



EUROPEAN

# POLYGRAPH

**PUBLISHED QUARTERLY**

Volume 2

2008

Number 2 (4)



Andrzej Frycz Modrzewski Krakow University College





## Contents

- Yury Kholodny, *Interrogations using polygraph in Russia: 15 years of legal application* ..... 73
- Mark Handler, Raymond Nelson, *Test Structure and Administration* ..... 83
- Frank Horvath, J. P. Blair, Joseph P. Buckley, *The Behavioral Analysis Interview: Clarifying the Practice, Theory and Understanding of its Use and Effectiveness* ..... 113
- Vitas Saldžiunas, Aleksandras Kovalenko, *The Event Knowledge Test (EKT) in Polygraph Examination (in case murder)* ..... 139

---

## Book review

- Frank Horvath, *JE Sullivan: Gatekeeper: memoirs of a CIA polygraph examiner* ..... 147
- Jan Widacki, *Jolanta Antas: O kłamstwie i kłamaniu (On lie and lying)* ..... 153

---

## Report

- Jerzy Konieczny, *AAPP Thirty First Anniversary Polygraph Seminar, Jacksonville, Florida, May 19-23, 2008, Seminar report* ..... 159

- 
- The basic information for Authors ..... 163





Yury Kholodny\*

Moscow State Technical University  
n.a. N.E. Bauman (BMSTU)  
RUSSIA

## Interrogations using a polygraph in Russia: 15 years of legal application

The method of psychophysiological detection of deception (PDD) has a long history in Russia. This history can be divided into six stages.

The first stage – the early history of the PDD method – ran from 1925 to 1930. The pioneer of the development of the PDD method in the Soviet Union was Alexander Luria, who later became a world-famous scientist. He investigated the psychological mechanisms of the PDD applied this method in practice in the Moscow district public prosecutor's office.

Alexander Luria discovered and formulated **the general principle of the PDD method using a polygraph** and all forms of the psychophysiological method of disclosure of hidden information. In the 1920s he wrote: “the only opportunity to research internal, ‘hidden’ psychological processes is to connect these ‘hidden’ processes with any physiological processes working simultaneously, observed from outside ... and these physiological processes should be chosen

---

\* kholodny@yandex.ru

in such a way that psychological processes could affect them” (1). In the early 1930s the second stage of the history of PDD in Russia began. All scientific investigations in this field were brought to a halt in the USSR. Soviet official jurisprudence declared that the PDD method was pseudoscientific and it was practically prohibited. The second stage lasted till 1993.

Though the PDD method was declared pseudoscientific, US achievements in polygraph application attracted the attention of specialists in the USSR, and the KGB leader Yuri Andropov (later the head of the USSR) approved the use of polygraphs. Today it is no secret that the main user of polygraph examinations in the Soviet Union was the KGB.

In the middle of the 1970s the third stage of the history began. This was the time of serious scientific investigations of the PDD method and its practical uses. These scientific studies and other works demonstrated the high effectiveness of the employment of polygraphs in law enforcement practice. The beginning of the employment of polygraphs in the USSR could not be concealed. During hearings in 1979 the US Congress discussed the appearance of polygraph examinations in the Soviet Union.

The second and third stages finished at the same time, and in the early 1990s the history entered its fourth stage – the stage of legal polygraph application. In spring 1993 the Ministry of Justice (MJ) awarded legal status to polygraph examinations in Russia. The Federal Security Service (FSS) of the Russian Federation was the first government department to employ the polygraph legally. That year marked the beginning of the broad spread of polygraphs in the country. The Ministry of Internal Affairs (MIA) and private commercial organizations began to use polygraphs in 1994. In the next six years the Ministry of Defense (MD), MJ and some other departments started to use polygraphs, including screening examinations.

In 1996 IUP were used only in Moscow and four other cities in Russia. Now polygraphs are used in more than a hundred large and small towns, from Kaliningrad (on the Baltic coast) to Petropavlovsk-Kamchatsky (on the Pacific Ocean coast).

During the 1990s in Russia a paradoxical situation surrounded polygraphs. On the one hand, several government departments were employing them in law enforcement practice. Polygraph use increased annually. But on the other hand, Russian criminalistic science (or – in English – forensic sciences) could not give the right scientific explanation of polygraph examinations or – in Russian – of the **interrogations using a polygraph (IUP)**.

That is why, according to Russian legal tradition, IUP could not be employed in law enforcement practice. This contradiction was successfully eliminated at the beginning of the new century.

The fifth stage of the history of the PDD method is the entry of the polygraph into criminalistics. Russian forensic sciences accepted IUP as a criminalistic method, and this step paved the way for a wide application of the polygraph in the everyday life of the state. A new branch of criminalistic science has the name “criminalistic diagnostic investigation using a polygraph”(2) (or “criminalistic polygraphology”) and under this name it entered modern Russian criminalistics textbooks (3).

The sixth stage of the PDD method history is developing now: it is a time of recognition of IUP as forensic psychophysiological expertise.

Before discussing current polygraph use in Russia, it is necessary to make some preliminary remarks.

Polygraph examinations began to develop in the Soviet Union in the years when any contacts with polygraphers of other countries were impossible. To develop polygraph technology, Russian specialists read books and articles of American polygraphers, trying to understand and to accept their experience in this field. The negative side of this process was a lack of information, which hampered development and use of the polygraph method. But the absence of the practical information was the positive side of that period too, because the lack of information forced specialists to study the PDD method more carefully. This has given the opportunity to develop an independent opinion about this method. That is why Russian IUP technology has some differences and novelties.

This year polygraphers are observing the 15-year anniversary of the first legal application of a polygraph in Russia, and during these years Russian polygraphers obtained some significant results.

So, what level have we achieved in Russia in the field of polygraph examination in the early 21st century?

## Polygraph examinations or IUP

The exact number of interrogations using polygraphs conducted in Russia this year or in previous years is unknown.

In 2005 state and police polygraphers carried out approximately 26,000–27,000 polygraph examinations. And at least 30,000 examinations were performed by commercial polygraphers.

According to expert opinions, polygraph examinations are increasing annually by 10–15%. Nowadays 60% of IUP are screenings, and 40% are examinations during state, police or private investigations. But these estimations

are very approximate. The main users of IUP are commercial organizations of Russia, MIA and FSS. IUP are actively used for investigations of crimes. Police officers and prosecutors are often the initiators of polygraph examinations.

At the end of 2005 the Prosecutor General's Office of the Russian Federation made a review of the practice of polygraph application during crime detections. This review came to the conclusion that the effectiveness of IUP was indisputable. And the Prosecutor General's Office recommended that all public prosecution bodies use polygraphs more actively in crime detection. The main achievement of recent years in polygraph application is the use of IUP as forensic psychophysiological expertise. The first precedent when the results of psychophysiological expertise were accepted as evidence in court took place in Russia in 2002. The law of criminal procedure of the Russian Federation allows the admissibility of the results of IUP in court as evidence. Since 2002 more than a hundred examples of such forensic expertise have been executed in different regions of Russia. The results of forensic psychophysiological expertise have been accepted by courts as evidence, including the Supreme Court of the Russian Federation. The Prosecutor General's Office believes that forensic psychophysiological expertise is the most persuasive way to use polygraph examination results in court.

## Examiners (polygraphers)

The exact number of polygraphers in Russia now is unknown. At least 300 polygraphers work in state and police organizations. And at least 200 specialists work in the private sector. Approximately 50% of Russian polygraphers are women.

In the recent past Russian polygraphers were police officers, government employees, psychologists, military men, students and so on and so forth. To be successful the polygrapher should have a good knowledge of the Russian language and national mentality. This is why there are no polygraphers in Russia who reside outside of this country.

## Training of polygraphers

In the USSR there was no polygraph training course, because the application of polygraph techniques was limited. Russian polygraphers of the first gener-



ation (the author of this article belongs to these polygraphers) are self-taught specialists. The first polygraph training course in Russia was commercial. It was opened in 1994. But the educational level was too weak and soon this course was closed. The MIA polygraph training course was founded in 1994 in Moscow, but in 1997 this training course left the capital for Krasnodar, and at the beginning the new century this MIA polygraph school started to work irregularly.

The polygraph training school of FSS was founded by the author of this article in the Institute of Criminalistics (IC FSS) in spring 1996, when polygraphers of the Institute had had more than two decades of experience in this field.

In 1998 Bill Thompson and his colleagues from Maryland Institute of Criminal Justice conducted the first polygraph training course in Russia. The course had been organized in the scope of the bilateral (Russia and USA) Cooperative Threat Reduction Program. This training course provided the opportunity to become acquainted with the educational technology of the American Polygraph Association. It was very pleasant for Russian polygraphers to find out what the technology of a Russian training course looked like compared to an American one.

The polygraph training school of the IC FSS is the only training school of federal standing in Russia, having functioned for 12 years on a regular basis and trained specialists for FSS, for other state departments, for commercial organizations in Russia and in some countries of the Commonwealth of Independent States (CIS) – Armenia, Belarus, Kyrgyzstan and Ukraine). Now commercial graduates of the school are working in more than 20 cities of the Russian Federation and Ukraine. The duration of formal training in the school of the IC FSS is about 500 hours. This training consists of lectures, laboratory training and an “internship” with professional polygraphers – schoolmasters or supervisors. Besides the polygraph training school of the IC FSS, which offers training courses regularly, there are three state (non-commercial) training schools – in MIA, in MD and Federal Penal Service.

Furthermore, there are about ten commercial polygraph training schools. Three of them hold training courses regularly, others from time to time. One of the commercial polygraph training schools is situated in Saint Petersburg, the second in Krasnodar, and all others (state and commercial) in Moscow.

The best known commercial polygraph training school in Russia is the National School of Lie Detection (NSLD), which was founded four years ago. The duration of the training course in the NSDL is about 500 hours.

Moscow State Technical University n.a. N.E. Bauman (BMSTU) – the first technical university in Russia and the CIS – begins this year a complex educational program concerning polygraphs. Its purposes are preparing polyg-

raphers, confirming the professional skills of polygraphers and teaching them to fulfill forensic psychophysiological expertise.

## Polygraphs

Like the specialists of other countries, Russian polygraphers have for a long time used the American “lie detectors”, produced by the Keeler, Lafayette and Stoelting companies, and medical polygraphs too.

It may be interesting to know that Russian polygraphers were the first to begin regularly to record a human voice on polygraph diagram paper. For this purpose a voice channel was added to American ink “lie detectors”. In the early 1980s Russian polygraphers stopped using sphygmography during polygraph testing, and for more than two decades used plethysmography. The information abilities of a plethysmography are the same, but it does not create the well known negative influence of a sphygmography on a person during testing.

The first computerized polygraph system was created in the Soviet Union in the middle of the 1980s. It was a predecessor of a computer polygraph.

In 1993 “INEX-company” created the first Russian commercial computer polygraph. The first “INEX”-polygraph had only 3 traditional channels.

Now, fifteen years later, at least 97 % of polygraphs in the country are Russian-made instrumentation. Besides these there are about two dozen American computerized and analog polygraphs of Lafayette and Stoelting.

Computerized polygraphs are manufactured by six companies, and none of them makes analog polygraphs for lie detection. These companies manufacture more than ten models of computerized polygraphs. All Russian polygraphs (like Americans) have 7, 8 or 9 channels and software-based scoring algorithms. Since 2004 the best models of Russian computer polygraphs had extra-channels of video and audio recording. Russian computer polygraphs give wide opportunities for fulfillment of an examination and are comparatively inexpensive. This is why Russian polygraphs have been bought by the polygraphers of Armenia, Belarus, Kazakhstan, Kyrgyzstan and Ukraine.

## IUP (polygraph examination) technology

The first generation of polygraphers of the Soviet Union accepted international (or American) polygraph examination technology as the only correct

technology, and diffused their polygraph knowledge in Russia. This is why most sophisticated Russian polygraphologists are now using the control question technique (CQT) and conceal information technique (CIT).

Besides accepting this technology, Russian polygraphers have critically assessed it and, as a result, have created some novelties. For example, several polygraph tests have been constructed on the basis of CQT and CIT. These tests are successfully applied both in screening and in criminal investigations.

In the middle of the 1980s the expert quantitative metrical estimation of reactions was created for Russian tests. When polygraphers estimate the IUP result, they often employed both the expert quantitative estimation and computer analysis of reactions. It should be emphasized that the last word in the decision-making process belongs to the specialist (polygrapher), and not to the computer program.

Fifteen years of domestic experience have demonstrated that if the polygraph examination is fulfilled with all methodological requirements, if the polygrapher is in no hurry and if he has enough time for examination, the tested person is unprotected before the polygraph and has no chance to conceal the information required.

The foundation of the successful polygraph application is a well-trained polygrapher who knows the polygraph examination technology that is correct and accepted by practitioners worldwide.

But in Russia there are "specialists" who have refused some fundamental requirements of polygraph examination technology. The main test and, as a rule, the only test which these "specialists" use during an examination is the searching guilty knowledge test. Such "specialists" employ this test both in screening and criminal investigations. Polygraphers, of course, strongly criticize their ugly technology, but without legal regulations can do nothing about it. Unfortunately, in the Russian polygraph community there are different points of view on polygraph examination technology.

## Legal Issues

As stated above, the first step of legal regulation of polygraph application in Russia was made in 1993. Now the use of polygraphs in federal departments is regulated by some special rules and regulations.

Russian labor legislation and federal law "Concerning a commercial secret" allows an employer to oblige an employee in some cases to take an IUP, but with the voluntary consent of employee, when commencing a job. So the

Russian legal system does not prohibit polygraph application in commercial practice. But there are no rules and regulations for application of polygraph examinations in the private sector.

Besides that nowadays there are no common standards for commercial polygraph training courses. This is why a significant percentage of persons using polygraphs in the private sector has poor training, and these “specialists” cannot be called polygraphers (professional polygraph examiners). The absence or imperfection of legal regulations for polygraph use inevitably leads to the breach of people’s human rights.

In 2007 the bill “Concerning use of a polygraph” was drawn up by experts of the Russian State Duma, and now this draft is being prepared for discussion. The law must enter for an employer two types of IUP – obligatory and voluntary. This law must guarantee the observance of the rights and liberties of a person in accordance with the Constitution of the Russian Federation. The use of a polygraph should be based on the principles of legality, humanism and confidentiality.

## Association

There is no national organization of polygraphers in Russia. Last year local associations were created in the Far East and in the Ural region of Russia.

## Political and Social Issues

IUP are developing in Russia very intensively in the state and private sectors. At present there are no public or political organizations which oppose polygraph application. And, in general, the Russian public understands the necessity of using polygraph examinations. Some Russian commercial organizations which manufacture polygraphs or train polygraph examiners have websites and advertise extensively their production. The number of articles about polygraphs in newspapers and magazines is increasing from year to year. But polygraph testing results in specific cases are not practically reported in the media.

## Scientific researches

When speaking about polygraph examination, many specialists see it as a method of diagnostic emotional or psychic stress. Theoretical and experimental research of domestic scientists, conducted in the 1970s and 1980s, defined the psychological and neurophysiologic mechanisms of the disclosure information hidden by a person. In 1988-1989 it was proved that IUP is a complex psychological-psychophysiological method of the research of traces of the events which are kept in a human memory (4). The traces of events, kept in the emotional memory, practically cannot be destroyed during the lifetime of a person. It is very important that a person can do nothing about these traces.

Real polygraph examinations confirmed that the traces of such an event could be found in the human memory decades after this event has happened. For example, in the real practice of the author of this article, a polygraph helped to discover the traces of events which had happened about 20 years previously.

This achievement had a great importance for understanding the place and role of IUP in the system of modern Russian criminalistic science and opened the way to using the polygraph as psychophysiological expertise.

In the middle of the 1990s scientific research was made to develop screening IUP, and this kind of IUP has become widespread in the country.

Now there is no academic or private research work. But at the beginning of the 21<sup>st</sup> century the economic situation has flattened out in Russia and it is expected that in the nearest future some universities, institutes and, maybe, private companies will begin scientific research in the field of polygraphy.

In conclusion it is necessary to say that this article gives only a short review of polygraph application in Russia, which is developing very intensively. Some aspects of this subject area – such as Russian polygraphs, IUP technology, fulfillment IUP as psychophysiological expertise and so on – may be the subject of separate articles.

## References

1. Luria, A. R., 1984, *Diagnostics of affect traces. Psychology of emotions*, Texts, M., MGU (Moscow State University), p. 231.

2. Kholodny, Y. I., 2007, *Criminalistic diagnostic investigation using a polygraph*, Polygraph, 5, 36, p. 91–99.
3. The article, Kholodny Y. I., *Criminalistic diagnostic investigation using a polygraph*, was published, as 15-th chapter, in the book *Criminalistics*, Textbook, M., Unity-Dana, 2008, 943 p.
4. Kholodny, Y. I., 2006, *Interrogation using polygraph and its scientific basis*, Polygraph, 5, 35, p. 1–21.



Mark Handler, Raymond Nelson\*  
Utah Approach to Comparison Question Polygraph Testing  
USA

## Test Structure and Administration

The Utah-CQT begins as other testing procedures do, with the pre-test interview, conducted in a non-accusatory manner. The examiner should obtain the necessary test release that includes a brief statement of allegations or issues to be resolved, and if applicable, a statutory rights waiver and then collects general biographical and medical information from the examinee. Rapport-building discussion gives the examiner a chance to evaluate the examinee's suitability for the examination. Interaction with the examinee also gives the examiner the chance to do a rough assessment of the examinee's verbal and mental abilities that will later be used to help word the examination questions. In the PLT version, the examiner uses this period of conversation to develop material for comparison questions to be used during the testing phase of the examination, although the nature of the issues to be resolved usually dictates the general content of the comparison questions. The examiner does not, however, lecture the examinee regarding past transgressions. This portion of the interview is conducted with open-ended questions and the careful use

---

\* raymondnelson@gmail.com

of suggestions as opposed to an interrogation of past deeds. The examiner points out any monitoring or recording devices in the examination room and explains the purpose for having the exam monitored and/or recorded. In the Utah-CQT approach all examinations should be recorded in their entirety. In an age in which video and audio recording technology is easily available and fully integrated into all modern field polygraph systems, there is no reason to forgo the advantages of a complete video and audio recording of all polygraph examinations. It is only through complete recordings that meaningful quality assurance is possible. Frankness regarding monitoring devices helps assure the examinee that the test will be conducted in a professional manner and may assist in convincing the examinee that the examiner is being open and truthful. Brief explanation of any quality assurance program also assists in establishing a professional and trustworthy atmosphere.

The examiner advises the examinee of the general nature of the allegations and the specific issues to be resolved by the examination. The examinee is then given the opportunity to provide a “free narrative” to discuss his or her knowledge of and/or role in the incident. The goal of the free narrative discussion is to obtain information from the examinee without confrontation or undue stress.

In general the examiner should allow the examinee to tell his or her story without interruption. The examiner informs the examinee of the case facts in a low-key approach and should advise the examinee that these are allegations and ensure the examinee understands the difference between allegations and facts known to be true.

The examiner should note inconsistencies or other matters to which he or she may wish to return once the examinee finishes the narrative. The examiner does not argue with the examinee nor does the examiner challenge the examinee’s version of the case facts. The examiner encourages the examinee to be candid in order to formulate the test questions in a succinct and clear manner.

In polygraph screening or monitoring programs (i.e., LEPET, security, PCSOT), the Utah-CQT may be used as a mixed-issue (multiple-issue) examination, similar to the AFMGQT, in the absence of a known allegation or known incident. In these programs discussion of the known allegation or known incident will be replaced with a structured interview protocol, which addresses content areas pertinent to the risk or compliance issues under investigation. It should be noted that these applications of polygraph testing have not been investigated as thoroughly as other uses, and scientific investigation and verification of such uses are more limited.



This low-key, non-accusatory approach presents the examiner as a neutral seeker of the truth and helps to allay fears of pre-conceived guilt. If there are inconsistencies or other matters that require follow-up or clarification before the examination, they are discussed at this time in a non-confrontational fashion.

After the narrative and the discussion of any other issues, the components are placed on the examinee. During this process, the functions of the various polygraph component sensors are discussed, and a general explanation of the psychophysiology that underlies the polygraph test is provided. This may be done through a general discussion of the anecdotes that illustrate psychophysiological responding and various possible causes of arousal (Handler & Honts, 2007). The goal of this portion of the interview is to ensure in the examinee an understanding that lying will inevitably be associated with physiological response.

Once the components are placed on the examinee, the examiner conducts an acquaintance test. The acquaintance test is generally a known-solution peak of tension test that is used to demonstrate the efficacy of the polygraph examination. Other approaches to the acquaintance test are not prohibited and would not invalidate an examination. In the known-solution acquaintance test, the examinee is told to select a number such that there will be some additional or padding questions before and after the selected number. This can be accomplished by directing the examinee to select a number between 3 and 6 and write that number on a piece of paper. The paper may then displayed in front of the examinee and the examinee is instructed to deny picking any number between 1 and 7 while the polygraph records his or her physiological reactions. The acquaintance test allows the examiner to ensure the production of adequate quality recordings and to take corrective actions to remedy any lack thereof.

The examiner can use the acquaintance test during the question review to demonstrate to the examinee that he or she is a suitable candidate for polygraph, and provide assurances that successful completion of the examination can be obtained by answering all of the test questions truthfully (in the PLC version of the examination) or that clear indication was found when the examinee was not answering truthfully (in the DLC version).

Following the acquaintance test the test questions are reviewed with the examinee for clarity. Some agency or local testing protocols may specify that the test questions be fully reviewed prior to attaching any components to the examinee. There is no theoretical rationale to suggest this difference would invalidate an examination result. Attaching the sensors earlier may allow them to stabilize, especially the electrodes for electrodermal recording.

The examiner begins with the sacrifice-relevant question followed by the relevant questions. The sacrifice-relevant question is used to introduce the relevant issue under investigation during the testing and is not scored. In investigative polygraph testing, relevant question targets are dictated by the circumstances of the investigation and are commonly formulated around the most salient or intense aspects of the allegation. In screening programs, relevant questions should describe the examinee's involvement in possible behavioral concerns to risk managers or adjudicators and should be designed to add incremental validity to their particular program.

Polygraph screening targets would ideally be selected to investigate content areas pertinent to actuarial or empirically derived protocols for risk assessment and risk management. The fundamental requirement for relevant question target selection is that the behavioral issue of concern provides information useful to the referring authority.

Effectively formulated relevant questions will directly assess the examinee's behavioral involvement in the issue of concern. Relevant questions should not introduce confusion through the use of language or concepts pertaining to psychological motivation or intent, as these are thought to introduce dimensions of excuse or rationalization on the part of examinees or skillful liars. Conversely, truthful examinees may produce spurious reactions because of the ambiguity and lack of concreteness of such questions. Direct questions with a simple grammatical structure are the best approach. Relevant questions should be free of idiomatic and legal jargon that is unfamiliar to the examinee, and should not include issues of psychological assessment or inference. Relevant questions are simple questions that can easily be answered "yes" or "no." Reluctance, on the part of the examinee to provide a simple answer to a simple question may be an indicator of a non-testable issue or an examinee who is unable to disambiguate the issue. Discussion and resolution of this should be non-accusatory, but persistent enough to achieve a simple testable answer to a question that is behaviorally descriptive of the examinee's possible involvement in an issue of concern. The prevailing practice preference for relevant questions is they are usually answered "no", though certain exceptions have been suggested such as alleged victims of severe sexual assaults (Hardy & Murphy, 1996). The current authors found nothing to support that using "yes" answered relevant questions would invalidate a test.

Next the examiner introduces the comparison questions. PLC questions are presented to the examinee as being necessary for further evaluating the examinee's character and the issue under investigation. PLC questions are based on transgressions whose subject matter is generally or conceptually related to the allegations of the examination and which virtually all persons

may have committed, but which are likely to be denied in the context of the examination. PLC questions are broad in scope and usually based on actions categorically similar to that of the issue under investigation. That is, relevant questions on theft would normally be associated with comparison questions about theft or general honesty. Relevant questions about violent acts are typically associated with comparison questions about causing harm. Standard comparison question construction, as taught in polygraph schools accredited by the American Polygraph Association and American Association of Police Polygraphists, is recommended for ensuring saliency. There is no reason, however, to prohibit the use of standard “lie” comparison questions in nearly any testing context.

Comparison questions in the Utah-CQT are traditionally “exclusive” in that they are separated from the relevant issue by time, place or category. Comparison questions not separated from the relevant issue are sometimes referred to as *non-exclusionary* type. Three studies (Horvath, 1988; Amsel, 1999; Palmatier, 1991) failed to establish any clear and consistent advantage of exclusionary comparison questions over non-exclusionary questions (Krapohl, Stern & Ryan, 2003). Podlesny & Raskin (1978) showed some superiority for exclusionary questions, in that Skin Conductance Response (SCR) half-recovery time, SCR recovery half time width and Skin Potential Response (SPR) amplitude were significantly more effective with exclusive comparison questions.

Podlesny & Raskin (1978) also reported that both types of comparison questions produced significant identification of innocent examinees, but only exclusive comparison questions produced significant identification of guilty examinees using numerical scores. Collectively these reports suggest that exclusionary comparison questions may hold no advantage over non-exclusionary comparison questions when data are evaluated using reaction criteria typically employed in field testing (for descriptions of those reaction criteria, see: Bell et al 1999; Handler, 2006; Raskin & Honts 2002; Kircher et al., 2005). The use of exclusionary comparison questions may avoid possible criticism that the PLC questions are also relevant and may cause a false negative result. The current authors found nothing to suggest a test would be invalid should an examiner choose to employ non-exclusionary type comparison questions.

As in other CQT techniques, the examinee is strongly, but indirectly, discouraged from making admissions to PLC questions. If the examinee makes an admission to a PLC question, the examiner notes that admission with some dismay, “*Really, you did something that would make me think you are a thief,*” and either minimizes the admission, “No, I am only concerned about serious things,” or modifies the comparison question. An example of

the latter is: “*Other than what you told me about*, before this year did you ever lie to anyone who trusted you?” Note the italicized modifier preceding the comparison question. The ultimate goal is to discourage admissions to PLC questions to ensure that the examinee perceives them as ambiguous and broad in nature. It is also important the examiner imply to the examinee that lying to any of the relevant or PLC questions will result in a failure of the polygraph test and the conclusion of deception to the relevant issue under investigation.

The examiner then introduces and reviews the neutral questions which provide time to return to a baseline when there is distortion or a physiological reaction to a specific question. Kircher, Kristjansson, Gardner, & Webb (2005) suggest inter-question intervals following a strong cardiovascular response should be increased to a minimum of 35 seconds to allow recovery, or a neutral question inserted. In general, the preferred approach is to wait to allow a return to, or at least toward, baseline levels.

The neutral questions should be non-emotional in nature and are generally answered “yes” to ensure the examinee is paying attention to the test questions. There is nothing to suggest, however, that an exam in which any neutral question is answered “no” would be invalid. The examiner may review additional neutral questions in case they are needed during testing to re-establish a baseline tracing.

The examiner next reviews the introductory question that is similarly worded to one of the “symptomatic” questions used in other CQT formats. The introductory question attempts to assure the examinee that no un-reviewed questions will be asked during the examination and may allow an orienting response at the beginning of an examination.

Research by Honts, Amato & Gordon, (2004) has failed to demonstrate the symptomatic question functions as described and may actually produce poorer accuracy, especially for innocent examinees. The consistent trend illustrated by these investigators and others suggests that the invention and addition of new types of questions should not be encouraged in an age of modern scientific polygraph testing unless research shows the efficacy of a new approach (Hilliard, 1979).

While it is wise for field examiners to adhere to the general principles and procedures taught in basic training, there is equal or greater wisdom in adapting field practices to conform to modern approaches with proven validity. We do not believe that minor departures from the above question sequences would cause a test to be invalid, and the varying formulations of the Utah-CQT since its emergence suggest that the scientists who developed the Utah-CQT method did not seek validity through simplistic adherence to

a “paint-by-numbers approach”, but sought demonstrable validity through the construction of CQT methods according to sound testing principles.

### Three-question format

The Utah-CQT has two versions, a three-question version and a four-question version (Raskin & Honts 2002).

The three-question version was the first designed and was primarily used for single-issue testing but can also be used for multiple-facet testing of a single known allegation. The three-question version of the Utah-CQT allows a great degree of flexibility in relevant question format.

The following describes an example of question numbering and type of question used in the three-question version of the Utah-CQT. For a *single-issue* examination, there will be three relevant questions, each slightly reworded.

#### Example of a Utah PLT 3-question wording

For an event-specific, *single-issue* test surrounding a bank robbery occurring last Thursday, one might ask the following questions:

Introductory 1	Do you understand I will only ask you the questions we discussed?
Sacrifice Relevant 2	Regarding whether or not you robbed that bank do you intend to answer all of these questions truthfully?
Neutral 1	Are the lights turned on inside of this room right now?
Comparison 1	( <i>Before turning X</i> ), Did you ever do anything that was dishonest or illegal?
Relevant 1	Did you rob that bank located at ____ in Austin?
Neutral 2	Are you now physically located within the State of Texas?
Comparison 2	( <i>Between the ages of X and Y</i> ), Did you ever take anything that did not belong to you?

Relevant 2	Did you rob that bank located at ___in Austin last Thursday?
Neutral 3	Do you sometimes listen to music while riding in a car?
Comparison 3	Did you ever take anything from a place where you worked, ( <i>before age X</i> )?
Relevant 3	Did you rob that bank at ___ on ___?

The examples above are shown with the exclusionary clause of the comparison question in brackets.

For a *multiple-facet* examination, the examiner has a choice of asking two reworded relevant questions with the same meaning and another relevant question that is directly related to the issue under investigation. This third relevant question can be an evidence-connecting, guilty knowledge or secondary involvement question.

A third alternative is to ask three separate relevant questions relating to the same specific issue under investigation. Readers are reminded that research has shown that accuracy rates are higher for tests in which the examinee is either completely truthful or deceptive to all of the test questions as opposed to just some of them (Honts, Kircher, & Raskin, 1988; Raskin, Kircher, Honts, and Horowitz, 1988; Barland, Honts and Barger, 1989).

The current authors would strongly recommend that examiners, if possible, attempt to limit the examination to one in which the examinee is truthful or deceptive to all of the relevant questions.

If one were to construct a *multiple-facet* polygraph examination surrounding a single crime event involving a bank robbery, examples of alternative relevant questions may be:

Introductory 1	Do you understand I will only ask you the questions we discussed?
Sacrifice Relevant 2	Regarding whether or not you robbed that bank do you intend to answer all of the questions truthfully?
Neutral 1	Are the lights turned on inside of this room right now?

Comparison 1	Did you ever steal anything from someone who trusted you?
Relevant 1	Did you rob that bank at ___ on ___?
Neutral 2	Are you now physically located within the State of Texas?
Comparison 2	Did you ever steal anything from a friend or family member?
Relevant 2	Did you plan or arrange with anyone to rob that bank at ___?
Neutral 3	Do you sometimes listen to music while riding in a car?
Comparison 3	Did you ever steal anything from a place where you worked?
Relevant 3	Did you participate in any way in the robbery of that bank?

Note that this example is provided with *non-exclusionary* comparison questions.

### Four-question format

The four-question format is similar in design to a version of the Air Force Modified General Question Technique (DoDPI 2006) using pairs of relevant questions that are bracketed by comparison questions. This allows the examiner greater flexibility covering more than one aspect of the relevant issues and in scoring by using the surrounding comparison questions. The relevant questions can range from one to four distinct behavioral aspects or facets of a single crime or allegation. The question construction rules are the same as those described above for the multiple-facet version of the three-question version.

The following describes an example of question numbering and type of question used in the four-question version.

I1	Introductory
SR2	Sacrifice Relevant
N1	Neutral
C1	Comparison
R1	Relevant
R2	Relevant
C2	Comparison
N3	Neutral (optional) This neutral question may be inserted at the option of the examiner to allow some decrease of tension and recovery to baseline. If inserted, the examiner will skip over this neutral question during scoring.
R3	Relevant
R4	Relevant
C3	Comparison
N2	Neutral

## In test operation

The examinee is instructed to sit still and answer each question truthfully. However, the approach is to avoid doing this in a heavy-handed manner. For example the following admonition would be typical for this approach:

“I need you to sit still during the asking of the questions. Movement will create distortion and artifacts in the recordings that will require me to repeat the questions and that will make the test longer.”

The examiner rotates the neutral, comparison, and relevant (if desired) questions during the next and subsequent presentations. The examiner may prefer to leave the relevant questions always in the same position, and rotating only the comparison and neutral questions, making it easier to score the charts by having a fixed order of relevant questions. Moving the questions helps to prevent pattern recognition and anticipation of a specific order of questions during the examination.

The following are examples of serial positioning in the question strings showing one example of question rotation.



### Three question version

First Chart	I1, SR2, N1, C1, R1, N2, C2, R2, N3, C3, R3
Second Chart	I1, SR2, N2, C3, R2, N3, C1, R3, N1, C2, R1
Third Chart	I1, SR2, N3, C2, R3, N1, C3, R1, N2, C1, R2

### Four question version

First Chart	I1, SR2, N1, C1, R1, R2, C2, N3 (N3 is optional), R3, R4, C3, N2
Second Chart	I1, SR2, N2, C2, R1, R2, C3, N3 (optional), R3, R4, C1, N1
Third Chart	I1, SR2, N1, C3, R1, R2, C1, N3 (optional), R3, R4, C2, N2

As can be seen above, each relevant question has an opportunity to be compared to each comparison question across the three chart series. As discussed above, if the results are inconclusive after three charts, two additional charts are run. The examiner may simply use the first and second serial positioning question strings for the fourth and fifth chart.

After the third chart, the charts are numerically scored. However, the examinee is only told that the examiner always stops at this point to carefully check the quality of the recordings before collecting more charts. If the scores meet the threshold of the decision criteria, the data collection phase is complete. If the test result is inconclusive following the first three charts, two additional charts are conducted following the same rotational patterns described above. Following the fifth chart, all scores are totaled to make a determination of veracity. The Defense Academy for Credibility Assessment (DACA), the Federal Training facility (2006) permits the examiner to conduct just a fourth chart, and if necessary a fifth test chart. We are aware of no theoretical rationale or evidence to suggest that this procedural difference would invalidate a test result.

The questions are presented to the examinee at least three times across three charts, with a brief discussion between charts to clarify and resolve any perceived problems raised by the examinee and to reinforce a focus on both the relevant and comparison questions (Raskin & Honts, 2002). After each presentation of the test questions, the examiner should ask the examinee if he or she has any concerns with the test questions. Honts (1999) reviewed data from 19 studies that involved 1092 polygraph tests and found between chart discussion (even when limited to only the comparison questions) reduced the

risk of error. In 9 of the 11 studies which included examinations in which the questions were reviewed between charts both the relevant and comparison questions were reviewed. In 2 of those studies, only the comparison questions were reviewed. Honts showed that between-chart stimulation and question review reduced the false negative rate (54%), had a modest reduction of false positive rates (2.9%) and a substantial decrease in inconclusive outcomes for truthful examinees (42%).

The following is typical of the type of exchange that might take place between charts.

Note how the examiner places equal emphasis on each group of questions during the stimulation and review.

Examiner: OK Roy, did you have any problems with any of those questions on the test?

Roy: No.

Examiner: Anything come to mind when I asked you those questions?

Roy: No.

Examiner: How about those questions about the drug transaction? Is it clear what I am asking you? Do you understand them?

Roy: Yep.

Examiner: How about those questions about lying? Any problem with any of those?

Roy: Nope.

There has been controversy surrounding the review of question between test charts. Abrams (1999) and Matte (2000) argued that review of comparison questions between tests is incorrect and Offe & Offe (2007) found no contribution to improved or degraded decision outcomes as a result of between test review of the test questions. In consideration of these findings, we feel it prudent to recommend a review of the questions between each chart, but find no reason to support an argument that the inclusion or exclusion of this review would cause a test result to become invalid or erroneous. Honts (1999) did not speculate as to the psychological cause of these findings. He

correctly stated that “The essence of science is empiricism. That is, scientific knowledge is built on data, not speculation nor authority.” Honts chose to accept the data for what it stated on its’ own merit.

Should an examinee make additional admissions to comparison questions or need to modify a relevant question, the examiner should do so and re-label the question. For example, if during a Utah PLT the examinee makes an admission to question C1 “Before this year did you ever steal anything from a business,” the examiner can modify that question to “Other than what you told me about, before this year did you ever steal anything from a business” and label that question C1a. The examiner should then review all test questions with the examinee. The examiner then conducts the next two charts and again starts by instructing the examinee to sit still and answer all of the questions truthfully.

## Test data analysis and decision criteria

The Utah Scoring System (Bell, Raskin, Honts & Kircher, 1999) is a simplified version of the numerical scoring techniques introduced by Backster in 1963 and modified by the US Army around 1970 (Weaver 1980; Swinford 1999).

The Utah scoring system is a simple and elegant scoring system designed to improve accuracy, reduce inconclusive results, and improve interrater reliability. It has fewer rules to follow and fewer criteria to score than the other scoring systems currently in use. The Utah Numerical Evaluation Scoring System was designed, refined and tested by Raskin and his colleagues.

The Utah scoring System is based on physiological response data that has been proven to be a valid and reliable indicator of sympathetic arousal. The inter-scorer correlations of results produced using the Utah Scoring system are typically around 0.90 (Bell et al., 1999). The accuracy of the Utah Scoring system from several analog studies was 90%, as reported by Bell et al., when averaged for programmed innocent and guilty examinees. The results of field studies using the Utah scoring system are consistent with analog study results (Bell et al., 1999).

Numerical evaluation of the test data is accomplished by comparing the relative strengths of responses to comparison and relevant questions. The Utah system uses a 7-position numerical scoring approach. The relative strengths of physiological reactions for each sensor are compared and a score is assigned. The possible scores range from -3 to +3. The reaction of each relevant question is compared to the reaction to the preceding comparison question in the 3-question CQT format or the stronger of the two surrounding

comparison questions in the 4-question MGQT format. If the relative strength of the relevant question is greater than that of the comparison question, a negative value is assigned.

Conversely if the comparison question strength exceeds the relevant question strength, a positive score is assigned. If there is no observable difference, a zero is assigned. In some components there are minimum relative ratios that must be achieved in order to assign a score.

For the three-question version shown above, the relevant question is normally compared to the preceding comparison question for evaluation. If the preceding comparison question is distorted by an artifact, the examiner may use the closest artifact-free comparison question for evaluation.

For the four-question version shown above, the examiner compares the relevant question to the two bracketing comparison questions, component by component. For example, in the first chart of the four-question version shown above, R1 is compared to C1 and C2. The examiner will find the strongest reaction channel separately of each channel for C1 and C2 and use that to compare to the corresponding channel of R1. Using the reaction of the stronger bracketed comparison question has been shown to produce valid field results (Honts 1996; Raskin et al., 1988).

Physiological tracings that are affected by artifacts are excluded for evaluation purposes. If the examinee answered "yes" to a comparison question during the test, the comparison question response may be used in scoring as long as the reviewed answer had been "no" (see Honts, Raskin & Kircher, 1992). The examiner may insert a neutral question routinely after the second comparison question or any other time needed to reestablish tracing stability. During test data analysis, the examiner will skip over that neutral question.

The Utah Scoring System uses a total of seven primary scoring criteria in the respiration, cardiograph, electrodermal, and peripheral vasomotor activity channels.

Values of -3, -2, -1, 0, +1, +2, and +3 are assigned by channel to each relevant question. As mentioned above, if the relevant question is the larger of the two, the score will be a negative number. If the comparison question is the stronger of the two, the score will be a positive number, and no difference yields a score of zero.

Only one score of +/-3 can be assigned per chart, in the cardio and electrodermal channel, and only if the baseline for the channel is stable and the reaction is the largest in that channel on the chart. The relevant question totals are calculated after three charts and, if inconclusive, after five charts.

For the respiration channel, there are four empirically confirmed features that are considered diagnostic (ASTM 2005). Three of those features are captured

by the phenomenon known as Respiration Line Length “RLL” (Timm, 1982). RLL is simply the measurement of the length of the respiration line for a fixed period of time. The total line length for the designated period of time between the relevant and comparison question or questions is compared. The greater the suppression the shorter the line length and thus the stronger the response. Those three features are suppression of respiration amplitude (Figure 2), reduction in the respiration rate (which includes changes in the inhalation/exhalation ratio if they result in respiration rate decreases, Figure 3) and apnea occurring near the exhalation cycle (Figure 4). The fourth respiration criterion is a temporary rise in the baseline of the tracing.

A respiration tracing is considered to be diagnostic if there are at least three successive cycles of an RLL feature or temporary baseline arousal. The exception to this is apnea, where there may not be any discernible cycles of respiration. While the thoracic and abdominal respirations are recorded separately, a single value is assigned. That value is based on the noted combined difference between the relevant and comparison questions.

The developers of the Utah-CQT have taught and practiced conservatism when evaluating the respiration channel. Bell and his colleagues used a sample of 50 polygraph examinations to conduct a survey that provided 450 numerical scores. Bell et al. (1999) tallied those scores to determine the distribution of scores, and reported that respiration scores of 0 were assigned about 75% of the time. Scores of  $\pm 1$  were assigned about 20% of the time, and  $\pm 2$  or 3 less than 5% of the time.

For the electrodermal channel, scores are based primarily on a comparison of the peak amplitude (Figure 6), a criterion that has been empirically shown to be diagnostic. Amplitude is measured from the pre-stimulus baseline to the highest peak achieved within the scoring window (Bell et al., 1999). The ratio of the relevant and comparison question is calculated.

A score of  $\pm 1$  is assigned if the relative strength is twice as large, a score of  $\pm 2$  is assigned if the relative strength is three times as large and a score of  $\pm 3$  is assigned if the relative strength is four times as large. If the electrodermal tracing is labile, a score of 3 should not be assigned.

Duration of response and complexity can be considered as secondary reaction criteria. Reactions that have clearly longer duration or complexity may increase a 0 to a  $\pm 1$  or a  $\pm 1$  to a  $\pm 2$  (Figures 7 & 8). If the amplitude ratios are at least 1.5:1 with complexity over no complexity or increased duration of reaction time, this allows an increase of a score of 0 to  $\pm 1$ . Similarly, a ratio of at least 2.5:1 to increase a score of  $\pm 1$  to  $\pm 2$  following the same rules regarding increased complexity or duration. Bell et al. noted in the electrodermal channel scores of 0 were assigned about 50% of the time,

scores of  $\pm 1$  about 25%,  $\pm 2$  about 20% and  $\pm 3$  less than 10% of the time. Krapohl and Handler (2006) demonstrated that additional diagnostic information can be gained from interpreting smaller differences in response to relevant and comparison questions using federal ZCT examinations. While these were not of the Utah-CQT format, the current authors know of no reason to believe improved ratio values for scoring would not be amenable in a Utah-CQT format. Improved ratios are being investigated at the time of writing.

For the relative blood pressure channel, relative strengths of reactions are assessed based on upward movement from baseline (baseline arousal) as shown in Figure 9. A minimum ratio of 1.5:1 is required for a score of  $\pm 1$ ; a ratio of 2:1 for a score of  $\pm 2$ ; and 3:1 for a score of  $\pm 3$ . Bell et al. (1999) reported duration of the response may be considered when evaluating the relative strength of the reaction and a reaction with greater duration may increase a score from 0 to 1 or from 1 to 2 (Figure 10).

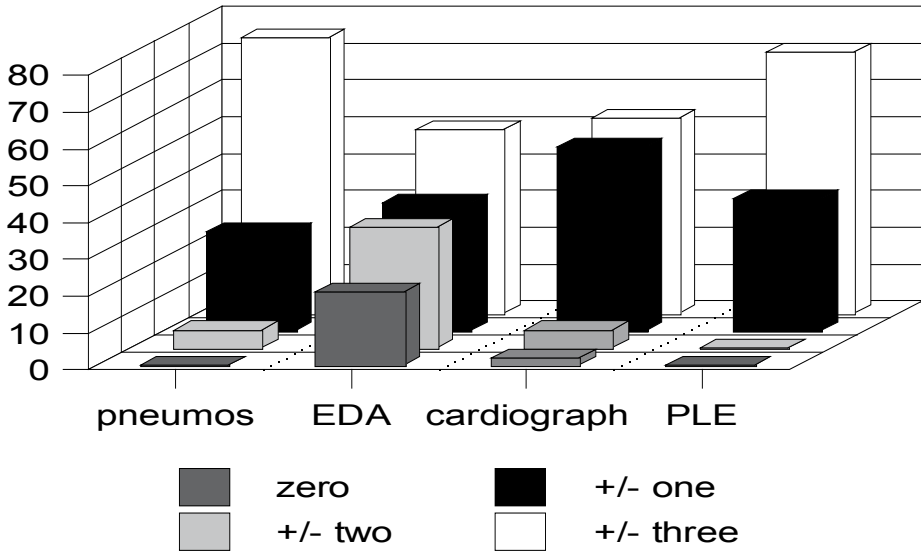
They did not, however, discuss the procedure for applying this rule and the current authors suggest limiting scoring of this channel to baseline arousal as suggested by Kircher et al (2005), Harris, Horner and McQuarrie (2000) and ASTM (2002). Bell et al. noted in relative blood pressure scores of 0 were assigned about 50% of the time, scores of  $\pm 1$  about 45%,  $\pm 2$  less than 5% of the time. Scores of  $\pm 3$  are rare and only one such score can be assigned per chart as explained in the electrodermal section (Bell et al. 1999).

For the peripheral vasomotor activity, the relative strength of the reactions is assessed by comparing the reduction in pulse amplitude (Figure 11). The source of this channel is a photoplethysmograph monitoring reduction in finger pulse amplitude. Numerical scores are based on the duration and degree of amplitude reduction.

Scores may be assigned when there is no difference in amplitude decrease but a discernible difference in the duration of the reactions (Figure 12). Bell et al. noted in scoring that finger pulse amplitude scores of 0 were assigned about 70% of the time and scores of  $\pm 1$  about 30%, (Bell et al. 1999).

The following graph shows the distribution of the numerical scores obtained during the survey by Bell et al. al (1999). As can be seen from the graph the majority of numerical scores assigned are zero or  $\pm 1$  for most channels.

Figure 1. The distribution of scores from the Bell et al. survey



## Decision criteria

The examiner proceeds through the charts and totals the score for each relevant question on each chart. The total score of each relevant question for the first three charts is then determined. For single-issue tests where the examinee must be truthful or deceptive to all of the relevant questions, the cutting score is  $\pm 6$ . In other words, when there is a grand total of  $+6$  or greater, the result is truthful. A grand total of  $-6$  or less would result in a determination of untruthful or deception indicated. Scores falling between  $-5$  and  $+5$  would result in a determination of inconclusive and the examiner would conduct an additional two charts as described above.

Following those two additional charts, the relevant question scores are once again totaled. The cutting scores of  $\pm 6$  remain the same for five charts.

The decision criteria are slightly different for multiple-faceted examinations where the examinee may be truthful to some, but not all, of the relevant questions. If the spot totals for all relevant questions are either all positive or all negative (ignoring spot scores of zero), use the  $\pm 6$  Grand Total rule described above for single-issue tests.

If any of the spots are opposite (some positive and some negative, again ignoring spot scores of zero), then use a Spot Score Rule (SSR) for each spot. The SSR is that each spot total must be +3 for a conclusion of no deception indicated (NDI), and any one spot total of -3 or less calls for a decision of deception indicated (DI) to the examination. However, if decisions are made on individual questions caution is warranted as research indicates when examinees answer some questions truthfully and some deceptively the accuracy for calls on individual questions is reduced (see the discussion in Raskin & Honts, 2002). The problem can be exacerbated when attempting to verify truthfulness to one or more questions when total scores for any relevant question have indicated deception (Raskin & Honts, 2002).

While there has been a consistent effort to evaluate the empirical validity of various cut scores, little emphasis has been placed on the determination of statistically determined cut scores, in the manner of a Gaussian signal detection model, as described by Barland (1985). Krapohl and McManus (1999), Krapohl (2002) and Nelson, Handler and Krapohl (2007) are exceptions to this trend.

## Utah directed-lie test

DLC questions are those which the examiner instructs the examinee to answer falsely (Honts & Raskin, 1988; Raskin & Honts, 2002). DLC questions may offer some relief to potential problems identified in PLC versions of polygraph testing. Examiners may experience difficulty in standardizing comparison questions in the PLC version. Each examinee brings with them their own life experiences and idiosyncrasies that may hamper maintaining a rapport while attempting to lay foundation for and set the PLC questions. Examinees who have prior polygraph experience or those who have researched polygraph techniques may not be naïve to the PLC principles. This sophistication could make laying the foundation for the comparison questions challenging. Non-naïve examinees may acquiesce to the procedure in order not to seem obstreperous in which case the PLC questions become similar to DLC questions.

DLC questions are easily standardized, require little psychological manipulation and have greater face validity.

Standardization and simplification of any technique can serve to increase inter-rater and test-retest reliability, and both of these dimensions constrain the potential validity of a technique. Excessive variability in test administration



or interpretation will necessarily compromise the reliability and validity of any test method. Inter-rater reliability is a concern that will remain of paramount importance to questions about polygraph validity. When standardized practices are based on principles that are consistent with validated constructs and data obtained through the objective study of data, we can more reasonably anticipate that improvements will contribute meaningfully to the test design goal of criterion validity and decision accuracy.

There are far fewer field and laboratory studies that address validity of the DLC than the PLC. However, the results of existing studies (Barland, 1981; Barland et al., 1989; DoDPI Research Division Staff, 1997; DoDPI Research Division Staff, 1998; Honts & Raskin, 1988; Horowitz, Kircher, Honts & Raskin, 1997; Kircher, Packard, Bell & Bernhardt, 2001; Reed, 1994; Raskin & Kircher, 1990) suggest that the DLC questions perform as well or better than PLC questions. DLCs require less complex administration practices than those associated with the PLC approach and offer greater potential for standardization. Studies using DLC techniques (DoDPI Research Division Staff, 1997; Research Division Staff, 1998) suggest that a DLC approach and other improvements in test administration structure and decision policies contributed significantly to polygraph testing program objectives of sensitivity to deception and specificity to truthfulness.

There are certain caveats that attend the use of DLC testing. First, examiners with no familiarity with DLCs should seek instruction in their proper development and introduction. Second, there is some indication in the research data that at least some examinees show unusual respiration responses with the DLC (see Horowitz et al. 1997; Kircher 2001). However, standard numerical scoring procedures in Horowitz et al.(1997) performed well with the DLC, although it may not be optimal and research exploring this issue is currently underway. Moreover, there are currently no computer algorithms available that have been trained on DLT data. Therefore, the results of those models should be viewed cautiously.

## Examples of scoring criteria

### RESPIRATION

Figure 2 below shows an example of suppression of respiration amplitude.

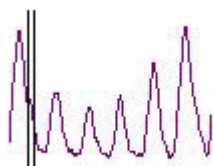


Figure 3 below shows an example of reduction in respiration rate.

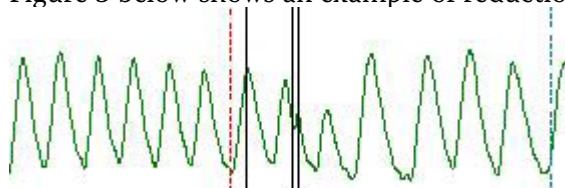
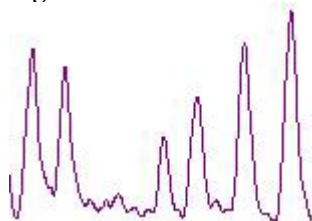
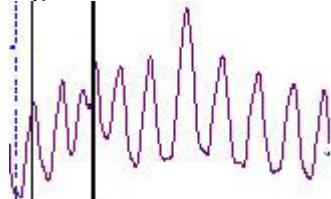


Figure 4 below shows an example of apnea occurring at or near exhalation.



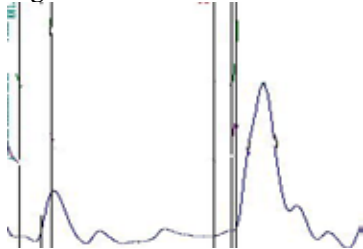
Note: The above three reaction criteria are those that are captured by the phenomenon known as RLL.

Figure 5 below shows an example of temporary baseline arousal.



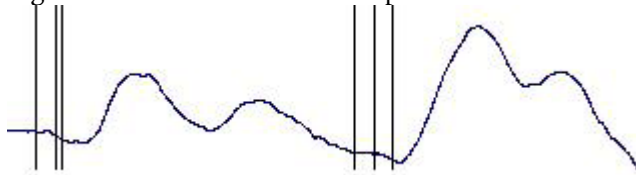
## ELECTRODERMAL ACTIVITY

Figure 6 below shows an example of amplitude Increase.



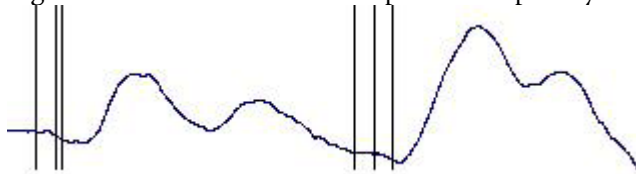
Note the ratio of the above tracings is about 2.6:1 (26 mm vs. 10 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing. This ratio would qualify for a score of  $\pm 2$  based on the increased duration and complexity.

Figure 7 below shows an example of increased duration.



Note the ratio of these tracings is about 1.8:1 (18 mm vs. 10 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing. The amplitude ratio does not exceed the 2:1 normally required for a score of  $\pm 1$ . This would qualify for a score of  $\pm 1$  based on the ratio of at least 1.5:1 with greater duration and complexity.

Figure 8 below shows an example of complexity.



Note the ratio of the tracings in figure 8 is about 1.8:1 (18 mm vs. 10 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing. The amplitude ratio does not exceed the 2:1 normally required for a score of  $\pm 1$ . This would qualify for a score of  $\pm 1$  based on the ratio of at least 1.5:1 with greater duration and complexity.

## CARDIOGRAPH

Figure 9 below shows an example of baseline arousal.

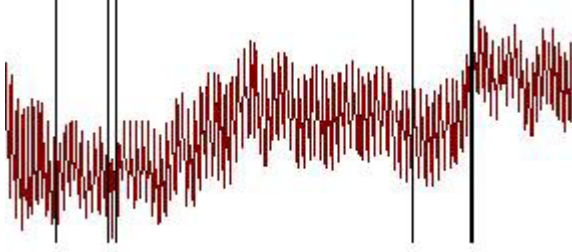
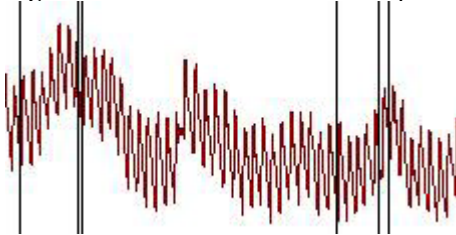


Figure 10 below shows an example of increased duration of response.



Note the ratio of the tracings in figure 10 is about 1.3:1 (8 mm vs. 6 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing.

The amplitude ratio in figure 10 does not exceed the 1.5:1 normally required for a score of  $\pm 1$ . This ratio would qualify for a score of  $\pm 1$  based on the greater duration observed in the first reaction.

## PHOTOPLETHYSMOGRAPH

Figure 11 below shows an example of amplitude reduction.

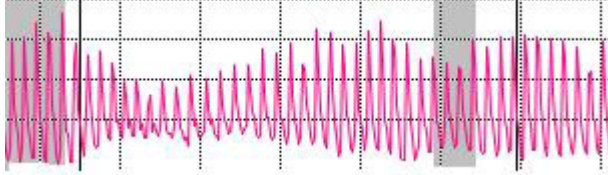
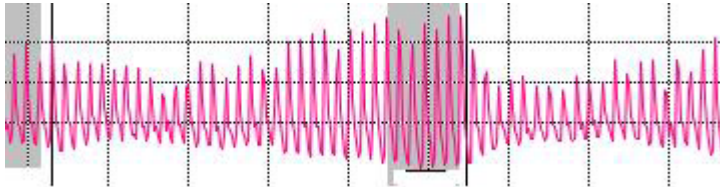


Figure 12 below shows an example of increased duration of amplitude reduction.



## Conclusion

The Utah-CQT was created by psychologist/examiners and founded upon known and proven principles of psychology and psychophysiology. The reliability and validity of the Utah-CQT has been demonstrated in many peer-reviewed and published scientific studies (see the review in Raskin & Honts, 2002). A number of writings may be found in scientific journals and texts discussing the Utah-CQT (for example, Raskin & Honts, 2002; Bell et al. 1999; Handler, 2006). We hope we consolidated some of those writings into a basic description of how to properly administer and evaluate the examination.

Those scientists who created and refined the technique took great pains to thoroughly research and assess the reliability and validity of the examination. This included numerous field and analog studies conducted over three decades. The Utah Scoring System (Bell et al., 1999; Handler 2006) takes a somewhat conservative approach to assigning values. This ensures that scores are assigned to reactions that are clearly different in comparison and not arbitrarily assigned. Some argue that this conservative approach may result in an inconclusive finding after three charts and thus require that the additional two charts be conducted. From a scientific standpoint, more data is better and the additional two charts should serve to increase

confidence in the results. It is the sincere hope of the authors that others in the field of polygraphy will consider learning and using the Utah-CQT. The more we move our profession toward techniques that employ scientifically validated principles, the more respect we will gain from others outside of the polygraph profession. As in any scientific field, progress can be made through the refinement of proven techniques already in place.

#### Acknowledgments

An earlier version of this paper was previously published by the American Polygraph Association in the journal, *Polygraph* 2006, 35, (3) and is reprinted here with permission of the editor, Tim Weber.

The authors thank David Raskin, Charles Honts, Don Krapohl, John Kircher and Frank Horvath for their thoughtful reviews and comments to an earlier drafts of and revisions to this paper. The authors would grant unlimited use and duplication rights to any polygraph school accredited by the American Polygraph Association or the American Association of Police Polygraphists. Questions and comments are welcome at [polygraphmark@gmail.com](mailto:polygraphmark@gmail.com) or [raymondnelson@gmail.com](mailto:raymondnelson@gmail.com).

#### References

- Abrams, S. (1999), *The role of polygraphy with the professional board*, *Polygraph*, 28, 300–303.
- Amsel, T. T. (1999), *Exclusive or nonexclusive comparison questions: A comparative field study*, *Polygraph*, 28, 273–283.
- ASTM (2002), E2229-02 Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data, ASTM International.
- ASTM (2005), E2324-04 Standard Guide for PDD Paired Testing, ASTM International.
- Barland, G. H. (1981), *A validation and reliability study of counterintelligence screening test*, Unpublished manuscript, Security Support Battalion, 902nd Military Intelligence Group, Fort Meade, Maryland.
- Barland, G. H. (1985), *A method of estimating the accuracy of individual control question polygraph tests*, Proceedings of IDENTIA, 1985, 142–147.

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

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

Department of Defense (2006), *Federal Psychophysiological Detection of Deception Examiner Handbook*, Retrieved on 03/13/2008, from: <http://antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf>.

DoDPI (2006), *Air Force Modified General Question Technique*, Presentation at the Senior Polygraph Examiner Course, Austin, TX.

DoDPI Research Division Staff (1998), *Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage*, Polygraph, 27, 68– 73.

DoDPI Research Division Staff (1997), *A Comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats*, Polygraph, 26, 79–106.

Handler, M. (2006), *Utah Probable Lie Comparison Test*, Polygraph, 35, 3, 139–148.

Handler, M. and Honts, C. R. (2007), *Psychophysiological Mechanisms in Deception Detection: A Theoretical Overview*, Polygraph, 36, 4, 221– 232.

Hardy, W. L., Murphy, V. T. (1996), *Victims and Polygraph Examinations*, Polygraph, 25, 4.

Harris, J., Horner, A., McQuarrie, A. (2000), *An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations*, Prepared under contract DABT02-96-C-0012 – CLIN0004-Phase 11 for The Department of Defense Polygraph Institute.

Hilliard, D. L. (1979), *Across analysis between relevant questions and a generalized intent to answer truthfully*, Polygraph, 8, 73–77.

Honts, C. R. (1996), *Criterion development and validity of the control question test in field application*, The Journal of General Psychology, 123, 309–324.

Honts, C. R. (1999), *The discussion of comparison questions between list repetitions (charts) is associated with increased test accuracy*, Polygraph, 28, 117–123.

Honts, C. R., Amato, S., Gordon, A. (2004), *Effects of outside issues on the Control Question Technique*, Journal of General Psychology, 151, 53–74.

Honts, C. R., Kircher, J. C., Raskin, D. C. (1988), *Patterns of activation and deception*, Psychophysiology, 25, 455.

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

Honts, C. R., Raskin, D. C., Kircher, J. C. (1992), *Effectiveness of control questions answered "Yes": Dispelling a polygraph myth*, Forensic Reports, 5, 265–272.

Horowitz, S. W., Kircher, J. C., Honts, C. R., Raskin, D. C. (1997), *The role of comparison questions in physiological detection of deception*, Psychophysiology, 34, 108–115.

Horvath, F. S. (1988), *The utility of control questions and the effects of two control question types in field polygraph techniques*, Journal of Police Science and Administration, 16, 198–209.

Kircher, J. C., Packard, T., Bell, B. G., Bernhardt, P. C. (2001), *Effects of prior demonstration of polygraph accuracies on outcomes of probable-lie and directed-lie polygraph tests*, DoDPI02-R-0002.

Kircher, J. C., Kristjansson, S. D., Gardner, M. G., Webb, A. (2005), *Human and Computer decision-making in psychophysiological detection of deception*, Submitted to the Department of Defense Polygraph Institute.

Krapohl, D. (2002), *Short Report: An Update for the Objective Scoring System*, Polygraph, 31, 298–302.

Krapohl, D. (2006), *Validated Polygraph Techniques*, Polygraph, 35, 3, 149–155.



Krapohl, D., Handler, M. (2006), *New Ratios for Scoring Electrodermal Responses*, *Polygraph*, 35, 2, 65–70.

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

Krapohl, D., Stern, B., Ryan, A. (2003), *Exclusionary vs. nonexclusionary: A review of the evidence*, *Polygraph*, 32, 4, 245–250.

Matte, J. (2000), *A critical analysis of Hont's study: The discussion (stimulation) of Comparison Questions*, *Polygraph*, 29, 2, 146–149.

Nelson, R., Handler, M., Krapohl, D. (2007), *Development and Validation of the Objective Scoring System*, version 3.

Offe, H., Offe, S. (2007), *The comparison question test: does it work and if so how?*, *Law and human behavior*, 31, 291–303.

Palmatier, J. J. (1991), *Analysis of two variations of control question polygraph testing utilizing exclusive and nonexclusive controls*, *Masters Abstracts*, 30, 380.

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

Raskin, D. C., Honts, C. R. (2002), *The comparison question test*, in: M. Kleiner (ed.), *Handbook of polygraph testing*, London, 1–49.

Raskin, D. C., Kircher, J. C. (1990), *Development of a Computerized Polygraph System and Physiological Measures for Detection of Deception and Countermeasures: A Pilot Study. A preliminary report under contract 88-L655300-000*, Scientific Assessment Technologies, Inc. Salt Lake City, Utah.

Raskin, D. C., Kircher, J. C., Honts, C. R., Horowitz, S. W. (1988), *A Study of Validity of Polygraph Examinations in Criminal Investigation*, Grant number 85-IJ-CX-0040, Salt Lake City, Department of Psychology, University of Utah.

Reed, S. (1994), Briefing on the Test of Espionage and Sabotage (TES), Expansion and Validation Study, Retrieved on 03/13/2008, from <http://antipolygraph.org/documents/tesexpansion-validation.shtml>.

Senter, S., Dollins, A., Krapohl, D. (2004), *A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute*, Polygraph, 33, 214–222.

Swinford, J. (1999), *Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute*, Polygraph, 28, 1, 10–27.

Timm, H. W. (1982), *Analyzing deception from respiration patterns*, Journal of Police Science and Administration, 10, 1, 47–51.

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

## Abstract

In 1970, a psychologist named Dr. David Raskin, a researcher at the University of Utah, began a study of the probable lie comparison question polygraph technique. Raskin and his colleagues systematically studied and refined the elements of polygraphy by determining what aspects of the technique could be scientifically proven to increase validity and reliability (Raskin & Honts 2002). Their efforts culminated in the creation of what is known today as the Utah approach to the Comparison Question Test (CQT). The Utah-CQT is an empirically consistent and unified approach to polygraphy. The Utah-CQT, traditionally employed as a single issue Zone Comparison Test (ZCT), is amenable to other uses as a multi-facet or multiple-issue (mixed-issue) General Question Technique (GQT) and the related family of Modified General Question Technique (MGQT) examination formats. The Utah-CQT and the corresponding Utah Numerical Scoring System (Bell, Raskin, Honts & Kircher, 1999; Handler, 2006) resulted from over 30 years of scientific research and scientific peer-review. The resulting technique provides some of the highest rates of criterion accuracy and interrater reliability of any polygraph examination protocol (Senter, Dollins & Krapohl, 2004; Krapohl, 2006). The authors discuss the Utah-CQT using the Probable Lie Test (PLT) as well as the lesser known Directed Lie Test (DLT) and review some of the possible benefits offered by each method.





**Frank Horvath\***

Michigan State University

**J. P. Blair\*\***

Texas State University

**Joseph P. Buckley\*\*\***

John E. Reid and Associates Inc.  
USA

# The Behavioral Analysis Interview: Clarifying the Practice, Theory and Understanding of its Use and Effectiveness

## Introduction

Everyone lies to the police (Ericson, 1981; Innes, 2003; Simon, 1991). Guilty suspects lie for obvious reasons. Witnesses lie out of fear of retribution; sometimes to cover for perceived personal failures. Victims lie to conceal their true involvement in matters that precipitated a crime. Even innocent

---

\* horvathf@daca.mil

\*\* pete.blair@txstate.edu or jpblairl@yahoo.com

\*\*\* info@reid.com

suspects lie, often because they perceive no need to cooperate with the police. And, of course, persons in all of these roles also tell the 'truth'. What and whom the police find credible are, at the core, the essence of the questions addressed in the police investigative task (Ericson, 1981; Horvath & Meesig, 1996; Innes, 2003; Simon, 1991). This is a dilemma. How do the police sort the truth-tellers from the liars; the credible story from the spurious? From the police point of view, this is a learned, on-the-job task, one in which they, if observant, ambitious and skilful, believe they eventually will become adept. From the perspective of a trained scientific observer, however, discerning between a 'lie' and a non-lie, even for those whose professional interest focuses on the problem, cannot be reduced to a simple, formulaic task. In fact, it is, according to some reports, as difficult for most 'professionals' as for others to determine when a lie is told (Bond & DePaulo, 2006; Ekman, O'Sullivan, & Frank, 1999). There is, however, some evidence that shows that, when 'high stakes' lies are assessed, police officers experienced in interviewing are able to detect truths and lies above chance levels, with an accuracy rate of about 65 per cent (Mann, Vrij, & Bull, 2004). Unfortunately, that study did not compare police officers with non-professionals; consequently, it did not show that 'professional' persons are better at lie detection than others. The accuracy statistic, however, suggests the validity of what seems to be truism, in the scientific literature at least, that detecting lies is quite difficult and not done, even in the best of circumstances, with a high degree of accuracy (Bond & DePaulo, 2006; DePaulo et al., 2003).

## The BAI

One approach that police officers can use to help them distinguish between those who are 'guilty' and those who are 'innocent' is called the Behavioral Analysis Interview (BAI). In fact, this is the only questioning method that has been developed specifically for this purpose. In its typical application the BAI is an interview process that the police use prior to a formal, accusatory interrogation. The purpose of the BAI is to help investigators sort those who are likely to be guilty from those who are not and thus to interrogate only those in the former category.

According to an article by Blair and Kooi (2004), training in the use of the BAI has been actively promoted as a part of a training protocol in the application of the Reid Technique, an approach to interrogation that may be the most well-known method in the world. Blair and Kooi maintain that over

150,000 police personnel have been trained in the use of the BAI throughout the world. But, they also point out that the basis for the BAI has not been solidly established in the scientific literature. This suggests that the BAI may lead the police to interrogate (using accusatory questioning) a truly innocent person and to fail to take appropriate action with respect to an actually guilty suspect. There is, therefore, a need for both practitioners and academics to engage in further research on the BAI and the Reid Technique in order that the police may rely on these processes with appropriate confidence.

The call for additional research on the BAI by Blair and Kooi (2004) would seem to be addressed in a recent study by Vrij, Mann, and Fisher (2006). They reported what is the first experimental, laboratory-based evaluation of the BAI. However, their report, coupled with other information found in the literature, demonstrated significant misunderstandings of the development, structure and research regarding the BAI. This article is intended to present an overview of the BAI and the research which supports it so that practitioners and researchers will not be misled by findings that have little, if any, ecological validity. In this paper a brief overview of the BAI process is offered first. Then the strengths and limitations of the current body of BAI research are discussed. Finally, suggestions are made about how to improve research on the BAI in the future.

Before getting to the heart of this paper, the key components and assumptions of the BAI will be described. More detail can be found in Blair (1999), Horvath (1973), Horvath and Jayne (1990), Horvath, Jayne, and Buckley (1994) and Inbau, Reid, Buckley, and Jayne (2001).

## Overview of the BAI

### General issues

A few special observations are in order at the outset. First, it is to be noted that it is common to find that the term 'guilt' is used synonymously with 'deception' and 'lying' in some contexts; whereas truthfulness is used to indicate 'innocence', 'honesty', 'truth-telling' and so forth. The difference between these terms is not always clear, nor is it evident from the context in which they are used precisely what is being referred to. For example, Vrij et al. (2006) stated that the BAI 'is designed to evoke behavioral responses to detect lying...' (p. 330). However, the BAI is not used to detect 'lying', at least not in the sense that there is an attempt to determine if a particular statement or message from a 'sender' is or is not true (as is commonly done in 'deception' research).

Rather, the BAI is used to detect 'liars', persons who deliberately withhold critical information regarding their knowledge of or involvement in a matter under investigation. (Maybe 'withholders' is a better term.) During the BAI a liar may, and usually does, tell a 'lie'. For example, she may say in response to a direct question, 'No, I did not steal that \$500' when, in fact, she did. But the detection of that 'lie' is not critical to the outcome of the BAI. It is only the completed BAI that leads to the inference that the 'liar' is indeed withholding relevant information and thus warrants additional investigative attention. In short, the BAI is perhaps best seen in the light of a 'guilty person' as opposed to a 'deception detection' paradigm as these terms have been used customarily in the 'lie detection' literature.

A properly conducted BAI leads to a decision of either 'Eliminated from suspicion', or 'Not eliminated from suspicion'. Outcomes in the latter category might result in follow-up questioning or the use of additional targeted, investigative resources. For this reason the terms 'guilty' and 'innocent' are used here to refer, respectively, to those who are not and those who are eliminated from suspicion.

A second point to be made clear is that the BAI is not exclusively a 'police' questioning approach, contrary to what some have suggested (Vrij et al., 2006). Although there is certainly widespread use of the BAI in policing, it is also the case that the BAI is used in many other environments. The major and most realistic assessment of the BAI carried out to date, by Horvath et al. (1994), did not involve police officers, police interviewers or 'suspects' who were in police custody. All interviewers were private employees of John E. Reid and Associates and all 'suspects' were employees of commercial firms suspected of involvement in independent specific thefts of money or property. All questioning took place on the premises of a private organisation without any connection to a police agency. The effectiveness of the BAI in a police environment has never been formally studied.

## Rationale of the BAI

It is believed that it is the 'guilty' person's underlying consciousness of involvement in the offence under investigation that is the foundation for the differential attitudes and behaviors of 'innocent' and 'guilty' persons during the BAI. Perhaps the use of a thought experiment will illustrate this. Imagine that you are at work as a bank teller and one day you are told that money is missing. You know you did not take the money. What are the first thoughts that enter your mind? It is likely that they will be 'I wonder what happened?'



Your initial thought might be that the money was misplaced. You do not want to believe that someone, a colleague, stole it. As your self-inquiry proceeds, however, and you learn that the money was not simply misplaced, you turn to 'what happened' and 'whodunit'. You replay the events of the day in your head. You also consider that since you are certain you are not responsible, someone you work with might be. These considerations will be based not only on your observations of co-workers' activities (i.e. opportunity and access), but also your personal knowledge of their personality and the circumstances of their lives (i.e. propensity and motive).

This response to such an event, the 'Sherlock Holmes Effect', is quite common. When an incident such as this occurs there is a tendency for involved persons to try to solve it. Observations over time show that innocent persons, though initially reluctant, are willing to discuss such thoughts (Horvath, 1973; Reid & Arther, 1953; Reid & Inbau, 1977). With prompting, they often reveal their suspicions of those whom they believe are most likely to be guilty.

Guilty persons on the other hand respond differently, and for good reason. They cannot share as much relevant information because that would ultimately lead to their detection, something they want to complicate, not simplify. Additionally, guilty persons do not experience the Sherlock Holmes Effect because they already know 'whodunit' and how it was done. Instead, guilty persons are consumed with thoughts of how to conceal their crime and how to misrepresent information that implicates them.

In the context of the established deception theories, the BAI can be viewed as a special case of the self-presentation theory presented by DePaulo et al. (2003). Both the innocent and the guilty try to craft a self-presentation that will make them appear as an innocent person. However, the guilty person's presentation will usually be less compelling, partly because of misperceptions regarding how truly innocent persons actually behave, and partly because the guilty person cannot share as much information as the innocent person, for to do so would lead to detection. The inability of the guilty person to share information typically leads to verbal responses to interviewer questions that are shorter and tend to indicate that the guilty person has not given a great deal of thought as to what happened and 'whodunit'; evidence of the Sherlock Holmes Effect is unlikely. The truthful persons responses tend to be longer, more detailed, and show that the person has given appropriate thought to the situation at hand.

In order for the Sherlock Holmes Effect to be observed, several assumptions seem to be necessary. The first is that some event of importance to the people who are involved must have occurred. This assumption is generally met whenever a criminal act occurs in a person's workplace or immediate

neighbourhood. The second is that the involved persons must be aware that the event did occur or is likely to have happened. Again, this assumption is easily met. In the workplace, people talk about shortages and the staff often engage in extensive attempts to locate missing money or objects before they are considered stolen. The third assumption is that the involved persons must have time to consider what happened, how it happened and who might be responsible. This is generally not an issue in police or corporate investigations because investigative interviews generally take place hours or even days after the event occurs. This time provides an opportunity not only for the involved persons to think about the event themselves, but also for them to gather information about the event from other people. Fourth, the involved persons must have some knowledge about others who are involved and about the environment in which the event occurred. In a corporate environment, the employees know the other people in their immediate working environment and they typically have a general idea of what is going on in each others lives. They also know the procedures and the day-to-day occurrences at their workplace. In a criminal case, the residents of a neighborhood may or may not know their neighbors and the general events that are occurring in each other's lives. The extent of this knowledge will depend on how tightly knit the neighborhood is. The more closely knit the neighborhood the more likely this condition will be met. Finally, the innocent person must have an interest in helping the investigation to succeed. In most cases this is due to a desire either to see justice done or to be exonerated of suspicion. However, in some situations it is possible that the innocent person has no interest in helping the investigation. For example, if someone close to the innocent person is known (by the person) to have committed the crime in question, the helpfulness typically observed may not be present. If these five assumptions are not met, the Sherlock Holmes Effect would be less likely to be experienced and therefore, the utility of the BAI would be more limited.

Viewed in this context, it is again important to point out that the BAI is not a procedure designed to evaluate directly at any particular point whether a person is lying or telling the truth. Rather, the BAI is used to assess the likelihood of involvement in a matter under investigation, often partially revealed by an attitude toward the investigation and the interview at hand. Field experience and field-derived data show that those who are not involved are more likely to be helpful, cooperative and confident; they show honest concern about the investigation and reveal sincerity and spontaneity in their responses during the BAI (Horvath, 1973; Reid & Arther, 1953; Reid & Inbau, 1977). Those who are involved in the matter at issue show other behaviors; they tend to be guarded and defensive; they are generally less helpful and

sometimes apathetic. These characteristics are revealed throughout the entire BAI, not just during a response to a particular question or set of questions, and they influence the message content that is evaluated during the BAI. However, it is important to note that because the research on the BAI to date shows a greater tendency for 'guilty' persons to be judged as 'innocent' than for the opposite outcome, it is possible that the BAI, even in the best of circumstances, may not be effective with persons who are sufficiently skilled at 'deceiving' others (Horvath et al., 1994).

## Conducting the BAI

The BAI is typically conducted in a private setting where the interviewer sits directly in front of the suspect at a distance of 4.5 to 5 feet. The interviewer's demeanor is non-judgmental and non-accusatory throughout the interview; this is so even when the interviewer believes that the suspect may have lied to a particular question or at a specific point in the BAI. It is believed that if this non-judgmental and non-accusatory demeanor is not maintained, accurate assessments of guilt (deception) or innocence (truth) are less likely because the behavior of the interviewer will lead the interviewee to distrust or become resentful of the interviewer and perhaps to manifest misleading behavioral indicators (Inbau et al., 2001). In addition to this concern, misleading behaviors could also result from the tendency of some interviewees to 'mirror' certain features of an interviewer's behavior (Chartrand & Bargh, 1999). While those who practice the BAI are aware of this and other possibilities that might distort accurate observations, there is no research of their effect, if any, on BAI outcomes.

During the course of the interview, the interviewer takes careful written notes documenting the questions asked of the suspect, as well as the suspect's verbal responses and non-verbal behaviors.

Because of this, there are short periods of silence separating each question. Of course, if an interviewee were to decide not to respond to the interviewer's questions during the BAI, or to respond with only a 'no comment' stance, the effectiveness of an assessment of behaviors would be greatly reduced, perhaps rendered valueless.

The first several minutes of the interview are spent obtaining background information from the suspect. This information, of course, establishes personal demographic data of interest to the interviewer. In addition, however, the collection of such relatively neutral data permits the interviewer to evaluate and note the suspect's 'normative' behavior, in particular, eye contact, response

timing, spontaneity and general nervous tension. During the remainder of the interview, the suspect is asked two different categories of questions, generally in separate time periods. These are investigative and behavior-provoking questions. Investigative questions concern such things as the suspect's actions, opportunity, access, motivations and propensity to commit the crime. It is important to note that these questions are as important to the BAI process as are the behavior-provoking questions. For example, consider a case wherein a suspect is seen on video entering the area from which money was stolen. The suspect denies being in that area during the asking of the investigative questions. In such an event, it is highly unlikely that a suspect would be eliminated from suspicion based solely on his or her responses to the behavior-provoking questions. The normative, investigative and behavior-provoking questions, in other words, are used in a complementary way.

The behavior-provoking questions are specifically asked to elicit differential verbal and non-verbal behaviors from 'innocent' and 'guilty' suspects. Over the last 45 years many different behavior-provoking questions have been developed. Verbal responses typically given by 'innocent' and 'guilty' persons to some of these were originally documented by Horvath (1973); they have been further developed based on empirical observations and theoretical expectations (Jayne & Buckley, 1999).

Specific guidelines govern the interpretation of messages and behaviors elicited during a BAI (Inbau, Reid, & Buckley, 1986; Inbau et al., 2001; Jayne & Buckley, 1999). These are designed to protect against errors, especially false positive outcomes (reporting an innocent suspect as guilty). These guidelines include the following:

1. Evaluate deviations from the suspect's normal behavior. There are no unique behavioral indicators which are always associated with 'innocence' or 'guilt'. It is the suspects deviation from his or her 'normal' behavior which is significant. Such 'normal' behavior is reflected especially in response to the questions regarding personal history, which are generally less emotionally provocative for 'guilty' and 'innocent' persons. For the 'innocent' person the items in the investigative and behavior-provoking categories are less provocative than for the 'guilty' person. Therefore, it is the consistency within and between categories that is important.
2. Evaluate the non-verbal behaviors as they occur in response to a particular question and be attentive to the repetition of the behaviors, that is, their consistency across questions.
3. Evaluate verbal and non-verbal behaviors in conjunction with each other. Look for discrepancies between the two channels.

4. Consider underlying factors which could affect the validity of behavior analysis. Examples of these include the seriousness of the offence, what the suspect was told about the offence, the suspect's emotional stability, and the suspect's cultural and social environment.
5. Consider the suspect's behavior in conjunction with factual analysis. When there is a discrepancy between known facts and a suspect's behavioral indicators, the interviewer must be cautious in rendering a definite opinion of the suspect's involvement.

## Examples of behavior-provoking questions

There are many behavior-provoking questions that may be asked during a BAI investigating an employee theft. Several examples of these, each dealing with a different aspect of the array of questions included in a BAI, are provided here to illustrate how such questions are used and how responses are assessed. More complete descriptions of these and other questions are found in Horvath (1973), Horvath et al. (1994), Inbau et al. (1986), Inbau et al. (2001).

In these three examples, the employee, Andy, is suspected of stealing a deposit from a department store. The interviewer's questions are preceded by the letter 'I' and are followed in parentheses with the abbreviated word (in italics) used to denote the question; the suspect's responses are indicated by an 'S'. Two responses accompany each behavior-provoking question; the first response is typical of a suspect not involved in the matter ('innocent') while the second is more indicative of a 'guilty' suspect (see Horvath, 1973; Inbau et al., 1986). It should be noted that during an actual BAI if the suspect's verbal response or the non-verbal behaviors are ambiguous, the interviewer may ask a follow-up question in an attempt to clarify the suspect's status.

I: (*Purpose*) Andy, what is your understanding of the purpose of this interview with me today?

S: (Innocent) Well, last Sunday morning when the bookkeeping department was counting up the deposits they found that the \$3,200 deposit from the men's department was missing and I know that I put it in the safe. So the reason I am here is to prove that I didn't steal it.

S: (Guilty) Well, I guess they may have misplaced a deposit envelope and I'm just here to help them find out what might have happened to it.

I: (*Attitude*) How do you feel about being interviewed concerning this missing deposit?

S: (Innocent) Oh, I don't mind at all. I want to prove to them that I didn't steal it and hopefully through these interviews they will be able to catch the thief.

S: (Guilty) I don't feel one way or the other. It's something that I have to do to keep my job.

I: (*Bait*) Andy, something you may not be aware of is that most drop safes have a counting mechanism on the underside of the drop slot. Very simply the force of an envelope entering the safe causes the counter to advance in increments of one. Now if you in fact did put that envelope in the safe last Saturday the counter should read 11, because that's the total number of envelopes that should have been dropped. Now I don't know if this particular safe is equipped with that mechanism, but if it is, can you think of any reason why the counter would indicate 10 drops instead of 11?

S: (Innocent) If it does it's not from my envelope because I know for sure I put that envelope in the safe.

S: (Guilty) Gee ..., I don't know very much about mechanical things, maybe it got stuck or something.

## Non-verbal behaviors

Many different non-verbal behaviors are also evaluated during the BAI. While Inbau et al. (2001) often use specific behaviors as examples of a guilty or innocent response, it must be remembered that it is the consistency of and the change from a person's normative behavior, not the behavior itself, that is evaluated. Commonly assessed vocalic behaviors include rate of speech, response latency and length of response. Commonly assessed kinesic behaviors include smiles, head movements, posture changes, illustrators, adaptors, foot and leg movements and eye contact. It must also be remembered that these behaviors are evaluated in conjunction with the suspects verbal statements.

## Research on the BAI

### Field studies

Historically, the foundation for the BAI was the early report by Reid and Arther (1953) regarding the behavior of persons undergoing polygraph examinations. Their report provided the grounding for an empirical evaluation of polygraph subjects' behaviors by Horvath (1973). His research on the 'structured pre-test interview', the portion of a polygraph examination prior to the collection

of physiological data, triggered the development of what is known as the BAI, a non-instrumental interviewing method used to assess the likelihood of a person's involvement in a specific matter under investigation.

Following the early Horvath (1973) evaluation, a second, larger scale and more carefully controlled assessment of the BAI was carried out (Horvath et al., 1994). In this study, four evaluators, all highly trained and experienced in the analysis of behavioral information using the BAI, made independent judgments of the innocence or guilt of real-life suspects who were undergoing BAI interviews. These interviews had been audio-visually recorded and were judged by the evaluators in four conditions:

1. A 'Written' condition in which they had access only to transcribed responses to the protocol of behavior-provoking questions asked of the suspects.
2. An 'Out-of-context, without audio' condition in which evaluators observed a video-taped replay of the suspects' non-verbal behavior displayed when responding to each of the behavior-provoking interview questions; evaluators were unable to hear the suspects' verbal responses but were aware of the question to which the suspect was responding (eg. 'Purpose', 'Attitude', 'Bait', etc).
3. An 'Out-of-context, with audio' condition in which the suspect's verbal and non-verbal behaviors were observable.
4. An 'In-context' condition in which each suspect's responses to the protocol of interview questions was observable in a sequence similar to that in the original BAI.

In the first three conditions evaluators made decisions about the suspects' 'innocence' and 'guilt', expressed the degree of confidence in those decisions, and rated certain behavioral and attitudinal dimensions of the suspects as they independently observed 786 segments of written or AV-taped questions. In the In-context condition, evaluators made two separate decisions of 'guilt' and, for each, expressed their degree of confidence. The first decision was made by evaluators as they judged separately each suspect's response to each question asked of each suspect; the question protocol was presented in the same, fixed sequence for each suspect. The second decision, a 'composite' judgment, was made after each evaluator had viewed the entire protocol of questions asked of each suspect. These 'composite' judgments were rendered by evaluators after they had observed each suspect in a context similar to the actual BAI from which the AV recording had been derived.

Evaluators' accuracy, calculated by combining the four evaluators' decisions of 'guilt' into a single-Group decision, showed that when inconclusive judgments were excluded, correct decisions on innocent suspects averaged 80 per cent, 79 per cent, and 88 per cent in the Written, No-audio, and With-au-

dio conditions respectively. Evaluators' In-context decisions were 88 per cent correct when they were made on observations of separate interview questions. On guilty suspects, evaluators' judgments, excluding that of 'inconclusive', were correct 53 per cent in the Written condition and 75 per cent, and 76 per cent in the No-audio and With-audio conditions, respectively. In the In-context condition when questions were scored separately, decisions were 64 per cent correct.

The different conditions of evaluation had a significant effect on evaluators' overall accuracy. On innocent and guilty suspects the lowest accuracy and the highest rate of inconclusive outcomes was obtained when only non-verbal behavior was observed (the Out-of-Context, without-audio condition). When evaluators were able to observe both verbal and non-verbal behavior in a context which closely resembled the real-life circumstance (In-context) and in which they rendered a 'composite' decision, they obtained the highest overall accuracy; excluding inconclusive judgments, they were correct in 86 per cent of their 'composite' judgments on innocent suspects and 83 per cent on guilty suspects. (Inconclusive judgments varied, of course, from evaluator to evaluator and across the different observational methods.

To summarize, however, the average percentage of inconclusive judgments in the In-context mode, composite judgments, on 'Truthful' persons was 14 per cent; on 'Deceptive' persons, 17 per cent.)

It is of interest to note here that analysis of objective scoring of the behaviors of the suspects in the Horvath et al. (1994) study, revealed significant differences between innocent and guilty suspects with respect to verbal and non-verbal behaviors.

In general, innocent suspects assumed different postures (uncrossed arms and legs, forward in the chair) and engaged in more frequent 'non-verbal' behaviors (movements of hands and postural changes) than guilty suspects; they also engaged in more smiles and head nods than guilty suspects. Innocent suspects were found to use more words per response, and were also more likely than guilty suspects to use words and phrases which were descriptive and reinforcing.

The findings in this study showed that in real-life decision-making both verbal and non-verbal behaviors observed during a BAI can be used in a complementary way. When these behaviors are observed in the context in which they occur they can be meaningfully assessed by persons who are trained and experienced in making these observations.

Additional research has indicated that training in some components of the BAI can enhance detection accuracy (Blair, 2007; Blair & McCamey, 2002). Both of these studies involved having participants view videotapes from the



Horvath et al. (1994) study, and both found accuracies that exceeded 70 per cent after training was conducted.

## Field study limitations

In all field deception or 'lie detection' research the ground-truth criterion problem is a difficult and complex methodological concern, and some have in large part dismissed the field research on the BAI because of this issue (Vrij et al., 2006). However, while the use of 'confessions' as the criterion for ground truth can be problematic in some circumstances, corroborated confessions have been and continue to be the 'gold standard' in 'lie detection' field research, especially that dealing with polygraphy. Perhaps one of the major reasons for this is that it is difficult to establish reasonably certain ground-truth criteria in field applications without at least partial reliance on corroborated confessions (Honts, 1996; Patrick & Iacono, 1991). Nevertheless, research in which direct comparisons between confession-based and other ground-truth criteria has been done has not revealed any effect on outcomes. For example, Horvath (1977) compared the 'accuracy' of blind and trained evaluators on two groups of polygraphic data, one confession-verified and the other unverified; accuracy did not vary as a function of verification. Similarly, Krapohl, Shull, and Ryan (2002) found no difference in outcomes when they compared the accuracy of polygraph results on a sample of guilty persons who confessed to their crime with that on guilty persons who did not confess but whose guilt was otherwise established. In this study, by the way, the polygraphic data were objectively scored and uncontaminated by the criteria that established ground truth. Moreover, it is to be pointed out that even in those cases in which DNA analysis has been used to demonstrate a wrongful conviction, a confession by the actual perpetrator often plays a role in buttressing the forensic testing.

In addition, it is exceedingly unlikely that in the Horvath et al. (1994) study any of the confession-based suspects were not actually 'guilty' or 'innocent' (when their 'innocence' was confirmed by the 'guilty' person in the same investigation). There are a number of reasons for this. First, false confessions are rare in 'real-life' cases. While some experimental findings may suggest otherwise, the best evidence, based on three different approaches to estimation of the rate of actual false confessions, shows that rate to be quite low. The first approach relies on United States national data to estimate the number of wrongful convictions produced by false confessions annually. This requires an estimate of the number of convictions annually, the error rate in those

convictions, and the proportion of errors due to false confessions. Using this method, two based studies estimated that the annual rate of false confessions ranged between 0.001 and 0.04 per cent for all FBI index crime convictions (Cassel, 1998; Huff, Rattner, & Sagarin, 1986). Some observers have rejected this method as being too error prone to produce accurate estimates (Kassin, 1997; Leo & Ofshe, 1998).

The second method of estimating the frequency of false confessions involves identification of such confessions in a random sampling of criminal cases. Cassell (1998) did such a search in 173 cases that were in the sampling frame of a previous study; he did not find any false confessions in that sample (Cassell & Hayman, 1996). Other studies, designed to examine the interrogation process, did not mention the presence of any false confessions in their sample of cases (Irving, 1980; Leo & Ofshe, 1998; Softley, 1980). It would be assumed that the presence of a false confession would have been noted, especially so since the authors of one of these studies have also done research on false confessions.

A third method involves surveying samples of court-convicted criminals who are asked to self-report false confessions. Using this approach, Gudjonsson and Sigurdsson (1994) surveyed 95 per cent of all inmates entering prison in Iceland in a one year period ( $n = 229$ ). They found that while 12 per cent claimed to have made a false confession in the past, none acknowledged a false confession to the current offence. In another study, Sigurdsson and Gudjonsson (1996) surveyed 509 prison inmates and 108 juvenile offenders in Iceland. None of the juveniles claimed to have made a false confession. Approximately 12 per cent of the adult inmates said they made a false confession to the police in the past; less than 1 per cent claimed to have falsely confessed to the current offence. Gudjonsson and Sigurdsson (1994) also noted that about 92 per cent of all Icelandic prison inmates confess to the crime for which they are convicted. Based on such information, Gudjonsson (2003) estimated that the rate of false confessions in Iceland is below 1 per cent, and approximately half of these are interrogation produced.

Finally, collections of false confession cases may also provide some insight here. If false confessions are common, a large number of cases should be available for analysis. Yet, the most comprehensive survey of such cases in the United States found only 125 proven cases of false confession in the last 30 years (Drizin & Leo, 2004). Because it is so difficult to establish ground truth in criminal cases, this number is probably an underestimate of the actual number of false confessions. Nonetheless, if the estimate is off by a factor of 1,000, the rate of interrogation-induced false confessions in criminal cases

in the US would still be less than 1 per cent. Moreover, it should be noted that some of the cases identified by Drizin and Leo as involving false confessions actually involved only admissions which might imply guilt, but were not confessions of guilt (Blair, 2005).

Each of the methods of estimating the frequency of false confessions has strengths and weaknesses. Taken together, however, the available data clearly show that false confessions are rare occurrences in actual criminal cases.

The second reason that it is unlikely that false confessions occurred in the Horvath et al. (1994) study is that all of the suspects were involved in private investigations, not police matters. The suspects were not in custody and had the right to leave at any time; therefore, the interrogators in the Horvath et al. study had significantly less power over the suspects than is common in police interrogations. Third, the confessions that occurred (all to thefts) led to resolution of the incidents wherein the 'innocent' were exonerated and the 'guilty' were dealt with as their employers thought appropriate. A false confession would have been obvious and immediately called to the attention of the investigative firm as the confessions were corroborated and found to be consistent with the facts of the case. Fourth, the confessions in these cases did not result from the kind of 'high pressure' interrogational tactics that have been demonstrated to cause false confessions in actual criminal cases (Blair, 2005). It is confessions in those situations, not private investigations of organizational thefts, which have generally led to the concern about false confessions. Finally, as in similar research, the confessions establishing ground truth had been corroborated by other evidence or circumstances; they did not stand alone.

Additionally, if it is assumed that most of the confessions were true, then a single false confession (or even several false confessions) would have served only to lower the power to find differences between the behaviors of the guilty and innocent. In other words, in the unlikely event that the group of 'confessors' contained persons who falsely confessed, say, in order to protect a coworker, it would be expected that the inclusion of such an innocent suspect (incorrectly identified as guilty) would reduce, not improve, the likelihood of finding statistically significant differences in behavior between the guilty and the innocent.

Factual analysis is a somewhat less used form of ground-truth verification. For this reason the process is described here in some detail. Factual analysis refers to the expert processing of the known information pertaining to an investigation (Inbau et al., 2001; Jayne & Buckley, 1999). In the Horvath, Jayne, and Buckley (1992) and Horvath et al. (1994) studies, factual analysis was carried out by two experts who independently assigned probabilities of 'guilt'

or 'innocence' in each of five separate categories for each suspect; these were: biographical, opportunity/access, behavior/attitudes, motivation/propensity and evidence. From the evaluators' separate evaluations a total probability of innocence or guilt was calculated for each suspect.

All 87 suspects' investigations were subjected to factual analysis. Of the 36 suspects whose ground truth was established by confession or other independent evidence, only one produced final factual analysis scores from both evaluators which were greater than 90 per cent and which were also inconsistent with ground truth. That is, when there was an agreement by both evaluators at levels of 90 per cent or higher, all but one of the confession/evidence verified suspects were correctly classified. Therefore, requiring at least a 90 per cent confidence level from both evaluators for including non-confession verified suspects appeared to provide a satisfactory criterion when ground truth could not otherwise be reasonably established. In this way, 24 suspects were included in the sample whose status was confirmed by factual analysis; 15 of these were 'innocent' and 9 were 'guilty'. In other words, the Horvath et al. (1994) sample of suspects included 36 persons whose 'guilt' or 'innocence' was confession-confirmed; 24 who were confirmed by factual analysis. Even though the criterion used to form the latter group is less certain as ground truth than is a confession, Horvath et al. (1994) reported: '...evaluators' scores on both the confession and the fact-analysis verified suspects were compared. This was done by separately subjecting each evaluators scores for the various assessments of suspects' behavior to a two-way ANOVA... These analyses did not reveal any significant effects for the Verification factor. Moreover, Chi-square tests showed no relationship between Verification and the frequency of correct, wrong or inconclusive judgments' (p. 801). In other words, the method of verification did not influence either the scoring of the data by the evaluators or the accuracy of the outcomes of their scorings.

Because false confessions are unlikely to occur in actual criminal cases and because corroboration of a confession with the facts of the case provides a safeguard against false confessions, confessions provide the strongest indicator of ground truth that is available in most field research. The use of factual analysis as an indicator of ground truth is clearly less certain and therefore less convincing. However, the similarity of the findings using the two types of ground-truth criteria in the Horvath et al. (1994) study suggests that factual analysis may be an effective substitute when confessions are not available. Additionally, multiple indicators of a construct are often used in research when, as was the case in the Horvath et al. study, they are internally reliable. This increases confidence in the outcome.

## Experimental laboratory studies

Two experimental, laboratory-based studies involving issues directly or indirectly relating to the BAI have been reported. The first of these, with only an indirect relationship to some aspects of the BAI, was done by Kassin and Fong (1999). In the first phase of this study, participants either committed a mock crime or engaged in another similar, but non-criminal, activity. The guilty participants were given an incentive not to confess to the crime. All of the participants were then interviewed (while being video tape recorded) by a civilian who pretended to be a detective. This interview began with the suspect signing a waiver of his or her Miranda rights (i.e. to legal representation, etc). Next, the interviewer explained to the participant how he or she had become a suspect, and asked the suspect to account for time when the crime was committed. Third, the interviewer refused to accept a suspect's denials, pounded his fist on the desk, and demanded the truth. Fourth, the interviewer attempted to break down the suspect's story, and finally, the interviewer asked the suspect to sign a confession. None of the suspects actually signed the confession. The interviews ranged from 3.5 to 6 minutes in duration.

During the second phase of the study, participants were randomly assigned to either a training or no-training group. Those in the training group watched two training videos developed and promoted by John E. Reid and Associates, Inc. These were video tapes designed to improve the ability to detect deception from observations of behavioral cues. Following the training, all of the participants, those in the trained and untrained groups, then assessed the tapes that were produced during the interviews with the suspects in phase one. Each person in each group judged whether or not the suspects in the videos were innocent or guilty. The reported results showed that the untrained evaluators were more accurate than the trained persons. Based on this finding the authors concluded that exposure to the Reid training tapes did not improve the accuracy of 'he detection' and may have actually reduced it.

A second laboratory-based study, one in which the BAI was directly evaluated, was conducted by Vrij et al. (2006). In this study, the subjects were randomly divided into a guilty group, who committed a mock crime, or an innocent group, who did not commit a crime. The innocent subjects completed a complex set of tasks which involved playing a game, 'Connect Four'. While playing the game, they observed several people entering and leaving the room.

Later, a person entered the room looking for a lost wallet, and the subject helped in a search for the wallet. The 'new' person then left the room, and after a brief time returned to claim that money was missing from the wallet.

The deceptive subjects in this experiment simply entered the room where the wallet was located, removed money from the wallet, and then, as instructed, concocted a story about playing 'Connect Four' using a detailed written account of the scenario provided by the experimenters. Persons in both groups were motivated by an offer of money if they were assessed as truthful regarding the theft of money from the wallet. Importantly, all of the participants were recruited for this experiment by being told that it was about telling a convincing story.

Following completion of their instructed actions, the participants were interviewed by a uniformed police officer. The interview began with some introductory questions and then the suspect had a chance to deliver an open-ended narrative about the event that was 'experienced'. Following that, the interviewer concluded the interview by asking the suspect an array of 15 behavior-provoking questions commonly included in a BAI (Horvath et al., 1994).

The 15 questions were then transcribed and assessed by coders. Each question was coded by two trained coders on two 5-point scales for the verbal assessments. The results pertaining to these assessments indicated that the answers given by the truthful participants to most of the BAI behavior-provoking questions were more likely to be coded as deceptive than the answers given by deceptive participants. In other words, the behavior-provoking questions in this study produced findings regarding message content opposite those said to occur in real-life BAI interviews.

Assessments of non-verbal behaviors in the Vrij et al. (2006) study were coded on only three of the 15 BAI questions. Vrij et al. reported that these results indicated that the coding of these behaviors was generally in the opposite direction to what has been reported by Inbau et al. (2001).

## Laboratory experiment limitations

While the primary limitation in field studies is the difficulty in establishing ground truth, the primary limitation of laboratory studies is external validity. That is, even though 'ground truth' is known with certainty in laboratory studies, the results of a given study may not accurately reflect what occurs in the real world. To limit this uncertainty it is necessary to design laboratory studies that closely mimic the key aspects of the real world phenomenon that is of interest. Both of the studies discussed above failed to do this.

The Kassin and Fong (1999) study, as already noted, did not directly assess the effectiveness of the BAI. It was instead an evaluation of the use of behavioral indicators in an 'interrogation', something which would only occur subsequent

to, not be a part of, an actual BAI in the field. However, because the behaviors that were assessed in this study were said to be those advocated by Reid and Associates, it is appropriate to consider these findings in the context of the BAI. As was discussed in the general overview of the BAI section, Inbau et al. (2001) explicitly state that an interview must be non-judgmental and non-accusatory if the behaviors of the interviewee are to be assessed accurately. Since the Kassin and Fong study was intentionally designed to include an accusatory component, it demonstrates a clear lack of understanding of the Reid BAI process and as such it is clearly unjustified to use either behavioral models that were designed as a part of the BAI or to generalize from those results to what happens in an actual BAI.

The external validity of the Vrij et al. (2006) study is also questionable. While the interviews were not accusatory, the scenario used by Vrij et al. did not resemble the typical real-life scenario in which the BAI is applied. First, the assumptions described earlier regarding the Sherlock Holmes Effect were not met. That is, the context of the Vrij et al. study was one in which all of the 'suspects' were involved in a situation in which they were isolated from other 'suspects'. This is very much unlike a real-life matter. In an actual case, say, for example, one in which money was stolen from a coworker's wallet, all of the suspects are in the work environment together; they would know each other; would have developed relationships and opinions about each other; would know who has access to what areas; would know who might have the need to commit such a theft and so forth. When they are asked questions such as: 'Who do you suspect?'; 'Who do you think would have had the best opportunity to commit this theft?'; 'Who would you vouch for and eliminate from suspicion?'; they would have a context to use to frame their answers. In the Vrij et al. study this was not true. The design did not provide enough of a context for the participants to provide meaningful answers to the behavior-provoking questions. Additionally, the participants were not given enough time to consider the situation (exactly how much time they were given is unclear as the Vrij et al. paper says only that the participants were given a few minutes). As was previously discussed, the suspect must have considerable time (at least a few hours) to consider what happened, who might have done it, and discuss this information with others if the Sherlock Holmes Effect is to be observed. It is also unclear whether or not a scenario in which the participants are aware that they are being assessed on their ability to 'tell a convincing story' can generate the concern on the part of the innocent persons necessary to produce the Sherlock Holmes Effect. This type of scenario merely creates a situation in which deception is a game with a cash prize. This is certainly not the situation in actual interviews of crimi-

nal suspects. It should also be noted that in the truthful condition there was nothing for the suspects to be suspicious about as no money was actually stolen from the wallet.

It is not clear what type of realistic situation Vrij et al. (2006) attempted to mimic; having guilty suspects make up a story about playing 'Connect Four' with another participant rather than actually involving themselves in an event similar to that in which the innocent participants engaged is not a common real-life experience. A design that more accurately reflected actual investigations could have been done. For example, having both the innocent and guilty suspects play 'Connect Four' while the guilty suspect attempted to steal the wallet would be a possibility. The Vrij et al. study also confounded any interpretation of the results. That is, having the theft of the wallet co-vary with the fabrication of a story does not permit 'guilt' for the incident to be sorted from the 'lying' about the story. Whether the results were as they were because of the stealing or because of the fabrication is not possible to know. The coding procedure utilized by Vrij et al. (2006) also represents a substantial misunderstanding of the BAI process. Vrij et al. incorrectly stated that non-verbal behavior is evaluated on only three of the behavior-provoking questions. In fact, non-verbal behavior is evaluated on all of the behavior-provoking and other questions that are asked during the BAI process (Inbau et al., 2001, pp. 126-127). Vrij et al. also separated the coding of non-verbal from the coding of the verbal process even though Inbau et al. clearly state that these behaviors must be considered in conjunction with each other (p. 126). Additionally, Vrij et al.'s representation of the way that non-verbal behaviors are evaluated during the BAI is inaccurate. While specific behaviors (such as adaptors) are sometimes used as examples, non-verbal behaviors are considered in the context of the BAI assessment criteria discussed previously. A behavior is not considered to be indicative of guilt or innocence until it has been assessed in the light of these criteria. In the Vrij et al. study all of the non-verbal behaviors that occurred during the three questions that were evaluated were coded in isolation from other criteria.

While the coders in the Vrij et al. (2006) study generally scored the verbal responses reliably, it is impossible to know whether their scorings would be consistent with what experienced BAI interviewers would assign. Information on the training of coders in Vrij et al. is sparse, and it is possible that the findings regarding the non-verbal behaviors would not generalize to a properly conducted BAI. It should also be noted that because only the behavior-provoking questions were analysed, the study was really an evaluation of only the behavior-provoking questions and not the BAI as it is commonly practised and described in the literature (Inbau et al., 2001).



## Conclusion

There is a clear need for additional evaluations of the BAI, both field- and laboratory-based. It is hoped that this paper provides a grounding for that research to move forward in two ways. First, the Sherlock Holmes Effect, its assumptions, and its relationship to the BAI process have been explicitly described. The reason this has not been done before is elementary (the bad pun is not intentional). The Inbau et al. (2001) manual was intended to be a training tool, written by practitioners for practitioners. As such there was little concern with or need to consider the underlying 'theory' and the associated assumptions. In reality the assumptions of the Sherlock Holmes Effect are almost always met in actual investigations; practitioners simply assume that this is understood. However, researchers do not necessarily make such assumptions, and the Sherlock Holmes Effect may not be given consideration in laboratory studies. It should, however, be seen as one of the critical design elements in any experiment that purports to test the BAI. Second, it is hoped that the overview of the BAI presented in this paper, and the detail given regarding the available field and laboratory assessments, will provide researchers with a better understanding of the BAI process and of the interrelationship of verbal and non-verbal indicators as they are synthesized by field practitioners. Directly dealing with these and related concerns will strengthen and, hopefully, encourage more laboratory-based studies in the future.

Well-designed laboratory studies can provide useful information. In the part of the 'lie detection' field dealing with instrumental methods of 'detecting deception', considerable attention has been devoted to the question of how best to simulate real-life conditions in the laboratory. It is generally recognized that in the latter environment it is difficult to replicate the motivation and the perceived consequences of the real world. In fact, it is precisely that difference between those two environments that has been the foundation for the controversy regarding how best to interpret empirical evidence in polygraphy (Faigman, Kaye, Saks, & Sanders, 2002; Lykken, 1998). Nevertheless, there are conditions which, if maintained in the laboratory, do appear to optimize detection and to be advantageous to replicating the real-life environment (Kircher, Horowitz, & Raskin, 1988). In addition, research has shown that there may in fact be only differences in degree, not kind, between good laboratory-based results and field data (Kircher, Raskin, Honts, & Horowitz, 1988; Pollina, Dolhns, Senter, Krapohl, & Ryan, 2004).

Given these findings in polygraphy it appears likely that continued research on widely used methods applied in the real world to distinguish between 'in-

nocent' and 'guilty persons', such as the BAI, will similarly reveal the optimal way to structure laboratory conditions to approximate those in the real world closely. This, of course, will enhance our understanding of the BAI and of the factors which influence its effectiveness.

It is especially important to produce convincing designs in this area that give specific attention to the interrelationship between gestures and speech, content and context, as is done in the real-life BAI. Such research should inform and perhaps alter the behavior of practitioners. For this to occur, however, practitioners must be convinced that the research is relevant to their practice. Given the available laboratory studies on the BAI, there is no reason at this time for them to hold this conviction. Further research, properly carried out and reported, will be welcomed, especially by practitioners who, on a daily basis, have to address the common but difficult task of 'lie detection'.

## References

- Blair, J. P. (1999), *Detecting deception: The effects of the Reid Behavioral Analysis Interview training*, Unpublished master's thesis, Western Illinois University.
- Blair, J. P. (2005), *A test of the unusual false confession perspective using cases of proven false confessions*, Criminal Law Bulletin, 41, 127–144.
- Blair, J. P. (2007), *Can detection of deception response bias be manipulated?*, Journal of Crime and Justice, 29, 141–152.
- Blair, J. P., Kooi, B. (2004), *The gap between training and research in the detection of deception*, International Journal of Police Science & Management, 6, 77–83.
- Blair, J. P., McCamey, W. P. (2002), *Detection of deception: An analysis of the behavioral analysis interview technique*, Law Enforcement Executive Forum, 2, 165–170.
- Bond, C. E., DePaulo, B. M. (2006), *Accuracy of deception judgments*, Personality and Social Psychology Review, 10, 214–234.

Cassell, P. G. (1998), *Protecting the innocent from false confessions and lost confessions and from Miranda*, Journal of Criminal Law and Criminology, 88, 497–556.

Cassell, P. G., Hayman, B. S. (1996), *Police interrogation in the 1990s: An empirical study of the effects of Miranda*, UCLA Law Review, 43, 839–931.

Chartrand, T. L., Bargh, J. A. (1999), *The chameleon effect: The perception-behavior link and social interaction*, Journal of Personality and Social Psychology, 76, 6, 893–910.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., Cooper, H. (2003), *Cues to deception*, Psychological Bulletin, 129, 74–118.

Drizin, S. A., Leo, R. A. (2004), *The problem of false confessions in the post-DNA world*, North Carolina Law Review, 82, 891–1007.

Ekman, P., O'Sullivan, M., Frank, M. (1999), *A few can catch a liar*, Psychological Science, 10, 263–266.

Ericson, R. (1981), *Making crime: A study of detective work*, Toronto, Butterworth and Company.

Faigman, D, Kaye, D., Saks, M., Sanders, J. (2002), *Modern scientific evidence: The law and science of expert testimony*, (2nd ed.), Eagan, MN, West.

Gudjonsson, G. H. (2003), *The psychology of interrogations and confessions: A handbook*, Hoboken, NJ, Wiley.

Gudjonsson, G. H., Sigurdsson, J. F. (1994), *How frequently do false confessions occur? An empirical study among prison inmates*, Psychology, Crime and Law, 1(1), 21–26.

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

Horvath, F.S. (1973), *Verbal and nonverbal cues to truth and deception during polygraph examinations*, Journal of Police Science and Administration, 1, 138–152.

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

Horvath, F.S., Jayne, B. P.(1990), *A pilot study to quantify behavioral indicators of deception in structured interviews of criminal suspects*, Report prepared for US Department of Defense, Contract M503-89-07.

Horvath, F. S., Jayne, B. P., Buckley, J. P.(1992), *Trained evaluator's judgments of behavioral characteristics of truthful and deceptive criminal suspects during structured interviews*, Report of grant #904–90-C-1164, US Department of Defense.

Horvath, F. S., Jayne, B. P., Buckley, J. P.(1994), *Differentiation of truthful and deceptive criminal suspects in behavioral analysis interviews*, Journal of Forensic Sciences, 39, 793–807.

Horvath, F. S., Meesig, R. (1996), *The criminal investigation process and the role of forensic evidence: A review of empirical findings*, Journal of Forensic Sciences, 41, 963–969.

Huff, C. R., Rattner, A., Sagarin, E. (1986), *Guilty until proven innocent: Wrongful conviction and public policy*, Crime and Delinquency, 32, 518–544.

Inbau, F. E., Reid, J. E., Buckley, J. P. (1986), *Criminal interrogation and confessions*, (3rd ed.), Baltimore, MD, Williams and Wilkins.

Inbau, F.E., Reid, J. E., Buckley, J. P., Jayne, B. C. (2001), *Criminal interrogation and confessions*, (4th ed.), Gaithersburg, MD, Aspen.

Innes, M. (2003), *Investigating murder: Detective work and the police response to criminal homicide*, Oxford, UK, Oxford University Press.

Irving, B. (1980), *Police interrogation: A case study of current practice*, Royal Commission on Criminal Procedure, UK, HMSO.

Jayne, B. P., Buckley, J. P.(1999), *The investigator anthology*, Chicago, IL, John E. Reid and Associates.

Kassin, S. M. (1997), *The psychology of confession evidence*, American Psychologist, 52, 221–233.

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

Kircher, J., Raskin, D., Honts, C., Horowitz, S. (1988), *Generalizability of mock crime laboratory studies of the control question polygraph technique*, Psychophysiology, 25, 462–463.

Krapohl, D. J., Shull, K. W., Ryan, A. (2002), *Does the confession criterion in case selection inflate polygraph accuracy estimates?*, Forensic Science Communications, 4, 3, 1–10.

Leo, R. A., Ofshe, R. J. (1998), *The consequences of false confessions: Deprivations of liberty and miscarriages of justice in the age of psychological interrogation*, Journal of Criminal Law and Criminology, 88, 429–496.

Lykken, D. (1998), *A tremor in the blood: Uses and abuses of lie detection*, New York, Plenum Trade.

Mann, S., Vrij, A., Bull, R. (2004), *Detecting true lies: Police officers' ability to detect suspects' lies*, Journal of Applied Psychology, 89, 137–149.

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

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., Ryan, A. H. (2004), *Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations*, Journal of Applied Psychology, 89, 1099–1105.

Reid, J. E., Arther, R. (1953), *Behavior symptoms of lie detector subjects*, Journal of Criminal Law, Criminology, and Police Science, 44, 106–112.

Reid, J. E., Inbau, F. E. (1977), *Truth and deception: The polygraph ("lie-detector") technique*, (2nd ed.), Baltimore, MD, Williams and Wilkins.

Sigurdsson, J., Gudjonsson, G. (1996), *Psychological characteristics of 'false confessors': A study among Icelandic prison inmates and juvenile offenders*, Personality and Individual Differences, 20, 321–329.

Simon, D. (1991), *Homicide: A year on the killing streets*, Boston, MA, Houghton Mifflin.

Softley, P. (1980), *Police interrogation: An observational study in four police stations*, Royal Commission on Criminal Procedure, UK, HMSO.

Vrij, A., Mann, S., Fisher, R. P. (2006), *An empirical test of the behavioral analysis interview*, *Law and Human Behavior*, 30, 329–345.

Copyright of International Journal of Police Science & Management is the property of Vathek Publishing Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download or email articles for individual use.

This article was published in: "International Journal of Police Science and Management", Vol. 10, No. 1, 2008, pp. 101–118, DOI: 10.1350/ijpsni.2008.10.1.68.

## Abstract

The Behavioral Analysis Interview (BAI) is the only questioning method that has been developed specifically to help investigators sort those who are likely to be 'guilty' from those who are not. In its typical application the BAI is a pre-interrogation interview that is used to focus interrogational effort; however, it also can be used independently in order to circumscribe investigative efforts in those cases in which there is a fixed and relatively large number of 'suspects'. In this paper an overview of the BAI process is provided and the findings and limitations of the extant bodies of field and laboratory research on the BAI are discussed. The paper concludes with suggestions to guide future research on the BAI.



Vitas Saldžiūnas, Aleksandras Kovalenko\*\*  
Ministry of the Interior  
LITHUANIA

## The Event Knowledge Test (EKT) in Polygraph Examination (in case murder)

In our last article we mentioned that in Lithuania the event knowledge test (EKT) is widely available (Saldziunas and Kovalenko 2008 ). Recently we made a psychophysiological test by polygraph which revealed the important circumstances of a crime, helping the police to investigate it.

In September 2007 in a Lithuanian city a 29 year-old car salesman disappeared in mysterious circumstances after meeting a friend. After a few days the missing businessman P's car was found. His disappearance met with a large response in the community.

Police looked over various versions and checked the evidence of witnesses and other participants in the criminal trial. They also conducted a detailed verification of this evidence and investigated objects or inspections. For instance there was rock oil and biological analysis. They analyzed businessman P's way of life. Police officers worked extremely hard, almost without days off. Finally they came to the conclusion that P. was dead, and accordingly

---

\* vitas.saldziunas@vad.lt; aleksandr.kovalenko@policija.lt

attempted to find the possible whereabouts of his body. To this end tracker dogs trained to detect dead bodies were introduced, as well as soldiers and a helicopter. After checking all possible versions of the crime officers concluded that only one man could have benefited from P's death. He was in debt with the businessman. They therefore checked the man's evidence: his whereabouts at the moment when the businessman went missing, other possible murder motives, intentions. His way of life was analyzed, his past investigated from school times to the present, and family members, the social and financial status of acquaintances, and possible connections all looked into.

Although during the investigation some facts led to suspicion of U's involvement in P's murder, because the body was not found, it was too early to bring judicial proceedings against him. The matter under investigation was at a standstill, with all possible methods and crime-solving measures exhausted. Detectives appealed to the VIP Security Department under the Ministry of the Interior asking to make a psychophysiological test with a polygraph on the suspect in businessman P's murder, citizen U. Intending to avoid any suspicions from police, friends and acquaintances, and hoping to 'cheat' the polygraph, U. agreed to be tested. For suspect U. a psychophysiological test was prepared over almost two weeks. Polygraph examiners examined all the materials of the investigation and familiarized themselves with the versions and ideas of officers who investigated the case.

According to the material of the investigation and information given by officers psychophysiological test (under EKT method) questions and answers were constructed.

**1. Where did businessman P. stay after his last meeting with you?**

- 0. K. village
- 1. S. city
- 2. P. village
- 3. L. village
- 4. D. village
- 5. A. village

**2. Do you know where P. is at the moment?**

- 0. On holiday
- 1. Hiding at a girlfriend's place
- 2. Gone abroad
- 3. Deceased
- 4. Went to buy a car
- 5. Went to his partner

**3. Do you know what time P. died?**

- 0. Before 2 pm



1. Before 3 pm
2. Before 4 pm
3. Before 5 pm
4. Before 6 pm
5. Before 7 pm

**4. Do you know what happened to P's wallet?**

0. Sold
1. Thrown in fields
2. P. has it
3. Given to a homeless person
4. Burned
5. At the businessman's house

**5. Do you know where P's body is hidden?**

0. In concrete
1. Fed to animals
2. Buried underground
3. Thrown in a pool
4. Loured
5. Quartered

**6. Do you know how P. was murdered?**

0. Suffocated
1. Poisoned
2. Shot
3. Beaten to death
4. Stabbed with a knife
5. Hanged

**7. Do you know where P's body is? (12 photographs are prepared taken near his house, all of them numbered.)**

**8. Do you know where P. was killed?**

0. By the lake
1. In a sport club
2. In the car
3. In the populated locality
4. In the forest
5. In the non-populated locality
6. In fields
7. By the river

**9. Do you know what type of weapon P. was shot by?**

0. An automatic rifle
1. A gun

2. A smoothbore

3. A pistol

4. A revolver

5. A crossbow

**10. Do you know what type of pistol P. was shot with?**

0. BERETA

1. WALTHER

2. MAKAROV

3. BAIKAL

4. CZ

5. TT

6. ASTRA

**11. Do you know how many times the trigger was pulled?**

0. One

1. Two

2. Three

3. Four

4. Five

5. Six

6. Seven

7. Eight

8. Nine

9. Ten

**12. Do you know what type of vehicle P's body was transported by?**

0. Bicycle

1. Car

2. Tractor

3. Truck

4. Trolley

5. Bus

Following the psychophysiological test some psychophysiological reactions to the answers were registered:

- Suspect U. left businessman P. in village P.
- U. thinks that P. is dead
- P. died between 4 and 5 pm
- P's wallet at the moment is near him
- P's body is buried, burned
- P's body might be in the place shown in photo no. 5
- P. was killed in the populated locality, forest, fields
- P's body was driven by tractor

- To the rest of the questions about possible shooting by a weapon, type of weapon and number of shots unambiguous psychophysiological reactions were not stated.

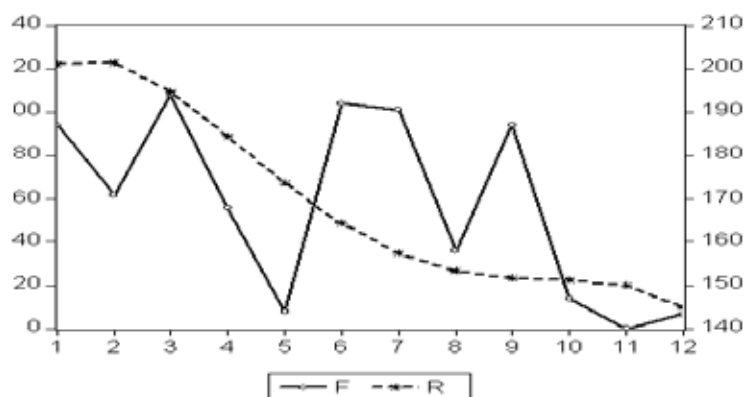
Though the psychophysiological test of suspect U. was made 4 months after businessman P's disappearance, it confirmed one of the police's versions of the disappearance. The results showed serious doubts about U's testimonies and version of his friend's disappearance. At the moment the circumstances of the murder have to be checked, as well as where the body is hidden. Unfortunately because of ambiguous psychophysiological reactions during the psychophysiological test the method of murder was not identified.

After conducting the psychophysiological test police officers investigating this case were instructed how the conclusion of this psychophysiological test could be used in making further investigative actions with suspect U. Using the conclusion of the test the suspect not only made a confession to murdering his friend P. but also confirmed all circumstances which were identified during the psychophysiological test. Later suspect U. showed the place where P's body was hidden.

All the answers to the questions from the described cases were searchable. Thanks to excellent police preparation it was possible to form good indirect questions. Before admeasuring the psychophysiological reactions to the answers the examiner read the questions with explanations one by one. Subject U. answered in each case that he did not know the answer. Accordingly it was suggested to him to say no to each version of the answer.

Thanks to the new computerized polygraph additional analysis was possible. Figure 1 shows how for a number of questions the subject's pulse rate and GSR tonic constituent fluctuated. Varlamov et al. (2001) write that GSR tonic constituent has a very slowly fluctuating skin resistance (potential) which depends on the metabolism in biological tissues. In figure 1 there are horizontally imaged numbers of questions, F – pulse rate (cycles per minute) and R – GSR tonic constituent (kilo ohm). These two values were measured after the question before the first answer was given.

Jaworski (2006) noticed some particular consistent patterns in the pulse rate fluctuation when measuring the pulse rate at the beginning and end of the sequence of questions. In the pulse rate F we did not notice any consistent pattern. As the number of questions was increasing GSR tonic constituent R gradually decreased. Value R does not depend on the question's social significance to the subject at all. The same tendencies were seen when investigating other subjects. It should be thought that the GSR tonic constituent R is decreasing because of adaptation processes.



According to Varlamov et al. (2007), if a person takes drugs R might increase to 400 kilo ohms. So it can be assumed that in the situation analyzed here the subject had not taken drugs. The special computerized program of the polygraph confirmed that during the psychophysiological test the examinee had not taken drugs and/or contra actions. Yet we did not have the opportunity to investigate subjects who use contra actions, and it is possible that in this case variation of the GSR tonic constituent can change by another consistent pattern. Only 12 questions were asked, as this investigation is not laboratory but outdoor. According to consistent pattern variation a solution is possible whereby with further questions the variation of R would decrease, and if this did not happen any changes in the investigation circumstances GSR tonic constituent would be steadied.

## References

- Jaworski R. (2006), *Situational sequencing tests in polygraph examination*, Wrocław, Wydawnictwo Uniwersytetu Wrocławskiego.
- Saldziunas V., Kovalenko A. (2008), *The event knowledge test*, European Polygraph, Vol. 1, No. 3.
- Varlamov V. A., Varlamov G. V. (2007), *Protivodejstviya poligrafu i puti ich nejtralizacii (Counteractions to polygraph testing and ways to neutralize them)*, Krasnodar, Kartika. [in Russian].
- Varlamov V. A., Korovin V. V., Varlamov G. V. (2001), *Testy poligrafnych prov-erok (PV Examination tests)*, Krasnodar, Prosvetschenije-Jug. [in Russian].

# Book review





Review of:

JE Sullivan,

*Gatekeeper: memoirs of a CIA polygraph examiner,*

Washington D.C., Potomac Books Inc., 2007, 273 pp.

Polygraphy – the use of a polygraph instrument to assess credibility, popularly known as “lie detection” – was developed along with many other forensic techniques that were widely applied in the United States early in the 20<sup>th</sup> century. While other techniques gradually became admissible evidence in judicial proceedings, polygraphy struggled for similar treatment. It continues to do so today. There are many reasons for this, not the least of which was and still is a real concern in the judiciary that a “lie detector” threatens the ostensible purpose of courtroom proceedings: a search for the “truth.” However, another factor is in play here also. Polygraphy for forensic purposes is one thing; when used in other environments it attracts controversy and, most importantly, political attention. It is this other application that Sullivan has written about in this book.

Ever since its original employment to protect the government’s atomic energy research programs in the late 1940’s, polygraph testing as a “screening” tool has been the source of controversy, scientific debate and vigorous challenge. At the federal level in the U.S. there have been periodic congressional reviews about every decade. Similar state-level reviews have also been carried out, though these have been confined to the last two decades or so. The most recent high-level review was the widely publicized report on “lie detection” by the National Research Council (NRC) of the National Academy

of Sciences (1). This was commissioned by the U.S. Department of Energy, largely in response to an FBI investigation in the Wen Ho Lee matter at the Los Alamos laboratory. When it released its report in 2003, the NRC did so with the plaint by at least one committee member that “The polygraph never caught a spy.” Sullivan, the author of this book, not only shows that this statement was not true but also says the committee members had been told before releasing their report of specific instances of detected spies. One member, however, was apparently determined to ignore “facts” in an effort to attract media attention. That issue notwithstanding, the NRC report clearly and forcefully brought to the forefront a legitimate concern about the use of “lie detection” to screen federal employees, contractors and those seeking employment. There is a real dearth of research on the topic. It is a curious fact that at least 80% of the polygraph testing done in or by federal agencies involves some sort of screening; at least 80% of the research studies done in or supported by federal agencies involves “lie detection” for other, usually forensic, purposes. Such, apparently, is the nature of government bureaucracy. In other words, we don’t have a lot of good, science-based knowledge about polygraph testing for screening purposes. If you’re a reader looking for technical substance on that topic then this book is not for you. However, if your interest lies in understanding why federal agencies are and continue to be the nation’s biggest consumers of the services of polygraph examiners, then Sullivan has quite a bit to tell you. And, if you’re especially interested in the “culture” of the small but influential polygraph examiner community in a very large intelligence agency, in this case, the CIA, Sullivan reveals a lot that heretofore has been largely sub-rosa.

It would serve no purpose to discuss this book by identifying the chapter titles, as they are not descriptive of commonly understood topics. They roughly set out in a somewhat chronological order, the author’s thirty-one years of work with the CIA, mostly as a polygraph examiner. During that time he spent “2,011 days overseas on agency business.” Much of this business involved conducting polygraph examinations in the most sensitive and protected areas of government concerns. Reading about these experiences reinforces many of the author’s points. In particular, he shows that polygraph testing is vital to our government processes. To carry out such testing, particularly in certain circumstances, is highly stressful in itself. Some persons excel at doing this; others have great difficulty. To complicate those problems with trying to deal with the vagaries of bureaucratic “politics” is, at times, debilitating for all. To add to all of these concerns, the author flavors his work with an oft stated understanding that the intelligence community operates in a world unique unto itself, sometimes black and white but often grey with vague boundaries and unspecified standards.



Spies and spying are, of course, central to the work of the CIA. Readers with an interest in knowing more about the real story behind some of the most damaging spies, e.g., Aldrich Ames (given an entire chapter), Robert Hansen, Harold Nicholson, as well as others who are less well known, will find it here. But they are the exception. The routine work environment of the author is at the heart of this book. Most of the material is personal and while it makes for an easy read, it is impossible to know what is unstated, what is being left out, not because the book was subject to pre-publication review (it was, but reportedly not much was redacted), but because the author simply wasn't privy to it.

One significant point to be made about this book, at least as it concerns polygraph testing, is that the author's view is based on experiences almost always in screening applications. It is likely that such limited exposure is the primary reason why he expresses the view that polygraph testing is "92% art and 8% science." His view does not square with the evidence; nor, is it consistent with the position of those who have taken the time to carry out, publish and digest the research studies which support contemporary polygraphy for forensic purposes. In the event that this point is not clear consider the differences. In forensic uses, there is a known event, usually investigative data, and a way, albeit limited, to verify outcomes. In screening applications the examination questions are event-free (e.g., "Did you ever give classified information to a foreign national?"), there are typically no available "facts" to link the examinee with an event, and it is difficult to determine the accuracy of the outcomes. This distinction is often not appreciated but it was a major feature in the review by the NRC and, though unstated, it provides the foundation for much of the author's narrative.

Another difference to be kept in mind is that in the intelligence field it is the fear of a false negative (i.e., a spy who evades detection, for example) that nags at examiners in their daily work. They, in their world, can cope with a few false positives but not with another Aldrich Ames, knowing in their heart of hearts that sooner or later it is going to happen. In forensic work, however, the opposite is the case. Avoiding errors on "truthful" persons, at the expense of some "deceptive" persons who go undetected, is not only the motivating force but it is also in line with the general philosophy of the judicial system. How an examiner weighs these two situations, one against the other, provides a backdrop for much of Sullivan's narrative.

The NRC drew two major conclusions from its work: First, with respect to screening applications, the accuracy of polygraph testing "in distinguishing actual or potential security violators from innocent test takers is insufficient to justify reliance on its use in employee security screening in federal agencies (p. 6)." Second, "Some potential alternatives to the polygraph show promise,

but none has yet been shown to outperform the polygraph. None shows any promise of supplanting the polygraph for screening purposes in the near term (p. 8).” This book shows the real-life tension between those two NRC conclusions. Policy in theory is one thing; policy in practice is another. The aphorism “That’s a fine idea in practice; but, it will never work in theory” is pertinent here. How Sullivan and his colleagues practiced their trade, with full awareness of this conflict is at the heart of his personal commentary. He shows that while all screening practices are imperfect and none is as effective as polygraph testing, the limitations in such testing are an ever-present source of daily-life stress of examiners in an agency such as the CIA.

One technical concern which Sullivan discusses that might be of interest to those who practice polygraphy is his focus on how the training and teachings of a person whose name is familiar to all, John E. Reid, dramatically changed for the better the examiner practices in the CIA program. “His ‘Reid test’ eventually replaced Keeler’s R/I test in the Agency,” and the “Reid School became the Harvard of polygraph training facilities. *Truth and Deception: The Polygraph (Lie-Detector) Technique*, the book Reid wrote in collaboration with Fred E. Inbau, is the polygraph examiner’s bible and the number-one reference source about polygraph” (p. 16). Although Sullivan himself was not trained in the Reid program he did attend some of their short courses. These experiences were beneficial. “I felt more confident interpreting charts, constructing tests, and interrogating subjects. Another benefit was that once I began applying what I had learned at Reid and saw that these methods worked, my faith in the polygraph process grew. IRD (The CIA’s Interrogation Research Division) was better because of the Reid training; IRD training was finally emerging from the Dark Ages” (p. 62). On the other hand, Sullivan also notes that this training may have had a down side. The increased confidence of the examiners led to greater certainty of testing outcomes; this, in turn, led to more aggressive post-test questioning of “deceptive” examinees which, given the nature of the workforce who were subjected to the testing, increased complaints about the polygraph testing process. Many of these, though apparently unfounded, nevertheless required some adaptive methods in the testing process. One can easily generalize from these CIA experiences to the external world. Such experiences feed the controversy about polygraphy. There is no doubt that the perceptions of this factor, whether factually founded or not, played a role, albeit not the consequential one, in the passage of anti-polygraph testing legislation in the United States (e.g., the Employee Polygraph Protection Act). Sullivan’s unstated point here is that the examiner community needs continually to be on-guard regarding how it deals with the use of polygraphy in screening environments.

A final point of interest in this volume is something that examiners know but outsiders may not recognize. Polygraph testing outcomes, considered as decisions of “truthfulness” or “deception,” in the screening context are subject to the “so what” effect. It is not the outcome, per se, that is of value; it’s the information that is developed that serves the consumer’s needs. A “lie detector” that merely serves objectively to determine if someone has “told the truth,” even if flawless, would not provide what Sullivan makes clear. The screening environment is structured so that an accurate personal history of the examinee can be constructed. Policy in practice is that there is more than “lie detection” at stake. “So what” if the screened person has not told the truth. What the “truth” is that is being withheld is more fundamental, more useful and indeed is the critical element in the process. It is certainly the unstated driver of screening applications of polygraph testing. Polygraph examiners who are able to develop information important to advancing the internal adjudication process in agencies such as the CIA are, to borrow from Sullivan, the Gatekeepers. It’s a squeaky gate though, and it needs attention.

Frank Horvath\*

## References

1. National Research Council, *The polygraph and lie detection*, Washington D.C., Committee to Review the Scientific Evidence on the Polygraph, Division of Behavioral and Social Sciences and Education, National Academies Press, 2003.

---

\* horvathf@daca.mil





Jolanta Antas:  
*O kłamstwie i kłamaniu (On lie and lying)*  
Universitas, Kraków, 2008, 348 pp. (in Polish)

It is Poland's first monographic work devoted to lie and lying. The author looks into the lie from the point of view of logic, semantics, language strategy, culture-related aspects, and signs and indications of lying from the point of view of verbal and non-verbal communication.

Until recently, one could believe that from the point of view of forensic sciences, only the indications of lying are material. While discussing them, the author makes use, however, of only a share of available literature. These are the generally known publications by Paul Ekman, David Lykken, and David Raskin. Works that are important, yet not the only ones. From the point of view of the traditional detection of deception, especially important is Chapter 7: *Znaki i oznaki kłamania. Z punktu widzenia komunikacji niewerbalnej (Signs and indications of lying. From the point of view of non-verbal communication)*.

The author believes polygraph examination to be unreliable, and following D. Lykken claims that "the successes of polygraph are achieved in fact only because experts convince the examined about the reliability of the device, and chiefly in this way achieve the intended results. For the examined believe in the efficiency of polygraph so much as to show all the symptoms of emotional agitation, which are actually characteristic for deception, yet which to the same extent may result from their panic fear that they will not be believed."

Further, the author claims that “the polygraph, which has been proven beyond doubt, is therefore in no way ‘a detector of deceit’, which means that it does not measure or prove directly the act of lying. What it does register directly are symptoms of emotional tension caused by the activity of the autonomous nervous system: physiological changes generated by this very emotional tension, which need not be specifically connected to lying.” (pp. 279–280).

All these are true and not true at the same time. If the author had reached at least to the generally-available manuals of forensic sciences, she would know that polygraph examination is a method of discovering emotional marks, which in turn, is a peculiar form of marks of memory that undergo a specific “processing” during the preparation to polygraph examination.

This “processing” means preparation of tests with questions formulated in such a way that the person examined may answer shortly “yes – no” or only “no”, and on preparing the examined person to answering such questions. Therefore, polygraphic examination in its classical understanding is not based on assessment whether any longer utterance by the examined in the form of a sentence in a logical sense is true or consciously false. In any form, polygraphic examination – be it the control questions techniques or the Guilty Knowledge Test technique – is a form of detection of deception. In the first case, deception means a knowingly false answer to a critical question, while in the latter – withholding the fact of possessing knowledge that the examined person is asked about.

The author lacks knowledge on the diagnostic value of polygraph examination as perceived this way, and her judgments in this field are ungrounded. Numerous experiments prove this value to be not lower than the diagnostic value of other methods of identification routinely applied in criminal investigation.

The remaining considerations of the author, the results of tests she refers to, and considerations of other authors are, on the other hand, of great interests for the new trends in lie detection research and practices.

The contemporary attempts at remote lie detection without the knowledge of the examined person must be based on indicators other than those used by the traditional polygraphic examination. Moreover, the utterances of the examined person are not stimulated by special lists of questions (tests), as is the case in a traditional detection of deception. Remote detection of lies requires the ascertainment and standardization of verbal and non-verbal symptoms accompanying longer utterances in the form of sentences (and not only short “yes – no” or “no” answers). The author presents four psychological states, described by M. Zuckerman and R. Driver, which the psychological theories of verbal deception associate with the act of lying:

**1) control** (trying to disguise what they really think, lying persons must control their behavior to a greater extent than persons telling the truth; which results in their being less spontaneous);

**2) general agitation** (the autonomic character of deceitful behaviors is related to the state of consciousness, described as “guilty knowledge”. It results in physiological agitation, as the lying persons remain in a permanent conflict between what they have in mind or memory and what they actually say. The indicators of this agitation, which accompany guilty knowledge, quoted by Antas after other authors (Scott, Wells, Wood, Morgan: *Pupillary response and sexual interest reexamined*, Journal of Clinical Psychology, 1967, 23, pp. 433–438, and: Simpson and Hale: *Pupillary changes during a decision making task*, Perceptual and Motor Skills, 1969, 29, pp. 495–498) include: pupil dilation, eye blinking, change of the basic voice frequency, linguistic errors, and language retardations. They can also be accompanied by pantomimic gestures;

**3) emotional states** (it is assumed that the action of lying is related to negative affections including a sense of guilt, involvement in the process of deceit, anxiety or fear of the lie being detected. They may be reflected in facial expressions, tone of the voice, and movements of the body;

**4) complexity of the cognitive processing** (as a conversation strategy, lying is a far more difficult mental process than telling the truth – true information). The lying person must build a non-existent image, and later communicate it, making sure that the message is coherent and non-contradictory, while the person telling the truth describes only a real image, which can be done spontaneously and without any stress. As Kahneman and Beatty (D. Kahneman: *Attention and effort*, Engelwood Cliffs, NJ: Prentice-Hall, 1973; J. Beatty: *Task-evoked pupillary responses, processing load, and the structure of processing resources*, Psychological Bulletin, 1982, 91, pp. 276–292) noticed, pupils contract in the phase of constructing lies, and expand in the phase of telling the lie.

Non-verbal reactions are accompanied by numerous verbal ones. Characteristic for the semantic structure of a deceptive utterance is the high frequency of phrases that reinforce/strengthen credibility, greater care for semantic and logical cohesion, lower concreteness, and increased brevity of the utterance (see: p. 302).

It seems that the symptoms listed above should be referred to the type of personality of the speaker whose truthfulness we want to assess. The minimum factors to be taken into account include extroversion/introversion of the speaker, and the level of the speaker’s emotional stability/lability. Probably general eloquence, being the derivative of a number of factors (extro-

version, IQ, the knowledge actually held, and language proficiency), would not be immaterial. Without these, the ascertainment of the verbal symptoms listed above tells nothing.

Nowhere in the book does the author refer to the dynamic changes in the brain that accompany lying, which can be ascertained through fMRI, and which have been subject of intensive research for over a decade, especially in the United States. This is certainly with a detriment to this precious and commendable book, as it deprives it of the attribute of completeness.

Jan Widacki\*

---

\* [biuro\\_poselskie\\_jwidackiego@interia.eu](mailto:biuro_poselskie_jwidackiego@interia.eu)









Jerzy Konieczny\*

## AAPP Thirty First Anniversary Polygraph Seminar

Jacksonville, Florida, May 19–23, 2008

Seminar report

American Association of Police Polygraphists (AAPP) is one of the two – next to the American Polygraph Association – great American organisations of specialists in the area of polygraph examinations. Its main aims include the development of cooperation among all American law enforcement organizations in the application and utilization of accepted polygraph techniques, development of highest standard of proficiency and promotion of highest standard of ethics, integrity, honour and conduct in the polygraph profession. One of the forms of activity of AAPP is the organisation of annual seminars; the 2008 seminar was held in Jacksonville, Fl.

The agenda was divided in two main sections. The first section was strictly connected with polygraph examinations, while the second one was devoted to the development of know-how and skills which are useful to polygraphists, although not directly necessary for the examinations themselves.

---

\* [jerkonieczny@wp.pl](mailto:jerkonieczny@wp.pl)

In the first section, the concept of two-tier polygraph examination certainly merits discussion; its key assumptions are outlined below. The starting point is the division of tests into diagnostic tests and screening tests. Diagnostic tests concern a specific event: the relevant questions focus on this very event, the number of suspects is relatively low, and decision accuracy in these tests is approximately 90%. Screening tests concern events yet undiscovered, the questions cover various aspects and various behaviour of the subject over a long period. Decision accuracy in screening tests is lower than in diagnostic tests.

The concept of the two-tier test requires that the procedure commences with screening tests. On subjects assessed as NDI (no deception indicated) no further tests are conducted, **but** since the screening tests are analysed using a method which minimises false negative errors, the NDI result is achieved only with respect to approximately 70% of subjects who are indeed truthful. These persons are “allowed through the net”, but the institution at whose requests the examinations are conducted must be warned that no decisions should be made with regard to these persons without additional information. Generally speaking, after the screening tests no DI (deception indicated) assessments are employed, but rather a dichotomy NDI/non-NDI. Diagnostic tests are only employed in the second stage and only then, if the interpretation suggests such a result, a DI assessment may be pronounced; in other cases the examination result is deemed to be IC (inconclusive). This procedure is recommended as good practice. Utah ZTC is recommended as the best diagnostic test by far (and the only one allowed for evidentiary purposes). Also allowed, but only of investigative purposes, are the CIT, Federal ZCT, Reid’s and RIT tests. It is worth noting that, while previously for all polygraph examinations 20% of IC results was considered the norm, the two-tier system of testing significantly lowers this number.

In terms of interpretation of the diagnostic tests, there has been a modification in recommendation with reference to the most popular seven-position numerical scale as far as assessment of GSR responses is concerned. Traditionally, the recommendation had been: where the proportion of intensity of responses is 2:1 between the relevant and comparison question or vice versa, the assessment was  $\pm 1$  (i.e.  $2:1 = \pm 1$ ), in cases of a 3:1 proportion  $= \pm 2$ , and  $4:1 = \pm 3$ . Currently, the recommendation is as follows:  $1.25:1 = \pm 1$ ;  $2:1 = \pm 2$ ;  $3:1 = \pm 3$ . This naturally results in no increase in decision accuracy, but changes the proportion of false negative/false positive errors, which is of fundamental importance for the two-tier procedure (as mentioned above, this is with regard to minimisation of false negative errors in the first stage of examinations).

The second important issue raised at the seminar was quality control in polygraph examination. The control covers **all** elements of the examination, and **each** polygraphist must take into account the possibility of a detailed control of **each** examination that he/she has conducted.

According to Elmer Criswell, AAPP–QC Director, the key principle of quality assurance is: The PDD examiner has excluded all factors other than lying as the cause of the physiological or deceptive responses. This is the driving principle of control. The chief elements that may be controlled, to mention only a few, include: the pre-test worksheet, consent form, Reid interview, used techniques, question sheet, questions reviewing with subject, test data, electronic disc (with charts and audio/video records), scoring (with rules followed), etc. A certain novelty is the necessity to conduct a Reid interview (not to be confused with the Reid technique) during the pre-test interview.

Issues of post-test interview were almost entirely missing from the seminar. However, the participants were informed that the Defence Academy for Credibility Assessment (former DoDPI) has contracted a consulting company (EASI Consult) to determine the optimal approach for conducting post-polygraph interrogation.

Presentations not directly related to polygraph testing covered interrogation methods, sexually motivated crimes and, last but not least, criminal profiling, which was of particular interest and value.

Producers of polygraph equipment had an opportunity for presentation too. Axciton, Lafayette and Stoelting were all present. Computerised equipment was mainly advertised. Lafayette displayed its model LX5000, promoted as the world's first wireless polygraph.

Overall, the seminar was held in good atmosphere, where discussion was to the point and presentations businesslike and factual. The location – Hyatt Regency – certainly contributed to the comfortable atmosphere, as did the majestic views over the nearby St. Johns River.



## The basic information for Authors

To publication will be accepts unpublished research papers as well as review article, case reports, book reviews and reports connected with polygraph examinations.

Submitted manuscripts must be written in English.

All papers are assessed by referees (usually from Editorial Board), and after a positive opinion are published.

Texts for publication should be submitted in the form of normalized printout (1800 characters per page) and in electronic form (diskette, CD), or sent by e-mail to Editorial Office.

The total length of research papers and review article should not exceed 12 pages, case reports – 6 pages, and other texts (book review, report) – 5 pages.

The first page of paper should contain: the title, the full name of the author (authors), the name of institution where the paper was written, the town and country.

Figures should be submitted both in printed form (laser print, the best) and electronic form.

Tables should be numbered in Roman numerals and figures in Arabic ones.

Figures, tables, titles of figures and titles of tables should be included on a separate page. The places in the text where they are to be included should be indicated.

The references should be arranged in the alphabetical order according to the surnames of the authors.

The references should be after the text.

Each reference should include: the surname (surnames) of the author (authors), the first letter of author's first name, the title of the book, year and place of the publication, the name of publisher, or the title of the paper, the full title of the journal, the year, the volume, the number and the first page of the paper.

For example (in references):

Reid J., Inbau F. (1966), *Truth and Deception: the Polygraph ("Lie-detector") Techniques*, Williams & Wilkins, Baltimore.

Abrams S. (1973), *Polygraph Validity and Reliability – a Review*, Journal of Forensic Sciences, 18, 4, 313.

and (Reid, Inbau, 1966), (Abrams, 1973) inside text.

Texts for publication in "European Polygraph" should be mail to:

"European Polygraph"

Andrzej Frycz Modrzewski Krakow University College

ul. Kanonicza 9

31-002 Kraków (Poland)

Or e-mail: [margерita.krasnowolska@kte.pl](mailto:margерita.krasnowolska@kte.pl)