

Utility and Numerical Evaluation of the Guilty Knowledge Test

Keith H. Gaines¹

Abstract

Eighty subjects were given polygraph examinations utilizing the Guilty Knowledge Test (GKT) in an effort to test the utility of that procedure in a mock crime situation. Additionally, a semi-objective numerical scoring system was implemented, modified, and evaluated. The subjects were broken down into two equal groups. The first group was programmed to commit the mock crime (a burglary/homicide). The second group did not commit the mock crime and was given no information regarding the details of the crime. Half the sample (twenty innocent and twenty guilty), was utilized to design the scoring system. It was then tested on the remaining half. The GKT with the developed numerical scoring system correctly identified fifteen of twenty programmed guilty subjects and seventeen of twenty programmed innocent subjects, with one inconclusive decision. The overall accuracy rate was 82%.

There is a deep sense of mystery surrounding the Psychophysiological Detection of Deception (PDD) as a forensic procedure. In fact, the terms "Lie Detector Test," are often used which results in PDD being viewed as a kind of "crystal ball" believed capable of reading the minds of those unfortunate enough to be subject to its cold inquiry.

The use of a polygraph instrument during interrogation has long been a source of great controversy in our criminal justice system (*Frye v. United States*, 1923; *United States v. Piccinonna*, 1988). This controversy is fueled by all the fear and distrust directed towards anything that threatens the privacy of our personal lives, thoughts, or actions. To many, PDD constitutes just that threat.

What is Polygraph?

The term "polygraph" or "polygraphy," the traditional terminology used when referring to PDD, is actually a misnomer. Taken from the Latin, "poly" meaning "many" and the word "graph" which means "to write," polygraph literally translates to "many writings." Therefore, the term "polygraph" should be reserved for the instrument used in PDD, and not the actual forensic testing procedure.

Polygraph instruments are not limited to lie detection and are, in fact, used for many types of physiological recordings. The polygraph generally used for PDD is a multi-channel physiological recording instrument which monitors respiration, cardiovascular activity, and perspiration activity of the person being examined (Iacono, 1988).

¹ Editor's Note: This report was originally released in August 1992. Due to space considerations the appendices have not been included in this reprinting. The full report is available at no cost at www.dtic.mil.

Acknowledgments

My profound gratitude goes to Dr. William J. (Bill) Yankee, without whom this study simply would not have happened. He taught me that anybody can find the answer to a question if they look hard enough. I am likewise grateful to Dr. Gordon Barland and Mr. John Schwartz, whose support and encouragement kept this project going through the toughest times, to Sam Pena, Greg Janz, David Sanders, and Tony Krum, the examiners of the U.S. Marine Corps Criminal Investigation Division, who worked long and hard collecting the data for this project, and to Dr. Barbara Carlton, who's expert advice regarding research design made everything come together. This study was supported by DoDPI92-P-0009 project funds from the Department of Defense Polygraph Institute. The views expressed in this article are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

The PDD examination as it is generally used today, is a systematic, highly structured procedure in which a Forensic Psychophysiol-ogist, traditionally referred to as a "Polygraph Examiner" or "Polygrapher" uses a polygraph instrument to monitor physiological activity while he asks the subject a number of questions.

There are a number of forensic PDD procedures. For many years, PDD examiners have been using various versions of the Probable Lie Control² Question Test (CQT). These versions have been the mainstay procedures for the majority of examiners conducting criminal issue examinations (Lykken, 1988).

The concept behind the CQT is a simple one. After an extensive pre-test interview, during which the subject denies involvement in a particular criminal act, the subject is administered a PDD examination where a number of specific questions are asked. These questions generally consist of Relevant questions, Control questions, and Irrelevant questions. Relevant questions are questions which specifically refer to the crime under investigation as, "Did you steal any of that money from that safe?" Control questions are similar to the relevant questions in issue and content, but refer to some past criminal act or indiscretion as, "Besides the magazine, six cans of soda, and \$10.00 you told me about, did you ever steal anything else in your life?" (Abrams, 1989). Irrelevant questions are generally innocuous questions, designed to be of no psychological threat to the subject as, "Are the lights on in this room?"

Theoretically, innocent subjects, no matter how threatened the relevant questions may make them feel, know that they are being truthful. However, they know or at least have doubts about the veracity of their answers to the control questions. (Raskin & Kircher, 1991). Therefore, innocent subjects show a greater response to control questions since they feel more threatened by them. Guilty subjects, knowing that they are lying to both

the relevant and control questions feel more threatened by the relevant questions and therefore have greater responses to the relevant questions than to the control questions. Many describe this concept as "psychological set" and define it this way: A person's fear, anxieties, and apprehensions are channeled toward the situation which holds the greatest immediate threat to his self-preservation or general well-being (Backster, 1974).

The CQT is quite versatile and can be used in nearly every criminal specific issue type examination. However, the CQT has been the target of a great deal of criticism from psychologists and psychophysiol-ogists, who question the validity of the CQT methodology (Lykken, 1981, 1985; Furedy & Heslegrave, 1991). Some claim that the "control questions" do not serve as controls in the "standard scientific meaning of this term" (Ben-Shakhar, 1991).

In addition, it is argued that the originators of the control question test made insupportable assumptions regarding the psychological processes of an individual taking a PDD exam (Furedy & Heslegrave, 1991), and that the CQT "Does not yield a test for which one can make a single stable estimate of accuracy" (Lykken, 1985).

Many psychologists encourage the use of an alternate procedure, widely known as the "Guilty Knowledge Test" (GKT) (Lykken, 1959; Furedy, 1991). A major strength of the GKT is that it is not used to determine if the subject is being "deceptive" or "non-deceptive", but instead uses the subject's physiological responses to neutral and key stimuli to reveal "guilty knowledge" concealed by the subject (Lykken, 1959; Iacono, 1988).

In a forensic application, "guilty knowledge" stems from information known only to the person(s) who perpetrated, participated in, or witnessed the crime, or who were informed of the details of the crime by someone who had access to them.

² Editor's Note: In 1998 the APA officially changed "control question" to "comparison question" in all of its publications. The original language of the report has been retained in this reprinting. Readers may also observe other concepts and terminology in this paper that have been superseded since its release in 1992.

During a GKT examination, the PDD examiner questions the subject about a particular case fact or item of evidence which only the guilty party should know, e.g., the caliber of the pistol used to shoot a murder victim. If the subject claims ignorance, then a PDD test is administered in which the caliber of the actual murder weapon, e.g. a .41 caliber pistol (hereafter referred to as the "key"), is presented with a group of other choices. For example: "Regarding the caliber of the weapon used to shoot that woman, was it a: .38 cal, .25 cal, 44 cal, .41 cal (key), .22 cal, .32 cal?"

During subsequent tests, the subject is presented with other similar facts like the location where the shooting occurred, the location of the weapon, the description of an item stolen by the perpetrator, and other pertinent items. If the subject's changes in physiological responses consistently occur to a minimum number of different "keys" during a PDD examination, then the probability is high that he has "guilty knowledge" regarding the case facts. The greater the number of keys, the lower the probability that an innocent person would consistently respond (Raskin & Kircher, 1991). In fact, the actual probability can be mathematically calculated, which is impossible to do with most other PDD procedures (Timm, 1989).

Taking these facts into consideration, it is easily argued that the non-critical items in a GKT would serve more as true "control" questions, than the control questions used in a CQT (Raskin & Kircher, 1991; Lykken, 1988).

The proponents of the GKT argue that this procedure has strong, construct validity and is a more theoretically sound alternative to other types of PDD testing (Iacono & Patrick, 1988). This appears to make the GKT a more understandable and defensible forensic PDD procedure. In fact, both Furedy and Heslegrave argue that "The GKT constitutes the most significant and encouraging prospect for long term improvement in the identification of the guilty by means of physiological measures." (Furedy & Heslegrave, 1991). The GKT as a forensic investigative procedure, has limitations. First of all, for a successful GKT, the examiner

must have at his disposal a number of (at least three) meaningful "key" questions (Forman & McCauley, 1986). These key questions must contain information not previously compromised to the subject by the news media, investigators, or the examiner himself (Raskin, 1989). Next, most GKT studies to date have been done in the laboratory using instrumentation frequently dissimilar to field polygraph instruments. Research personnel conducting the laboratory PDD examinations were generally not professional PDD examiners (Lykken, 1974). Lastly, much of the GKT chart analysis to date has been through the use of "global evaluation," although some limited numerical scoring has been applied (Lykken, 1959, 1988). Global evaluation is highly subjective and is less desirable than a standardized numerical scoring system (Honts & Driscoll, 1988). A semi-objective numerical scoring approach can be inter-rater evaluated, and thus provides the option of a "second opinion" of the examiner's decision. Research has revealed that the use of semi-objective numerical scoring systems have proven to be more reliable than any intuitive evaluation of PDD charts (Honts & Driscoll, 1987) and have yielded consistently higher accuracy (Raskin, Barland, & Podlesny, 1977).

The purpose of this research was to examine the accuracy of the GKT methodology as an investigative PDD procedure for the detection of deception; and, to develop a numerical scoring system for evaluation of the GKT data.

A preliminary numerical scoring system was designed. This system, essentially a modified Rank Order Scoring System (ROSS) (Honts & Driscoll, 1987), was an expanded version of Lykken's numerical evaluation procedure for GSR responses (Lykken, 1959). It was employed by original examiners to test inter-rater numerical scoring and was later modified.

Method

Subjects

Eighty-nine healthy subjects, (males and females between the ages of 19 and 33, with a median of 21) were obtained from the population of troop trainees at Fort McClellan,

Alabama. The subjects had no prior PDD experience.

Examiners

Four U.S. Marine Corps Criminal Investigation Division (CID) certified forensic psychophysiolgists conducted the examinations. All were graduates of the Department of Defense Polygraph Institute (DoDPI)³ and were trained and experienced in standard PDD testing procedures. The examiners were male, averaging 2.75 years of experience in forensic PDD techniques. Upon arrival at DoDPI, the examiners spent their first three days learning the CODAS computerized recording system, and two days practicing the particular GKT methodology utilized in this study.

Apparatus

Lafayette "Factfinder" model field polygraph instruments were used in this study. Recordings consisted of two electronic pneumograph channels to measure changes in thoracic and abdominal expiration and inspiration, and a skin conductance channel with the finger plates being placed on the hand opposite the arm used for a conventional pressure cuff cardiovascular activity channel.

These polygraph instruments were connected to a personal computer which digitized the analog data using CODAS Level Five software by Dataq.

Procedure

Upon arrival at the Institute, subjects who had been randomly selected for this project by their respective training commands, were met and briefed on the purpose of this investigation (Appendix A).⁴ The subjects were asked if they had prior polygraph experience. If they did, they were returned to their unit and not included in this study. The purpose and procedures of the study were fully explained to all subjects (Appendix B), and they were asked to read and sign a volunteer affidavit (Appendix C), which informed them

that their participation was voluntary. Basic background information regarding subject medical and physical condition (Appendix D) was completed and they then waited in a room until called for individual programming.

Scenario

A mock "homicide" scene was arranged in a near-by room. The "victim" was a female mannequin, placed in a bed prior to the arrival of the programmed guilty subjects. In addition to the bed, the room contained a dresser, night table, chair, a dresser mirror, and personal items such as jewelry, make-up, handbags, and purses, consistent with what might be found in a woman's bedroom. There were no windows in the room.

The scenario setter, acting as a "director" took each programmed guilty subject into the crime scene room. The scenario setter, working from a script (Appendix E), instructed each subject to pretend that they had entered the room to commit a theft. As they looked around the room, they were told that the victim was waking up. They were then ordered to take a broom from a corner of the room and to strike the mannequin over the head with the broom. The scenario setter then "checked the pulse" of the "victim" and informed the subjects that their actions "killed" the woman. The subjects were then told to conceal the broom handle under the bed. The subjects were instructed to remove a ring from the woman's finger and conceal the ring in their pocket. The subjects then wrapped the woman in a bright orange sheet (provided), and hid the woman behind the bed. They then took a woman's lipstick from a dresser top and wrote "Satan" on the mirror. The subjects then exited the room with the scenario setter.

After the programming, the guilty subjects were given a questionnaire regarding the crime scenario (Appendix F). The questionnaire consisted of six questions concerning crime scene facts and evidence,

³ Now the National Center for Credibility Assessment (NCCA).

⁴ None of the appendices are included in this reprinting. Interested readers can obtain the complete report at www.dtic.mil.

and was used to determine what facts the subjects remembered from the crime scene. If any subject failed to remember certain key fact(s), they were reminded of that information prior to their PDD examination.

The programmed innocent subjects were provided no information regarding the scenario and were not allowed to interact with any of the programmed guilty subjects. The innocent subjects were informed that they would be given a PDD examination regarding a homicide investigation, but since they were not involved in any way, they had no information or knowledge of the details of the crime (Appendix G). The innocent subjects were then taken to a room separate from the programmed guilty subjects to avoid contamination.

All subjects were informed that they would soon be administered a polygraph examination regarding a homicide investigation. Regardless of their programming, they were instructed to cooperate fully with the examiner and if asked by an examiner about the case, to say that they had been informed by DoDPI staff that a homicide had been committed, and that they know nothing more about it.

Piloting

Before data collection for the actual study, two pilot studies were conducted. The first pilot (Pilot 1), involved the participation of 22 guilty subjects in the mock crime scenario referred to previously. After participating in the scenario, the subjects were asked to write a statement describing their actions and observations during the scenario. The subjects answered a multiple choice questionnaire to determine which crime scene facts they remembered. The crime scene facts most often reported were:

1. The weapon used to murder the victim (22).
2. The location the victim's body was concealed after the murder (22).
3. The item stolen from the room (22).
4. The color of the sheet used to wrap the victim (21).
5. The message left by the perpetrator (21).
6. The location in which the murder weapon was concealed (20).
7. The item on which the message was written (20).
8. The item used to write the message (19).

These eight subject areas were then utilized for the second pilot (Pilot 2) study. During Pilot 2, 20 programmed innocent subjects were administered PDD examinations to determine if they responded to key question choices more frequently than chance. During these examinations, eight PDD tests were collected from each subject. In keeping with the design of this study, each test contained one key item and five alternate choices. An evaluation of the collected data indicated none of the subjects responded to the key items more frequently than chance. The question sequences selected for use in the core study are located in Appendix F. The locations of the key questions in the question sequence were randomly assigned from a table of random numbers. Additionally, the order in which the question sequences were asked, was rotated by the personnel responsible for assignment of subjects to examiners.

The GKT Examination

Eighty-nine subjects were administered the GKT examinations. Nine were removed from the data set for reasons detailed in the discussion section. The remaining eighty subjects were equally distributed between innocent and guilty conditions.

The GKT construction rules were as follows:

1. Each of the questions selected (key questions) were combined with five other like items for a total of six questions per sequence.
2. The test questions were reviewed immediately before each test with the subject, but not in the order of their appearance. The test questions were not given in any logical (i.e., numerical) sequence.
3. Each sequence was asked once for a total of six questions per test.

4. If the subject moved or caused some other distortion during the administration of a test, the examiner utilized an alternate test prepared with a different key and alternate choices. If the subject continued engaging in that behavior, the examination was terminated.

5. The key question was not asked in the first position on any test.

6. The location of all critical items were randomly assigned utilizing random assignment tables.

7. A stimulation test was used which required the subject to choose a number in a given range. When asked about the number, the subject was instructed to deny choosing that particular number. No feedback was given to the subject concerning his responses.

All subjects were taken to laboratories at DoDPI, asked to sign a rights waiver and polygraph consent form (Appendixes H and I), and were then given a series of instructions regarding the PDD examination (Appendix J). They were then questioned regarding their knowledge of the crime scene. The programmed innocent subjects had no knowledge and the programmed guilty subjects had been instructed to deny any knowledge of the crime.

The sensors were then attached. Each subject was instructed to sit quietly and to answer "no" to all the test questions, while the physiological recordings were collected.

A stimulation test was collected first. It was not used in the conventional manner, but used to make sure that the subjects could follow directions and were physically suitable for testing. Five GKT PDD tests were collected on each subject, with the key questions and alternate questions being asked once during each test.

After the examination, the subjects were asked to fill out the "subject questionnaire" (Appendix F). If programmed innocent subjects properly identified three or more of the key items on the questionnaire, they were interviewed by the principal investigator to determine the source of their

apparent knowledge. Three subjects exceeded this threshold. In all three cases, the subjects denied any prior knowledge of the scenario and stated that the correct answers simply "made sense." All subjects were debriefed before their release (Appendix K).

Test Evaluation

The initial test evaluation strategy called for the original examiners to utilize a modified version of Lykken's scoring system for Galvanic Skin Response (GSR) data (Lykken, 1959). In this approach, a numerical value of "2" was credited to each physiological recording if the physiological response was greater to the critical item than the other choices. If the response to the critical item was the second greatest, it was credited a value of "1." If the responses were less than the above, they received a value of "0." Therefore, the maximum value the items could earn was "6" (2 points for each physiological recording) and the minimum would be "0." All physiological channels were weighted equally. Responses to the first item in the tests were not evaluated. A numerical score of "3" was arbitrarily selected as the cut-off for a decision of "guilty knowledge." A score of "3" or greater was needed on at least three tests for a decision of "guilty knowledge" to be rendered. No inconclusive decisions were allowed.

Since the GKT is designed to determine if a person is concealing information, it was decided that the diagnostic categories would be "Concealment Indicated" (CI), for scores of "3" or above and No-Concealment indicated (NCI) for scores below "3".

Results

Upon completion of the data collection phase of this project, half of the tests from the completed examinations, 20 innocent and 20 guilty, were put aside (designated Group 2) to be used as a validation group for any numerical scoring system designed during this study.

The remaining 40 examinations were designated Group 1 and evaluated utilizing the scoring approach described in the test evaluation section. The original examiners (OE), were correct in 13 of 20 cases involving programmed guilty people, and correct in 18

of 20 programmed innocent subjects for an accuracy rate of 65% for guilty, and 90% for innocent, and an overall accuracy rate of 77.5%.

A blind evaluation by another examiner, utilizing the same scoring system as the original examiners, was correct in 10 of 20 cases (50%) involving the programmed guilty; in 18 out of 20 (90%) programmed innocent and an overall accuracy rate of 70%.

These examinations were used to test various evaluation systems in an effort to optimize scoring accuracy of the GKT methodology. A total of nine individual scoring approaches were applied as follows:

1. "OE"- Original Examiners scoring as set forth in the methods section.
2. "S1"- System 1 was blind evaluation scoring as set forth in the methods section.
3. "S2"- System 2 was blind evaluation scoring a "2" or higher on three or more tests was considered concealment indicated (CI). If a score of "2" or higher occurred on 2 of the five tests, the test was deemed inconclusive.
4. "S3"- System 3 was blind evaluation scoring the most significant skin conductance response as a "3", second most significant as a "2," the third most significant as a "1"; other recordings unchanged. A "2" or better on three or more tests for a decision of CI. No inconclusives were allowed.
5. "S4"- System 4 was the same scoring strategy as System 3, but with a minimum cut-off for CI decision being a "3" or better on 3 or more tests. A test was determined to be

inconclusive only if there were no skin conductance responses.

6. "S5"- System 5 was the same scoring strategy as System 4, but the respiration recordings receive a value of 1 if the response is greatest to the key question and lesser responses receive a "0."

7. "S6"- System 6 was the same scoring strategy as System 5, but the greatest response in the cardiovascular recordings receive a maximum value of "3," second greatest a "2," and the third is scored as a "1."

8. "S7"- System 7 was conventional GKT scoring as specified by Lykken (1959). Maximum points obtainable on the total test was 30 (2 points per recording = 6 points per test X 5 tests). If within 0-15 total test points, subject was diagnosed as NCI. If within 16-30 total test points, subject was diagnosed as CI.

9. "S8"- System 8 was the same scoring strategy as System 7, with the exception that the greatest skin conductance response was scored as a "3," the second as a "2," the third as a "1." The other recordings are unchanged. The maximum points per examination was 35 (due to the added point given the maximum skin conductance response X 5 tests). If within 0-15 total points, subject was diagnosed as NCI. If within 16-35 total points, subject was diagnosed as CI.

The results from each scoring system is depicted in Tables 1 through 3.

Table 1 depicts evaluator decisions for group one innocent subjects by the scoring system.

Table 1. Examiner Decision for Group 1 Innocent Subjects

Subject		Scoring System							
#	OE	S1	S2	S3	S4	S5	S6	S7	S8
3	G	G	G	G	G	G	G	G	G
6	I	I	INC	I	I	I	I	I	I
9	I	I	I	I	I	I	I	I	I
15	I	I	INC	G	I	I	I	I	I
17	I	I	G	G	I	I	I	I	I
19	I	I	INC	G	I	I	G	I	I
21	I	I	I	I	I	I	G	I	I
25	I	I	INC	G	I	I	I	I	I
27	I	I	INC	G	I	I	G	I	I
30	I	I	I	I	I	I	I	I	I
36	I	I	I	I	I	I	I	I	I
39	I	I	INC	I	INC	INC	INC	I	I
48	I	G	G	G	G	I	G	I	I
52	I	I	INC	I	I	I	I	I	I
56	I	I	I	I	I	I	I	I	I
63	I	I	G	G	I	I	I	I	G
65	I	I	G	G	I	I	I	I	I
71	I	I	INC	I	I	I	I	I	I
79	G	I	G	G	G	G	G	I	G
83	I	I	INC	I	I	I	I	I	G

(G) Guilty Decision

(I)= Innocent Decision

(INC)= Inconclusive Decision

Table two depicts evaluator decisions for group one guilty subjects by scoring system.

Table 2. Examiner Decisions for Group 1 Guilty Subjects

Subject		Scoring System							
#	OE	S1	S2	S3	S4	S5	S6	S7	S8
2	G	G	G	G	G	G	G	I	G
8	G	G	G	G	G	G	G	I	G
12	I	I	I	I	I	I	I	I	I
26	G	G	G	G	G	G	G	G	G
29	G	G	G	G	G	G	G	G	G
33	G	G	G	G	G	G	G	G	G
38	G	G	G	G	G	G	G	G	G
43	I	I	G	G	G	G	G	I	I
44	G	I	G	G	G	G	G	I	G
46	I	I	G	G	G	G	G	I	I
49	G	G	G	G	G	G	G	I	G
55	I	I	INC	I	INC	INC	INC	I	I
59	G	G	G	G	G	G	G	I	G
62	I	I	G	G	G	I	I	I	I
67	I	I	I	G	I	I	I	I	I
69	G	I	G	G	G	G	G	I	G
76	G	I	G	G	G	I	I	G	G
78	G	G	G	G	G	G	G	G	G
85	I	I	INC	G	INC	INC	INC	I	I
88	G	G	G	G	G	G	G	G	G

(G) Guilty Decision

(I) Innocent Decision

(INC) = Inconclusive Decision

Table 3. Accuracy per Scoring System (N=40)

Decisions:	Guilty (N=20)			Innocent (N=20)			Inc	Accuracy
	CI	NCI	Inc	CI	NCI	Inc		
OR	13	7	0	2	18	0	--	78%
SI	10	10	0	2	18	0	28%	72%
S2	16	2	2	6	5	9	--	70%
S3	18	2	0	10	10	0	8%	86%
S4	16	2	2	3	16	1	8%	84%
S5	14	4	2	2	17	1	8%	73%
S6	14	4	2	6	13	1	--	65%
S7	7	13	0	1	19	0	--	72%
S8	13	7	0	4	16	0	--	70%

(CI) Concealment Indicated

(NCI) No Concealment Indicated

(INC) = Inconclusive

Table 3 depicts the results of evaluator's decisions of the Group one tests and overall scoring system accuracy.

After analyzing the eight scoring strategies, the "S4" system had the best overall accuracy. The forty examinations originally set aside (Group 2) were then evaluated utilizing the S4 scoring system. After recording the S4 scores, the scores of the original examiners (OE) were recorded. The scores of a blind evaluation using the original examiner's (SI) scoring system were also recorded. The results, as seen in Tables 4 through 5, show that 15 of the 20 guilty

subjects and 17 of the 20 innocent subjects were correctly identified, with one exam being declared an inconclusive test.

The S4 scoring approach, now designated the "232" scoring system (for the value given its channel weighing), netted an accuracy rate of 75% on guilty and 89.5% on innocent, with an overall accuracy rate of 82% for the GKT procedure, utilizing the "232" system.

Table four contains examiner decisions concerning Group 2 subjects, based on scoring systems.

Table 4. Group 2 Examiner Decisions for Three Scoring Systems

Innocent Subjects	Scoring System			Guilty Subjects	Scoring System		
#	OE	SI	S4	#	OE	SI	S4
Decisions							
4	I	I	INC	1	I	I	G
5	I	I	I	7	G	I	G
10	I	I	I	11	I	I	I
18	I	I	I	13	I	I	G
20	I	I	I	28	I	G	G
22	I	I	I	31	I	I	I
23	I	I	I	35	G	G	G
24	I	I	I	41	I	G	G
34	I	I	I	45	G	I	G
37	I	I	I	47	I	I	I
42	I	I	I	54	I	I	G
51	I	I	I	57	I	G	G
53	I	I	I	60	G	G	G
61	G	G	G	66	I	I	I
64	G	I	I	68	I	I	G
70	I	I	I	73	I	I	G
74	I	I	I	77	I	G	G
81	I	I	G	80	I	G	G
82	I	I	I	87	G	G	G
84	I	I	I	89	I	I	I

(G) = Guilty Decision

(I) = Innocent Decision

(INC) = Inconclusive Decision

Table five presents the combined results of evaluator's decisions of the 40

Group 2 examinations based on the scoring system utilized.

Table 5. Accuracy per scoring system.

Programmed: Decisions:	Guilty (N=20)			Innocent (N=20)		
	CI	NCI	INC	CI	NCI	INC
OR	5	15	0	2	18	0
SI	8	12	0	1	19	0
S4	15	5	2	2	17	0

	Correct	Incorrect	Inconclusive	% INC	% Accuracy
OR	23	17	0	--	58
SI	27	13	0	--	68
S4	32	7	1	2	82

CI = concealment indicated

NCI = No concealment indicated

INC = inconclusive.

Binomial and Chi² (Goodness of Fit) tests were conducted on the OE, S1, and S4 scoring of the Group 2 PDD tests. As can be seen on Table 6, the S4 (232) scoring system results were significant.

Table 6 presents the results of a binomial test of each group two condition (Guilty & Innocent) and the Chi² analysis of the evaluator's decisions based on analysis of the test results from both conditions.

Table 6. Statistics

Binomial Test		Chi ²	
(Guilty)	(Innocent)	(Both Groups)	
OE/ p=.21	OE/ p<.001	OE/ X ² =0.9	
S1/ p=.252	S1/ p<.001	S1/ X ² =4.9	(p<.05)
S4/ p=.021	S4/ p<.001	S4/ X ² =16.0	(p<.001)

Using the 232 scoring system on all 80 examinations revealed that 31 out of 40 guilty subjects and 33 out of 40 innocent subjects were properly identified. Two of the guilty and two of the innocent subject examinations were declared inconclusive. Excluding the inconclusive, the net accuracy rate was 82% for the forty programmed guilty subjects and 87% for the forty programmed innocent

subjects for an overall accuracy rate of 84% across the 80 subjects.

Each channel was evaluated individually to ascertain if any one physiological measurement was any more accurate than the others in identifying either the guilty or innocent. For this analysis, all 80 examinations were evaluated and the 81

numerical scores from each physiological channel were added horizontally across each test. The 81 system was selected for this purpose, since this was the system specified for scoring prior to the initiation of the study. Since there were five tests per exam, the greatest score any one channel could receive would be a "10," (2 points for each test). If the over-all score for a particular channel was 0 to 5, the exam was scored innocent. If the score was 6 to 10, the exam was scored as guilty. No inconclusive zone was utilized.

As seen in Table 7, the skin conductance was the only channel that discriminated between guilty and innocent better than chance.

Table 7 presents the accuracy of each physiological recording as evaluated independently of the others. The results are depicted in raw numbers of total examinations over both groups (one & two) and in terms of programmed condition (Guilty & Innocent). The results of a binomial test are also depicted.

Table 7. Accuracy for Each Physiological Channel

Programmed <u>Component</u>	Guilty (N=40)		Innocent (N=40)		Both Groups (N=80)	
	<u>CI</u>	<u>NCI</u>	<u>CI</u>	<u>NCI</u>	<u>Correct</u>	<u>Incorrect</u>
Pneumo	9	31	4	36	45	35
SCR	10	30	0	40	50	30
Cardio	5	35	2	38	43	37

-Binomial Test -
(Two-tailed)

Pneumo	Z=1.006	p = n.s.
SCR	Z=2.12	p = .034
Cardio	Z=1.949	p = n.s.

(CI) = concealment indicated

(NCI) = no concealment indicated

Discussion

The results of this study indicate that the Guilty Knowledge Test does discriminate more towards clearing the innocent than identifying the guilty. This result is consistent with earlier studies (Lykken, 1981; Forman & McCauley 1986).

It should be noted that the subjects examined in this study were assigned to their respective conditions of guilt or innocence by research personnel. According to Forman & McCauley, the majority of analog studies are designed in this manner (e.g., Forman & McCauley 1987; Barland & Raskin, 1975; Lykken 1959; Honts & Hodes, 1982). In

assigning subjects to commit (or to not commit) a "crime," the question of external validity naturally becomes an issue.

By assigning a subject to condition, and then instructing them to lie or to withhold information, the end result is that the subject has little to nothing to lose if a lie (or concealment) is detected. In a real situation, the subject risks possible imprisonment, financial loss, and personal embarrassment if a deception is discovered. Even in studies such as this one, that attempt to simulate actual testing conditions through the use of mock crimes, the emotions associated with deception are simply not the same and not of the same degree (Abrams, 1972). Indeed,

many PDD examiners and researchers believe that greater motivation to deceive leads to greater detectability (Abrams, 1972; OTA, 1983; Forman & McCauley, 1986).

The subjects utilized in this study not only lacked a true criminal's fear of getting caught, but presented the additional problem of extreme physical fatigue stemming from the strenuous regimen associated in a basic training situation. Several of the subjects started to fall asleep during the operational phase of the test and some had to be stood up and walked around in between the collection of individual tests. As reported in the methods chapter, a total of nine examinations had to be removed from the data set. Of these nine examinations, eight were terminated due to lack of subject cooperation, based at least in part on the inability of the subjects to remain awake and alert during the test. Of these eight, four had been programmed innocent and four guilty. Additionally, an over-all lack of physiological arousal appeared to be present on many of the PDD tests. Although there may be a number of reasons for this lack of arousal, it is suspected to be connected with the absence of threat, and the physical exhaustion previously discussed.

The final subject dropped from the data set was an extra subject discovered in the innocent population during data analysis. Since the research protocol called for equal numbers of innocent and guilty, (40 innocent & 40 guilty), the data from the very last innocent subject tested (#86) was removed from the set. For informational purposes, examination #86 was subsequently evaluated and was determined to be a true negative by all scoring approaches.

As with any controlled experiment, a number of the procedures utilized during this study would not be used in a live field situation. For example, during this study, there was only a short delay (usually less than 15 minutes) between the commission of the crime and the administration of the PDD test. In a field situation, months or years could indeed pass between these two events. In this study, if a subject had a prior polygraph examination, he was not tested. In the field, many criminal suspects have had PDD examinations before. The practice of having

the guilty subject fill a questionnaire prior to the PDD examination would have no applicability in the field, nor would the deliberate coaching of the subject through the crime scene. It could be argued that these control measures enhance the detectability of the subject. Certainly, this may be true. However, it is as easily argued that the actual criminal in a field setting is coming into the PDD examination with far more emotional involvement in the outcome of the test. Therefore, he is certainly more likely to be psychologically and physiologically aroused than his laboratory equivalent.

Countermeasures

Countermeasures are deliberate attempts by a guilty subject to alter his physiological reactions, recorded for analysis on a polygram, to appear non-deceptive (Stephenson & Barry, 1988).

Because a minimum amount of physiological arousal must be present on the key question for a GKT to be evaluated as "Concealment Indicated", then the issue of physical, psychological, and pharmaceutical countermeasures must be considered. Theoretically, a guilty subject could reduce his arousal level by use of a number of "relaxation techniques" (Honts & Hodes, 1982), or by ingesting a substance that so altered his normal physiological responses (e.g., depressants, stimulants, hallucinogens, etc.), that any reactions to the key question would be insufficient to score. Physical countermeasures could come into play in the situation where a guilty subject would deliberately distort the physiological recordings on the key and all alternate choices, so as to make responses indistinguishable from each other. Research has shown that various forms of deliberate physical distortions (e.g., tongue biting, pressing toes against the floor, and other muscle movements) can be effective in diminishing the ability of a PDD test to identify the guilty subject, if the subject has been extensively trained in these techniques (Abrams & Davidson, 1988; Honts & Raskin, Kircher & Hodes, 1988; Stephenson & Barry, 1988).

Proponents of the CQT argue for that reason, the CQT is a superior test to the GKT.

They maintain that the only end result of many countermeasures, would be that the test would be declared "inconclusive" since these countermeasures would affect the relevant and control areas equally (Raskin, 1986). Lykken (1981) disagrees, arguing that an individual properly trained in countermeasures would artificially enhance only his reactions to the control questions, thereby in fact passing the CQT test. Regardless of the position taken, it should be noted that most countermeasure studies involved individuals formally trained and that such training is not readily available to most individuals who are tested in the criminal justice system (Honts & Raskin, Kircher & Hodes, 1988).

Whether or not caused by countermeasures, the lack of physiological arousal could add to the false negative error rate of a GKT examination. In an effort to address this lack of responsivity, several of the test evaluation strategies examined during data analysis contained an inconclusive area which was not driven by numerical cut-offs (84, 85, 86). Instead an examination required an inconclusive decision if the PDD tests showed a significant lack of physiological responsiveness, specifically in the skin conductance channel. This decision was ultimately made on four of the eighty total examinations resulting in a 5% over-all inconclusive rate. It is interesting to note that the condition of the subjects ultimately declared inconclusive during this study, were equally divided at two innocent and two guilty.

It may be prudent to suggest that if sufficient responsivity is present in the other channels (pneumographs and cardiograph), to cause a numerical decision of concealment indicated (CI), then that decision should be made, regardless of any lack of responsivity in the skin conductance channel.

This approach was not considered during analysis of the physiological recordings collected during this study. However, inspection of the four inconclusive PDD examinations revealed a lack of arousal in all three parameters. Therefore, on these particular examinations, such a rule would not have made a difference.

Field Application of the GKT

Even the strongest proponents of the GKT acknowledge difficulties in the versatility of this procedure. The use of the GKT requires a special set of conditions, where only the guilty have knowledge of the crime (Furedy & Heslegrave, 1991).

For example, generally the GKT cannot be used in cases where the subject admits being present at the crime scene, but denies committing the act. Unless the GKT examiner has access to sufficient GKT information to construct test questions able to identify criminal knowledge beyond what the subject would have obtained from simply being at the crime scene. Guilty Knowledge Tests cannot be used in sexual assault cases where consent is the issue (Raskin, 1989). Next, as with any test that depends on crime scene events or evidence, the question of what facts the subject actually perceived or remembered is raised (e.g., The subject may not have observed the color of the victim's clothing).

In many cases, the key facts which would be used by a PDD examiner to construct a GKT have been compromised by the media, police, or even the examiner himself. Additionally, the very circumstances under which most forensic PDD examiners test, is not conducive to the administration of the GKT procedure. For example, in the U.S. Department of Defense, it has been directed by most criminal investigative commands that the forensic PDD examination be administered as the last or near the last step of a criminal investigation (e.g., AR 195-6; AFOSIR 124-40). These agencies direct that one or more thorough interrogations of the subject be completed before the PDD examination is conducted. During the course of these interrogations, the investigators often divulge the very information that must be kept secret from the suspect in order to construct the GKT (Raskin, 1989). It is no surprise then that many PDD examiners believe that they cannot frequently utilize the GKT and automatically select to forego concealed information type tests in favor of what they consider to be more versatile formats.

With some small modification of current criminal investigative procedures, the GKT could become a far more frequently

administered test. Instead of waiting until the very end of a criminal investigation to consider the forensic use of PDD exams, they should be considered an available tool from the outset of the case.

With a closer working relationship, investigators and examiners could follow the logical progression of an investigation. If possible, examiners could visit the more significant crime scenes to get a first-hand view of possible GKT key material. Where not possible, crime scene photographs and video tapes, together with case facts and statements, could help provide this information to the PDD examiners. As the time approaches for suspects to be interrogated, liaison between the investigator and examiner could lead to an interview strategy designed not to reveal certain case facts needed for a GKT examination to be administered. In cases where a significant number of suspects would make other forms of PDD testing impractical, the GKT could be used as a screening process to identify those who have guilty knowledge about the crime scene facts or evidence. Because the GKT has a high false negative rate, clearing the test should not result in complete removal of suspicion. However, those identified by the GKT would become excellent subjects for further investigation and interrogation.

Summary and Conclusion

The results of this study suggest that the Guilty Knowledge Test with the "232" scoring system would be a viable addition to the testing procedures at the forensic psychophysiologicalist's disposal.

Furthermore, the development of the "232" scoring system appears to provide an easily applied and objective evaluation of GKT test results by field PDD examiners. In cases where there is a quality control system in place, the "232" scoring system allows for an objective second opinion of the field examiner's decision.

The Guilty Knowledge Test represents one of the very few forensic psychophysiological detection of deception procedures which would be easily understood and accepted by most criminal justice professionals. The GKT is logical and theoretically sound. Unlike other methodologies, the GKT does not require any psychological manipulation of the subject and its success is not as dependent on examiner/subject interaction. The GKT is not difficult to present to other criminal justice professionals nor is it hard to explain to the layman. This fact is crucial if PDD procedures are ever to be widely accepted by the court system in this country.

The "art" of polygraphy as practiced by many, must be augmented by a more scientific group of forensic psychophysiological tests. Although, the GKT is by no means the last step in this evolution, it can be considered a step in the proper direction.

For the GKT to be valued as a forensic tool, field studies need to be conducted into this procedure. Regardless of the results of any one analog study, there is a notable lack of field data concerning the GKT (Raskin, 1989). One way in which to field test the viability of the GKT methodology, would be to dedicate one PDD examiner from a large metropolitan police department, such as New York City, or Los Angeles, to travel to major crime scenes with the forensic laboratory personnel. Once on the crime scene, the PDD examiner could examine the scene first hand for GKT material. The examiner would maintain liaison with the detectives working the case, and then use all the key material acquired to conducted GKT examinations of any identified suspects. As with other field studies, the examiner's findings could be validated by confession, conviction, or by other scientific means. Whatever methodology is chosen, the field data must be collected.

References

- Abrams, S. (1972). The polygraph: Laboratory v. field research. *Polygraph* (1), 145-150.
- Abrams, S., & Davidson, M. (1988). Counter-countermeasures in polygraph testing, *Polygraph*, 17(1), 16-19.
- Abrams, S. (1989). *The Complete Polygraph Handbook*. Massachusetts: Lexington Books.
- Barland, G. H., & Raskin, D. C. (1975). An evaluation of field techniques in the detection of deception. *Psychophysiology* (12), 321-330.
- Backster, C. (1974). The anticlimax dampening concept. *Polygraph* (3), 48-50.
- Ben-Shakhar, G. (1991). Future prospects of psychophysiological detection: Replacing the CQT by the GKT. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds), *Advances in Psychophysiology*, (4, 194). London: Jessica Kingsley.
- Forman, R. F., & McCauley, C. (1986). Validity of the positive control test using the field practice model. *Journal of Applied Psychology*, 62, 127-136.
- Frye v. United States*, 293 F. 1013 (D.C. Cir 1923).
- Furedy, J. J., & Heslegrave, R. J. (1991). The forensic use of the polygraph: A psychophysiological analysis of current trends and future prospects. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds), *Advances in Psychophysiology*, (4, 157-186). London: Jessica Kingsley.
- Honts, C. R., & Driscoll, L. N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16(4), 241-245.
- Honts, C. R., & Driscoll, L. N. (1988). A field validity study of the rank order scoring system (ROSS) in multiple issue control question tests. *Polygraph*, 17(1), 1-13.
- Honts, C. R., & Hodes, R. L. (1982). The effects of multiple physical countermeasures on the detection of deception. *Psychophysiology*, 19, 564. (Abstract)
- Honts, C. R., Raskin, D. C., Kircher, J. C., & Hodes, R. L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. *Journal of Police Science and Administration*, 2, 91-94.
- Iacono, W. G., & Patrick, C. J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed), *Clinical assessment of malingering and deception* (206-227). New York: Guilford Press.
- Larson, J. A. (1932). *Lying and its detection*. Chicago: University of Chicago Press.
- Lykken, D. T. (1959). The GSR in the detection of guilty. *Journal of Applied Psychology*, 43, 385-388.
- Lykken, D. T. (1974, October). Psychology and the lie detector industry. *American Psychologist*, p. 728.
- Lykken, D. T. (1985). The probity of polygraph. S. M. Kassin & L. S. Wrightsman (Eds), *The Psychology of Evidence and Trial Procedure* (pp 95-98). Beverly Hills, CA: Sage Publications.

- Lykken, D. T. (1988). Detection of guilty knowledge: A comment on Forman and McCauley. *Journal of Applied Psychology*, 73(2), 303-304.
- Office of Technology Assessment (1983). Scientific validity of polygraph testing: A research review and evaluation--a technical memorandum (OTA-TM-H-15). Washington DC: U.S. Congress, Office of Technology Assessment.
- Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding application and acceptance of polygraph evidence. *Utah Law Review*, 1, 50. University of Utah College of Law.
- Raskin, D. C. (1989). Polygraph techniques for the detection of deception. *Psychological methods in criminal investigation and evidence* (p 281). New York: Springer.
- Raskin, D. C., & Kircher, J. C. (1991). Comments on Furedy and Hesslegrave: Misconceptions, misdescription, and misdirections. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds), *Advances in Psychophysiology* (pp 216-219). London: Jessica Kingsley, 216-219.
- Raskin, D. C., Barland, G. H., & Podlesny, J. A. (1977). Validity and reliability of detection of deception. *Polygraph*, 6, 1-39.
- Stephenson, M., & Barry, G., (1988). Use of a motion chair in the detection of physical countermeasures. *Polygraph*, 17-21.
- Timm, H. W. (1989). Methodological considerations affecting the utility of incorporating innocent subjects into the design of guilty knowledge polygraph experiments. *Polygraph*, 18(3), 145.
- US Army Regulation 195-6 (1989). Criminal investigation, DOA polygraph activities (2.0). Headquarters, Department of the Army, Washington, DC.
- US Air Force Office of Special Investigations Regulation 124-40 (1988). AFOSI Polygraph Program (p 19). Department of the Air Force, Headquarters, Air Force Office of Special Investigations, Bolling Air Force Base, Washington, DC.
- United States v. Piccinonna*, 885 F. 2d 1529 § 12 (11th Cir 1989).