. 1985); People v. Angelo, 666 N.E.2d 1333 (N.Y. 1996); State v. Grier, 300 S.E.2d 351 (N.C. 1983); Fulton v. State, 541 P.2d 871 (Okla. Crim. App. 1975); State v. Brown, 687 P.2d 751 (Or. 1984);

Commonwealth v. Brockington, 455 A.2d 627 (Pa. 1983); In Re Odeil, 672 A.2d 457 (R.I. 1996); State v. Hart, 911 S.W.2d 371 (Tenn. Crim. App. 1995); Tennard v. State, 802 S.W.2d 678 (Tex. Crim.App.1990); State v. Hamlin, 499 A.2d 45 (Vt. 1985); Robinson v. Commonwealth, 341 S.E. 2d 159 (Va. 1986); State v. Beard, 461 S.E.2d 486 (W.Va. 1995); State v. Dean, 307 N.W.628 (Wis. 1981), declined to follow on other grounds by State v. Davis, 645 N.W.2d 913 (Wis. 2002); Contee v. United States, 667 A.2d 103 (D.C. 1995).

These *per se* states ban polygraph evidence, including test results, offers to take the test, as well as refusals to take the test, for a variety of reasons. These courts found that the polygraph has not been proven valid or reliable or that it has not been generally accepted in the scientific community.[11] But a more salient reason for the outright ban is that the prejudice in a jury trial outweighs the probative value of corroborating a witness's credibility. See State v. Porter, 698 A.2d 739 (Conn. 1997)("State appellate courts, for whom Daubert is not mandatory authority, largely agree with our assessment that the prejudicial impact of polygraph evidence outweighs its probative value.") Id. at 773.

Four of the above states (Massachusetts, North Carolina, Oklahoma, and Wisconsin) had admitted polygraph evidence for years, but have since returned to a *per se* ban. See Commonwealth v. Mendes, 547 N.E.2d 35, 41 (Mass. 1989)(citing *inter alia* dangers of confusing jury and usurping jury's role and the "overwhelming authority throughout country") and State v. Dean, 307 N.W.2d 628, 653 (Wis. 1981)("Adequate standards have not developed in the seven years since [the decision to admit polygraph

---

[11] *The Legal Relevance of Scientific Research on Polygraph Tests, Per se* Exclusion § 19-1.2.1 in 2 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (David L. Faigman, David H. Kaye, Michael J. Saks & Joseph Sanders eds., 2002)

evidence on stipulation] to guide the trial courts in exercising their discretion in the admission of polygraph evidence. The lack of such standards heightens our concern that the burden on the trial court to assess the reliability of stipulated polygraph evidence may outweigh any probative value the evidence may have.")

Seventeen (17) states admit polygraph evidence at trial only when its admission is stipulated to in advance by all parties. See Ex Parte Hinton, 548 So.2d 562 (Ala. 1989); State v. Valdez, 371 P.2d 894 (Ariz. 1962); Holcomb v. State, 594 S.W.2d 22 (Ark. 1980); People v. Fudge, 875 P.2d 36 (Cal. 1994); Melvin v. State, 606 A.2d 69 (Del. 1992); Delap v. State, 440 So.2d 1242 (Fla. 1983); Fargason v. State, 467 S.E.2d 553 (Ga. 1996); State v. Fain, 774 P.2d 252 (Idaho 1989); Sanchez v. State, 675 N.E.2d 306 (Ind. 1996); State v. Losee, 354 N.W.2d 239 (Iowa 1984); State v. Webber, 918 P.2d 609 (Kan. 1996) Corbett v. State, 584 P.2d 704 (Nev. 1978); State v. McDavitt, 297 A.2d 849 (N.J. 1972); State v. Stevenson, 652 N.W.2d 735 (S.D. 2002); State v. Crosby, 927 P.2d 638 (Utah 1996); State v. Renfro, 639 P.2d 737 (Wash. 1982); Schmunk v. State, 714 P.2d 724 (Wyo. 1986).

In these states, stipulation usually means both parties agree prior to a subject taking a test that the results will be admissible and that the adversely affected party retains the right to cross-examine the polygraph examiner and otherwise to attempt to impeach the polygraph evidence. See, e.g., State v. Valdez, 371 P.2d 894 (Ariz. 1962). Generally, these appellate decisions do not claim that the evidence is probative or becomes reliable due to the stipulation. See Delap v. State, 440 So.2d 1242, 1247 (Fla. 1983). Some courts, however, have concluded that the stipulation makes the test reliable

– it raises the examinee's fear and leads to the selection of more impartial examiners, tending to produce more accurate results.[12]

Two (2) other states admit stipulated results but in limited circumstances. See State v. Yodsnukis, 281 N.W.2d 255 (N.D. 1979)(post-trial proceedings) and State v. Souel, 372 N.E.2d 1313 (Ohio 1978)(for corroboration or impeachment only).

Louisiana and Michigan allow the admission of polygraph evidence without stipulation but only in post-trial proceedings. See State v. Catanese, 368 So.2d 975 (La. 1979) and People v. Barbara, 255 N.W.2d 171 (Mich. 1977).

South Carolina generally bars admission of polygraph evidence, but the decision is now left to the discretion of the trial judge after a hearing applying Rules of Evidence 702 and 403. See State v. Council, 515 S.E.2d 508 (S.C. 1999).

ADMISSIBILITY OF POLYGRAPH EVIDENCE IN FEDERAL COURTS

United States v. Scheffer, 523 U.S. 303 (1998) held that military courts' *per se* rule excluding polygraph evidence did not violate a defendant's right under the Fifth or Sixth Amendment to present a defense. Beyond this holding, the decision lacks precedential value, given the fractured makeup of the Court's three opinions.

In contrast to the majority of state courts, only two federal circuits have a *per se* rule barring admissibility. See United States v. Prince-Oyibo, 320 F.3d 494 (4th Cir. 2003), Petition for Certiorari Filed, (July 11, 2003)(NO. 03-5297) and United States v. Skeens, 494 F.2d 1050 (D.C. Cir. 1974)(citing the Circuit's decision in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Most federal appellate courts leave admission of polygraph evidence to the discretion of the trial courts, but generally such evidence is excluded on the basis of

[12] Faigman et al. at § 19-1.2.3, fn. 73 and 74.

Daubert/Rule 702 or Rule 403 or both. See United States v. Black, 78 F.3d 1, 7 (1st Cir. 1996)(generally inadmissible); United States v. Santiago-Gonzalez, 66 F.3d 3, 6 (1st Cir.1995)(admissible if agreed to in plea bargain); United States v. Kwong, 69 F.3d 663, 668 (2nd Cir.1995)(balancing test under Rule 403); United States v. Lee, 315 F.3d 206, 214 (3rd Cir. 2003)(noting lack of per se exclusionary rule and admissibility to rebut claim of coerced confession but declining to rule on admissibility at trial or revocation hearing), Petition for Certiorari Filed, (June 2, 2003)(NO. 02-11166); United States v. Posado, 57 F.3d 428, 434 (5th Cir.1995)(must meet Rule 702 and Rule 403 standards); United States v. Sherlin, 67 F.3d 1208, 1216-17 (6th Cir. 1995)(Rule 403 standard, but results generally inadmissible, especially if unstipulated); United States v. Lea, 249 F.3d 632, 640 (7th Cir. 2001) ("[W]e continue to hold that a district court need not conduct a full Daubert analysis in order to determine the admissibility of standard polygraph evidence, and instead may examine the evidence under a Rule 403 framework. Nonetheless, we posit that the factors outlined by the Supreme Court in Daubert remain a useful tool for gauging the reliability of the proffered testimony, as reliability may factor into a 403 balancing test.").

See also United States v. Williams, 95 F.3d 723, 729-30 8th Cir. 1996)(suggesting non-stipulated evidence may be admissible under Daubert if Rule 403 is met) and United States v. Waters, 194 F.3d 926 (8th Cir. 1999)(Daubert hearing unnecessary where 403 not met despite defendant passing test requested and given by prosecution); United States v. Cordoba, 194 F.3d 1053 (9th Cir. 1999)(must meet 702 and 403); United States v. Call, 129 F.3d 1402 (10th Cir. 1997)(evidence properly excluded under 403 where requested Daubert hearing not held); United States v. Gilliard, 133 F.3d 809 (11th Cir. 1998)(Honts-administered polygraph inadmissible under 702 and under 403).

"Leaving discretion to trial courts rather than prescribing a *per se* rule does not seem to have changed practice substantially."[13] That is, "even when presented with an opportunity to admit polygraph evidence, most [federal] district courts are decidedly reluctant to do so." See State v. Porter, 698 A.2d 739, 776-77 (Conn. 1997).

One rare case admitting polygraph evidence was United States v. Galbreth, 908 F.Supp. 877 (D.N.M. 1995). In Galbreth, Judge Vasquez admitted the expert opinion testimony of Dr. Raskin, the nation's leading supporter of the validity of polygraph evidence, after finding it met the reliability criteria of Rule 702 and Daubert as well as being more probative then prejudicial under Rule 403. Dr. Raskin had given Galbreth a polygraph test, which the court described as "a properly conducted examination by a highly qualified, experienced, and skillful examiner." Id. at 896. However, this ruling carries little weight due to its procedural placement.

The judge ruled from the bench after a hearing in March, 1995. In July, 1995, the case went to trial. At the conclusion of the Government's case-in-chief, the Government dismissed the charges (income tax evasion). Galbreth's polygraph evidence was never presented to the jury. On October 4, 1995, the judge issued a "Memorandum Opinion and Order" that detailed her ruling on the admission of the polygraph evidence. The Order was therefore unappealable and *dicta*.

State v. Porter, 698 A.2d 739, 777, n. 76 (Conn.1997) described Galbreth this way:

> The most substantial of the few federal opinions permitting polygraph evidence at trial comes from the District Court of New Mexico. United States v. Galbreth, supra, 908 F.Supp. 877. The Tenth Circuit Court of Appeals had only addressed the question of polygraph admissibility before Daubert had been released; see United States v. Soundingsides, 820 F.2d 1232, 1241-42 (10th Cir.1987); so the court in Galbreth felt free to formulate its own standard. The

[13] Faigman et al. at § 19-1.2.2.

> court accepted that Daubert provided the proper threshold standard; id., at 878; and then relied largely on testimony by Raskin to conclude that polygraph evidence satisfied Daubert and rule 403 of the Federal Rules of Evidence. Id., at 895. Although the court in Galbreth did address many of the concerns that have motivated us to retain our per se rule of exclusion, it did so by recounting only the most propolygraph studies and information. Id., at 885-93. We believe that a more balanced review of the polygraph literature, such as we have conducted in the present case, reveals substantially more uncertainty regarding the effectiveness and prejudicial impact of the polygraph test than the court in Galbreth acknowledged.

Dr. Raskin scored the test as +29, and Dr. Honts scored it as +32, indicating a high probability of truthfulness. The Government's expert, Dr. Barland, found the charts to be inconclusive. Galbreth at 894.

A critical issue was whether Galbreth knowingly failed to report income. Had Dr. Raskin testified, he would have been permitted to state that Galbreth's "answers to the *relevant* questions regarding his knowledge and intent [were] consistent with a truthful polygraph outcome." Id. at 895. (Emphasis added.) As the judge put it:

> Dr. Raskin concluded that Defendant was truthful in his statements that he did not realize his returns under reported his taxable income. At trial, Defendant intends to call Dr. Raskin as an expert witness to testify about the testing procedures, to explain how the test was evaluated and to explain his interpretation of the results. Dr. Raskin is expected to testify that the results are indicative of a truthful polygraph test outcome with regard to the relevant questions. Dr. Raskin will not testify as to his personal opinion that Defendant was in fact telling the truth.

Id. at 878.

The testimony would therefore not be limited to Galbreth's credibility but would cover his substantive answers to questions concerning his guilt or innocence. The judge would have allowed the assistant U.S. Attorney to cross-examine Dr. Raskin and to present the Government's expert to "refute any of Dr. Raskin's testimony relating to the polygraph technique in general or to the specific application of that technique in this case." Id. at 896. There was no mention of permitting the Government to give Galbreth a polygraph exam.

By contrast, another district court in United States v. Crumby, 895 F. Supp. 1354, 1363 (D.Ariz. 1995) admitted the evidence with severe limitations while noting that "the prejudicial effect of permitting the jury to hear the specific responses to the question of whether Defendant committed the ultimate crime in the case is overwhelmingly prejudicial." That is, Crumby could introduce evidence that he took and passed the test if (1) he gave notice to the prosecutor, (2) took a government-administered test, (3) introduced the evidence only to support his credibility, if attacked, under Rule 608(a), and (4) the specific questions and physiological data were not introduced into evidence, although the general nature of polygraphy could be discussed by the experts under Rule 702. Id. at 1365. In Crumby, Dr. Raskin again testified, but unlike the Galbreth prosecutor, the U.S. Attorney did not offer any expert testimony as to the validity of the theoretical basis for the polygraph, nor contest Dr. Raskin's testimony regarding the known error rate. The Crumby decision failed to mention any of the studies that challenge the validity of polygraph tests.

Galbreth and Crumby are exceptions, even within their own federal circuits, to the general rule that polygraph evidence is not admitted in federal courts. See United States v. Call, 129 F.3d 1402 (10th Cir. 1997) and United States v. Cordoba, 194 F.3d 1053 (9th Cir. 1999)(barring evidence under Rule 702 due to lack of known error rate for real life exams, controversy in scientific community regarding validity of theory behind test, and lack of controlling standards).

FINDINGS OF FACT

Decision theory and base rates

1. Measuring validity of polygraph test results is crucial to determining their admissibility. The following definitions come from PALD, page 29, *et seq.*

2. *Decision theory* is a scientific approach that applies basic statistics to real world problems. It is used to attempt to predict the utility of a test when there is a high degree of uncertainty before a test is conducted.

3. *Reliability* is a term used to indicate repeatability across different times, places, subjects, and conditions.

4. *Test-retest reliability* is the extent to which the same procedure, including the examiner, test format, and equipment used to examine the same subject for the same purpose yields the same result on repetition.

5. *Inter-rater reliability* is the extent to which different examiners would draw the same conclusions about a given subject at a given time for a given examination.

6. A measurement is considered *valid* if it measures what it is supposed to measure.

7. *Criterion validity* refers to how well a measure captures what it is supposed to capture. In the case of a polygraph test, does it show deception when the test subject is in fact deceptive and show lack of deception when the subject is truthful. This is synonymous with *accuracy*.

8. Without accuracy or criterion validity no test or procedure can be considered valid.

9. *Construct validity* refers to how well explanatory theories and concepts account for performance of a test. Users can have greater confidence in a test when evidence of its accuracy is supported by evidence of construct validity. In other words, when there is a chain of plausible mechanisms that explain both the empirical findings of the test and evidence that each test mechanism operates as the theory prescribes.

10. A *positive* polygraph test result means that the test indicates deception. A *negative* polygraph test result means that the polygraph indicates no deception. Therefore, a *false positive* result means the test indicates deception when the test subject is being truthful and a *false negative* result means the test indicates no deception when the test subject is not being truthful.

11. *Decision threshold* is the cutoff point for deciding whether a result is positive or negative. Even though polygraph test results, like other diagnostic tests, are usually presented in a yes or no answer format, the actual score is not presented in that fashion. In other words, there is a cutoff point, below which or above which the test is not scored as a positive or negative. These cutoff points are policy choices made by polygraphers. If they are set incorrectly, it increases the chance for a false negative or false positive result.

12. The literature and the presentations focused to a great extent on the issue of base rates. Base rates are an essential element in establishing a level of confidence in

the outcome of a diagnostic test. Base rates dictate whether a diagnostic test is worth considering at all.

13. *Base rate* refers to the proportion of people in a population as they relate to a particular trait in issue. For example, in polygraph testing, the percent of truth tellers versus deceivers would result in the base rate. While the cases refer to the rate of error, that is not the only number that a court should consider in determining admissibility under Rule 11-403 NMRA 2003. Even though a particular piece of information may have some slight tendency to make the existence of a fact of consequence more or less probable, the confidence one could have in that information in relation to the circumstances of the case may be so low as to render the evidence inadmissible under Rule 11-403 NMRA 2003.

14. The confidence level in decision theory is a function of the error rate and base rate. To be complete in evaluating any diagnostic test, accuracy has two components. In the polygraph context, these components are: How likely is the test to be positive (indicating deception) if lying is present; and, how likely is the test to be negative (indicating a lack of deception) if lying is not present.

15. In the world of medicine, for example, Dr. Zelicoff noted that in diagnosing strep throat that the disease is seasonal. During certain seasons, strep is so rare that the test result does not significantly add to our confidence level. That's because due to seasonal fluctuation, the base rate of possible strep is so low, that even though the test accuracy is high, a positive test result does not increase our confidence that a decision made based on the test result will be correct.

16. In polygraph use, knowledge of the base rate can help decide whether the result of a polygraph test is worthy of consideration in making an important decision. In the employee screening contest, the NAS focused on base rate since the percentage of spies is assumed to be very low. Dr. Zelicoff quoted the former Secretary of Energy as saying 1 in 10,000 employees of the Department of Energy are spies.

17. The accuracy rates of polygraph examinations are, at best, debatable in real life contexts. However, even if one assumes a high accuracy rate, the test is of little utility because of the low confidence level in the test result.

18. The NAS noted that if you use a test with 90% accuracy and an 80% threshold value (see p.61, PALD) and the test is used in a population with .1% (one in 1000) spies, the test would identify an average of 1606 as deceptive, only 8 of whom would be spies. PALD p.47.

19. Dr. Iacono used a similar example to illustrate the problem as it might apply in the criminal context. If you assume a base rate of 90% guilty and 90% test accuracy (and a maximum threshold value) and apply those assumptions to 100 criminal defendants who take polygraph tests, the resulting confidence level in the test result is notable. Of the 90 guilty, 81 will fail the test and 9 will pass. The 81 test failures will not be disclosed to the jury, the court or the prosecution, of course, but the 9 passed tests will be disclosed. Of the innocent, 9 will pass and 1 will

fail. The passes will be disclosed and the one failure will not. Of the 18 passed tests, there are only 9 (50%) who are factually not guilty. In other words, the confidence level of the test in its application is only 50-50. See Resp. Exhibit 4.

20. Petitioners have some arguments to address this illustration. First, they note that the base rate is not truly knowable. A defendant is, after all, presumed innocent and to clump an individual in with all others accused is to violate basic principles of American jurisprudence. Second, petitioners argue that the standard under Rule 11-401 NMRA 2003, is any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

21. The argument points out that even though the confidence level of the test result in the context of these assumptions is only 50-50, it still makes a fact in issue more or less probable. In other words, even though the confidence level is merely 50%, the argument goes, it is still a 40% improvement over the pre-test 90% figure. To rephrase, before the test result, given the population, one could be confident that any one of the population who denied culpability was 90% likely to be not telling the truth. After passing the test, one could be only 50% confident that the denial was untruthful. That move from a 90% confidence the testimony is false to a 50% confidence the testimony is false makes it more probable it is truthful than it was before and, so the argument goes, it is relevant.

22. The base rate issue is part of this Court's analysis of the field study reliability and is a major issue raised directly by Respondents. Therefore its effect must be considered as it relates to polygraph evidence. This Court finds that, *if* polygraph testimony is reliable enough to be admissible, it would be deceptive to testify to the type of testimony offered in the past, such as claims that there is a 90% chance the test subject was truthful or that the test is 90% accurate.

23. Dr. Raskin and Dr. Honts both testified that in the absence of a known base rate, a base rate of 50% should be assumed. Both also testified that juries tend to work out their own base rates. In other words, in considering the strength of other evidence, juries give more or less weight to polygraph evidence.

24. The Court agrees that the base rate in an individual case is basically either unknowable or, at best, is a moving target based on the strength of all of the non-polygraph evidence. Yet it exists. To assume a base rate of 50% is no more reliable than any other assumption and is misleading. If any level of accuracy is testified to, it is either directly or inferentially suggestive of a confidence level in the result that is directly tied to a base rate most appropriately to be determined by the finder of fact. If the art of polygraphy were to ever achieve sufficient reliability for admissibility, it would be appropriate to prohibit any percent of accuracy to be introduced on direct examination. In other words, it would be inappropriate to testify that the test reflects a 90% probability that the test subject was truthful if it is not possible to accurately express how confident the jury could be in that number given the population of test subjects. Any probative value of such testimony would be substantially outweighed by the danger of confusion of the issues, misleading the jury, and undue waste of time.

Known rate of error in operation

25. The only way to determine the "rate of error in operation" of the polygraph test procedure is to test the operation of the procedure and determine its reliability or accuracy.

26. To test a theory, one must start with a hypothesis.

27. There is no sound scientific theory upon which polygraph is based.

28. Dr. Honts claims to have a hypothesis that is being tested, that of whether a comparison question test accurately diagnoses truth and deception. However, there is no explanation as to why it does so if it does indeed do so.

29. There is no lie response. There is no one testable physiological manifestation of a lie. Polygraphs test physiological responses to questions and, if there is a physiological response, the thinking is that if the response is greater for a relevant question than for a comparison question, then it means the response to the relevant question is likely to be deceptive. However, any physiological response to any question could be caused by any one of a number of emotions such as shame, anxiety, guilt, fear, tension, or other emotional responses not understood. There is no single underlying process reflected in responses to questions that are measured by the polygraph. The polygraph measures a variety of psychological and physiological processes, including some that can be consciously controlled.

30. In the comparison question test, one emotional or physiological response to the relevant question could cause a measurable result on the polygraph and a completely different emotional or physiological response to the comparison question could cause a measurable result on the polygraph. Yet the level of response for each of the two responses is what is measured and compared, resulting in the gauge of truth-telling.

31. The comparison questions are not determined in advance and are either directed lie or probable lie questions. A directed lie means in the pre-test interview the test subject is told to lie to the question which will supposedly result in the physiological response. A probable lie is similar in operation, but is a question like: "Have you ever taken anything of value that did not belong to you?" Pre-test procedures sometimes include card tricks or similar techniques to convince the test subject that the test is working and will detect deception. No standards exist for how the pre-test procedures will be conducted or for how the comparison question will be formulated.

32. The vast majority of the tests upon which the claimed accuracy of polygraph examinations is based are laboratory tests, as opposed to field tests.

33. In most laboratory tests, the subject is given a series of written instructions and during the course of following those instructions will or will not "steal" an item. Then the subject is immediately subjected to a polygraph examination.

34. In most field tests, results of polygraph examinations by various law enforcement agencies are examined to determine if they were correct.

35. The accuracy of a test in the field can only be determined if objective truth is known. If objective truth is not known, then you can not determine if the test accurately detected deception.

36. The method for determining objective truth in field tests is usually based on whether or not there was ultimately a confession either by the subject of the polygraph or by others who then exonerate the test subject. If nobody confesses, then the test result is not considered in determining accuracy.

37. This technique effectively limits the ability to measure polygraph accuracy in the field, since all test results are thrown out if there is not a confession. It is highly unlikely that subjects in a field study would confess if they passed the polygraph. A fair assumption is that a guilty subject would have a vested interest in passing the polygraph. That is one of the ideas proponents assert to argue that the stress of facing the relevant question would result in a more pronounced response than the control question. If it's so important to pass, why would anyone who's successfully passed the polygraph in a real life setting then decide to reveal the truth? Why would the subject bother taking the polygraph in the first place if the point wasn't to try to get away with it? If that assumption is correct, and this Court, based on years of experience on the bench and in a criminal practice, as well as after reviewing all of the evidence and testimony in this case, finds that it is, field studies do not produce a reliable error rate. None of the errors are likely to admit they were "errors".

38. Conversely, the truly innocent person who is scored as having failed the polygraph examination is also highly unlikely to confess to the crime they did not commit. Again, this error would not reach the final tally of test "success" since the result would not be considered at all as there was no confession. If the innocent person falsely confessed, which appears to happen from time to time, that would also inflate the accuracy figures of the field study and distort the claimed error rate.

39. Experimental field studies are the most compelling type of field validation study. This would be a study in which a variable of interest is manipulated among polygraph examinations in real-life settings. No experimental field studies are found in any of the literature on polygraph validity. PALD at 109-110.

40. At the top of research hierarchy is the peer reviewed publication. No specific-incident field investigations are found in the higher levels of research hierarchy. PALD at 114.

41. The field test results suggest that polygraph examinations are an effective interrogation tool because they seem to produce a significant number of confessions. This utility is separate from polygraph validity. According to NAS: "There is substantial anecdotal evidence that admissions and confessions occur in

polygraph examinations, but no direct scientific evidence assessing the utility of the polygraph. Indirect evidence supports the idea that a technique will exhibit utility effects if examinees and the public believe that there is a high likelihood of a deceptive person being detected and that the costs of being judged deceptive are substantial. . . . there is no evidence to suggest that admissions and confessions occur more readily with the polygraph than with a bogus pipeline – an interrogation accompanying the use of an inert machine that the examinee believes to be a polygraph." PALD at 214-215.

42. Because there is no underlying theory explaining why polygraphs detect deception, it limits the ability to determine effectiveness in contexts that vary from the lab settings or the limited number of field tests. For example, the majority of polygraph test results offered in evidence in New Mexico (all of the test results in the cases in issue in these appeals) are offered by the defendant.

43. Because laboratory tests are so dissimilar from the complex matrix of variables that can occur in real life, they are not sufficiently useful for determining the accuracy of polygraph testing in real life contexts.

44. The context of a polygraph test offered by a defendant differs in many material ways from the lab setting and field tests. First, the delay between the targeting of the suspect and the test is often significant. Second, the pressure to perform is different since the result of the test will not be disclosed if the defendant fails the test. Third, given the delay, the defendant may become habituated to answering questions about the pending charges and therefore may not react as strongly to relevant questions during the polygraph test. Fourth, the polygrapher is "friendly" to the defense. Fifth, the opportunity for the defendant to learn and utilize counter-measures is increased.

45. An example of the types of problems that are inherent in most laboratory studies was demonstrated by a laboratory study conducted by Dr. Iacono which was designed to introduce some real stressors into the test dynamic, stressors that are more likely to mimic real life situations. Dr. Iacono went to a population that Dr. Raskin used for one of his lab studies, prisoners. But instead of using the traditional Raskin approach of offering a nominal financial reward if the test is "beaten", Dr. Iacono generated some real pressure. He told the prisoners that he would pay them if they "beat" the polygraph, but that the payment would be to all of the prisoners or none. He told them that he expected a certain percentage to be successful and that if they fell below that percentage nobody would get paid and he would publish the names of the prisoners who failed to pass the polygraph in the prison. At the conclusion of the test he paid everyone and didn't publish any names. However, the test accuracy fell from Dr. Raskin's 94% to 72%, even though it was the same population group. As Iacono described it, he set up a group contingency threat, where each test subject would be concerned about the consequences of the test outcome. The study was published in The Journal of Applied Psychology, a peer reviewed publication. TT, 6/24/03, 46-48.

46. The Iacono prisoner study is one example of what can happen if a key and relevant variable is altered to more closely approach real life. Unfortunately, there are not enough studies that try to answer these types of questions.

47. No scientific field studies of the friendly polygrapher scenario have been conducted. Given the variables, the risk of significant impact is great. In the normal scenario, the scenario from which the field studies have been derived, the test is conducted in an adversarial setting. The goal of the police officer conducting the test is to catch somebody. The focus is intense and the consequences of failing the polygraph are great.

48. In the friendly polygraph there is no adversarial atmosphere.

49. The Rosenthal Effect is a phenomenon that has been recognized in psychology for approximately thirty years. It recognizes that psychologists and scientists and others who have an investment in a theory are likely to unconsciously arrange an experiment in such a way that they get favorable results. It is the reason that it is necessary that test results need to be replicated by an independent researcher.

50. The Rosenthal Effect can affect an individual polygraph examiner because the hypothesis in an individual test involves the examiner's sense of whether the test subject is guilty or not. The examiner necessarily has access to the case facts and interviews the examinee in a pre-test interview. Based on the case information and how the interview develops – for example the examinee might seem truthful – it can affect the attitude of the examiner. The Court noted the following statement from Dr. Honts: "In my experience in New Mexico in testifying before juries clearly indicates that, (the jury will make use of the polygraph as they see fit) and that they have decided to convict despite a polygraph that showed the person was truthful." TT, 7/3/03, 114. The context of the statement and the observation of the witness led the Court to conclude that Dr. Honts was invested in the outcome and that he was surprised that a jury could reach a different conclusion.

51. The risk of the Rosenthal Effect is exacerbated by the lack of standards in the profession.

52. There is no requirement that the test subject be drug free. However, drugs that act to decrease responding in a general way will not normally affect the control question test because the scoring is based on comparing responses to two types of questions. The problem is, there is at least one study that indicates that alcohol could reverse the responses in a control question setting. Dr. Iacono was unable to duplicate the result of the study. More research needs to be done in this area.

53. Since it is not clear what emotional triggers will result in a particular reading in a polygraph chart and since different emotions may produce a given polygraph response in the control versus the relevant question, there is no way to determine if the drug may affect one emotional response, but not another.

54. There are no standards which dictate whether an examiner should use a probable lie versus a directed lie versus a relevant-irrelevant test.

55. There is no restriction regarding testing mentally ill individuals. However there is at least one study that indicates that psychopaths are not more able to defeat the polygraph than others.

56. While there are supposed guidelines that dictate the form of relevant question, they seem to be subject to unreasonable interpretation by practitioners. Dr. Raskin, on the one hand takes the position that intent is not a proper subject for a relevant question, yet claims that asking a relevant question regarding whether touching the victim's penis was for "sexual purposes" is not problematic. TT, 7/1/03, 217-218. (Regarding the questions asked in State v. Robinson, one of the pending cases).

57. At this point there remains no licensing requirement for polygraphers in New Mexico.

58. There is no blind proficiency testing requirement in New Mexico.

59. Covert counter-measures consist of simple techniques such as biting the tongue, flexing the toes, or performing mentally stressful math exercises. These activities, if timed to take place during the control question phase of the test, can artificially augment the "involuntary" physiological response.

60. Counter-measures are effective in affecting polygraph test outcomes. One laboratory study indicates that with less than a half hour training or explanation, the likelihood of a false test result increases by 50%. There is a consensus among scientists that counter-measures are effective. Some studies indicate that merely reading about countermeasures is insufficient to affect test outcomes, but more research is necessary in this area. See, State v. Porter, 241 Conn. 57, 113, 698 A. 2nd 739, 768 (1997).

61. This Court shares the concern of the Connecticut Supreme Court in Porter, noting the informal study cited in that case where twenty-seven inmates were given fifteen minutes of instruction by a fellow prisoner (who had been instructed by Dr. Lykken) before reporting for a polygraph exam regarding an alleged infraction of prison rules. All twenty-seven privately admitted their guilt and twenty-four passed the polygraph. Id., at 241 Conn. 114, 698 A.2d 768. Although that study is appropriately criticized by Dr. Raskin, see. Faigman, § 19-2.2.2 FN 72, the specter of the ease of communicating how to successfully utilize counter-measures remains.

62. Experienced examiners could not detect counter-measures in the lab study.

63. There are no properly conducted studies regarding the effectiveness of counter-measures in real life by sophisticated test subjects.

64. In PALD, the authors note: "Notwithstanding the limitations of the quality of the empirical research and the limited ability to generalize to real world settings, we conclude in populations of examinees, such as those represented in polygraph

research literature, untrained in counter-measures, specific instance polygraph tests for specific investigations can discriminate lying from truth well above chance and well below perfection, and accuracy may be highly variable across situations ." Id. at 214.

65. However, there is no guarantee that the populations of test subjects that are likely to offer the test in evidence in New Mexico are "untrained in counter-measures." Also, it must be kept in mind that the context of all of the research referred to was in relation to specific investigations in either laboratory settings or field studies based on adversarial test situations. As a result, the conclusion that tests in those situations can discriminate lying from truth "well above chance" is irrelevant to the inquiry of this Court.

66. Computer scoring of test results is a recent development. However, the algorithms for the programs are based on certain assumptions:

- that the probability of truth or deception in real-world situations can be determined from the score on a control question test (the basic assumption of lie detection);
- that the scores stored in the computer accurately represent the scores to be expected from truthful or deceptive subjects obtained under circumstances similar to those in the instant test;
- that 50 percent of those who are tested with the instrument are deceptive (the base rate problem discussed elsewhere)

See, Faigman, § 19-3.3.9. Because of the problems with field studies no database meeting the above criteria exists. The computer scoring results in an expressed confidence level presented as a percent likelihood that the test subject is truthful. Examiners will testify, for example, that the test score shows the likelihood that the subject was truthful is 93.3%. As discussed above, this is without a scientific basis and deceptively ignores the problem with base rates.

Acceptance in the Relevant Scientific Community

4. The relevant scientific community is The Society for Psychophysiological Research and Fellows in Division One of the American Psychological Association, a division of the American Psychological Association General Psychology Group broadly versed in principles of psychology.

5. There have been four attempts to survey the relevant scientific community for its views of the validity of polygraph examinations.

6. Of the four attempts, the most reliable is the survey conducted by Dr. Iacono and published in The Journal of Applied Psychology, a peer reviewed publication.

7. While Dr. Honts is critical of the methodology, the response rate was the highest by far, and the survey clarified potential ambiguities found in the other surveys.

The Court finds it significant that the article relating to the Iacono survey and the results were selected by the publisher of a book on research methodology to be used as an exemplar of how to do similar types of research. Further, unlike the Iacono survey, the other surveys did not distinguish between control question tests and guilty knowledge test.

8. 36 % of those responding felt the control question polygraph test was based on scientifically sound psychological principles and theory. This compares with 22% who agreed with that statement regarding the directed lie test and 77% who agreed with the question in the guilty knowledge test.

9. A significant majority also agreed that a "friendly" test was more likely to be passed than an adversarial test. 99% believed that counter-measures might work.

10. On the issue of the weight to be given laboratory studies as opposed to field studies, only 17% believed that results of laboratory studies should be given substantial weight.

11. The Iacono survey results were consistent with the NAS view that the high levels of accuracy claimed by practitioners have rarely been reflected in empirical research. NAS, p. 107.

12. Control question polygraph tests do not enjoy general acceptance within the relevant scientific community.

13. This finding is even more significant given the length of time the polygraph has been in use. The polygraph is not "cutting edge" technology that would tend to be esoteric. It is technology that would be familiar to members of The Society for Psychophysiological Research and Fellows in Division One of the American Psychological Association.

CONCLUSIONS OF LAW

1. Polygraph test results and the conclusions derived from them are not based upon an overarching theory. To the extent it is merely argued that there is a hypothesis that the test reliably detects deception, that hypothesis has not been subjected to field research. The existing laboratory research, given the problems described above, is woefully inadequate to support admissibility in court in real life contexts.

2. There is no theory, as stated above. The technique has been subjected to limited peer review publication. The conclusions of the relevant publications do not enhance confidence in the test results, particularly considering the effectiveness of counter-measures.

3. The potential rate of error is vague and unreliable. Given the effect of ignoring base rates as endorsed by proponents, the reliability of test results as reflected in an actual percentage misrepresents the confidence level in the test.

4. There are no set standards other than those set out in Rule 11-707 NMRA 2003. Those standards are insufficient for the reasons set out above.

5. Control question polygraph tests are not accepted in the relevant scientific community at a significant level, particularly considering the age of the technique.

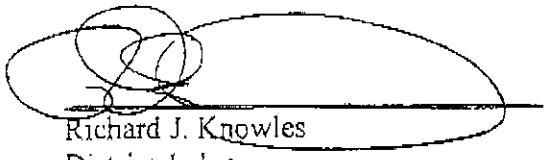
6. The technique is not based upon well-recognized scientific principles and is not capable of supporting opinions based upon reasonable probability rather than conjecture.

7. If the risk of counter-measures is ignored, there is an argument that all of the studies taken together support a conclusion that a successful polygraph result makes a fact in issue more or less probable. However, given the state of the art of polygraphy, the limited probative value polygraph test results is substantially outweighed by the danger of confusion of the issues, undue delay, and waste of time and therefore polygraph evidence becomes inadmissible under Rule 11-403 NMRA 2003.

8. At least one court has found that testimony that someone has passed a polygraph examination is extrinsic evidence of a specific instance of conduct (passing the polygraph) that supports a witness's credibility, and is therefore inadmissible under Rule 11-608 B. US v. Piccinonna, 729 F.Supp. 1336, 1338 (S.D.Fla. 1990), aff'd by U.S. v. Piccinonna, 925 F.2d 1474 (11th Cir. 1991).

9. Because of the inherently subjective nature of the test procedure, the polygraph examination can not be repeated. Successful repetition of a test is the cornerstone of the scientific method. It lacks test-retest reliability.

10. The results of polygraph testing are not sufficiently reliable for admissibility in courts in New Mexico.



Richard J. Knowles
District Judge