Polygraph & Forensic Credibility Assessment:

A Journal of Science and Field Practice

VOLUME 47 **2018** **NUMBER 1**

Contents

Published Semi-annually

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice

Editor-in-Chief: Mark Handler
E-mail: Editor@polygraph.org
Managing Editor: Nayeli Hernandez
E-mail: polygraph.managing.editor@gmail.com

******

Subscription information: *Polygraph* is published semi-annually by the American Polygraph Association. Editorial Address is Editor@polygraph.org. Subscription rates for 2018: One year $150.00 (Domestic), $180.00 (Foreign). Change of address: APA National Office, P.O. Box 8037 Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN *POLYGRAPH* DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

Instructions to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript Submission

Manuscripts must be in English, and may be submitted, along with a cover letter, on electronic media (MS Word). The cover letter should include a telephone number, and e-mail address. All manuscripts will be subject to a formal peer-review. Authors may submit their manuscripts as an e-mail attachment with the cover letter included in the body of the e-mail to:

Editor@polygraph.org

As a condition of publication, authors agree that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere. Authors also agree to give reasonable access to their data to APA members upon written request.

Manuscript Organization and Style

All manuscripts must be complete, balanced, and accurate. Authors should follow guidelines in the *Publications Manual of the American Psychological Association*. The manual can be found in most public and university libraries, or it can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Writers may exercise some freedom of style, but they will be held to a standard of clarity, organization, and accuracy. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psychophysiological Detection of Deception* (2012) which is available from the national office of the American Polygraph Association. Legal case citations should follow the *West* system.

Manuscript Review

An Associate Editor will handle papers, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 4 weeks of receipt of manuscript. Articles submitted for publication are evaluated according to several criteria including significance of the contribution to the polygraph field, clarity, accuracy, and consistency.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be granted if the senior author approves.

Technique for Psychological Assessment of the Testimony Reliability in Legal Practice

Vitalii Shapovalov[1]

The technique for psychological assessment of the testimony reliability in the legal practice is a set of special procedures developed for the purposes of psychological assessment of verbal and nonverbal expressions of the person under investigation (the person in question) during an interrogation, a crime reenactment, or other investigative activities recorded on videotape.

The technique is based on Undeutsch's hypothesis (Undeutsch, 1967) asserting that statements derived from the memory of real-life experiences differ considerably from fabricated information. This difference is based on the assumption that the feigned information requires a greater cognitive effort, more creativity, and, certainly, self-control on the part of the subject. Fabrication of a lie takes more energy than telling what actually happened. This unavoidably affects the specifics of the narration and its content. An account of real-life experiences and an account of fictitious events are based on different *text-forming regularities*[2] (Shapovalov, 2015). Identifying these regularities offers a way to determine with a degree of probability whether the provided information is true or false. Every procedure of the technique is aimed at identifying specific text-forming regularities.

Application of this technique in an investigation is most appropriate in cases when the person under investigation admitted their guilt in the early stages of investigative activities only to recant their testimony later, claiming that he was subjected to undue psychological and/or physical pressure during the investigation, which compelled him to slander himself.

The comprehensive approach employed by this technique also makes it possible to detect the truth and lie combinations in the testimony. There is a paradox in that a truthful testimony can include elements of deceit, for example, when the person in question may attempt to diminish their role in the event being investigated, present their position in a socially desirable light, or conceal the real motives behind their actions or other significant participants of the event being investigated, etc. Likewise, completely fabricated testimony may be subtly interlaced with some actions or events that had happened in reality and can be corroborated by other persons, except that they had actually occurred at a different time and were deliberately included by the person in question to concoct a more believable lie. The set of procedures of this technique makes it possible to identify these regularities of truth and lies combinations of, thereby considerably minimizing the risk of error in findings.

Thus, the main objective of an investigation undertaken using this technique is to determine the psychological indications of *reliability*[3] or *unreliability*[4] (Shapovalov, 2015) in the testimony of the person under investigation. Based on a recorded set of psychological

1 Vitalii is a polygraph expert, psychologist, and the chairman of Ukrainian Polygraph Collegium in Kiev, Ukraine. He is the developer of this technique.

2 Psychological text-forming regularities in giving testimony are mechanisms of text generation from the perspective of the theory of text-forming, rules of speech communication, psycholinguistic regularities of communication, and principles by which reality is reflected in speech. These regularities are at the core of the structure, content, and expression of speech, as well as the mimic and pantomimic manifestations accompanying it (nonverbal communication).

indications, this makes it possible to conclude whether the person under investigation actually experienced the event recounted by him, or if the account of the event has been fully or partly fabricated.

The following questions can be answered using this technique:

1. Are psychological indications of reliability or unreliability present in the video recorded testimony of the person under investigation given during an interrogation, crime reenactment, or other investigative activities?
2. Has the person under investigation given testimony during an interrogation, crime reenactment, or other video recorded investigative activities independently or under somebody's influence?
3. Has the person under investigation been subjected to any undue psychological pressure during an interrogation, crime reenactment, or other video recorded investigative activities?

Summary of the Technique

At least two video recordings of different investigative activities, such as an interrogation and a crime reenactment, are needed to perform a psychological analysis and prepare findings. If the person under investigation was subjected to a video recorded polygraph examination, this video can also be included in the psychological analysis of the testimony.

A transcript is made using the available video recordings of investigative activities - a verbatim recording of the statements made by the person under investigation and by other participants of investigative activities. The statements are recorded word for word, showing the pauses and individual specifics of phraseology and pronunciation.

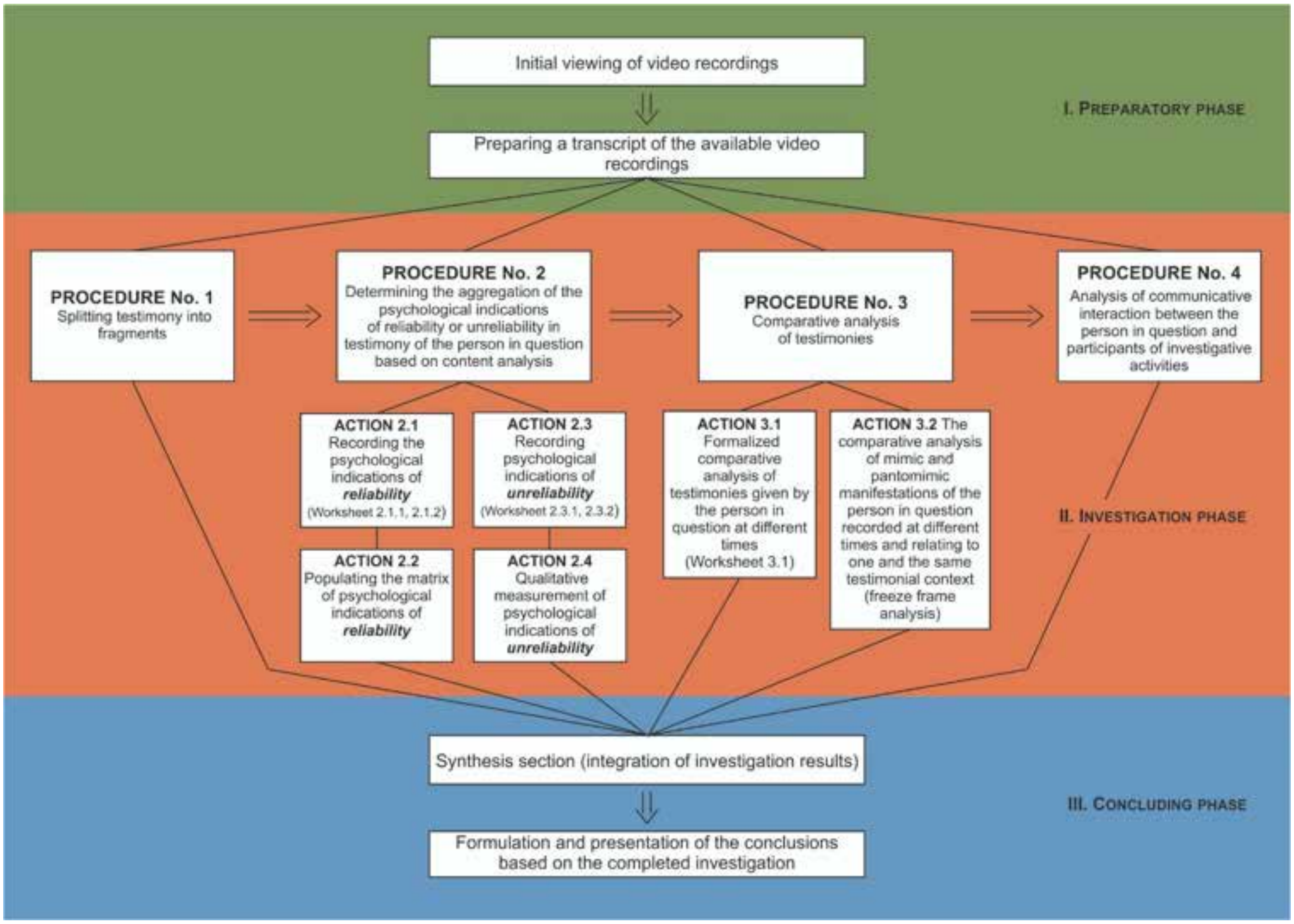
The expert also uses the video recording in addition to the speech of the person under investigation to add descriptions of emotional, mimic, and pantomimic manifestations to the transcript. They play a diagnostic role and can be interpreted on the basis of the aggregation of recorded manifestations.

A verbatim recording of the actual statements made by participants of investigative activities makes it possible to apply content analysis, elements of semantic analysis, comparative analysis, etc.

3 Psychological indications of **reliability** are units of speech that include perceptive information (visual information, sounds, smells, taste, and sense of touch), contextual information (details regarding the location and time of the event), information about actions / interactions and affective information (description of emotions experienced during the event), which in some cases can be accompanied by illustrator gestures and mimic/pantomimic manifestations. These indications reflect the different aspects of the objective reality. A total of 43 indications are used.

4 Psychological indications of **unreliability** are content and vocal units of speech, mimic-pantomimic manifestations that act as indicators of stress and an active thinking process, which complicates a free narrative in testimony. A total of 25 indications are used.

Fig. 1. Structure and Sequence of Procedures within the Technique of Psychological Assessment of the Testimony Reliability in Legal Practice



Procedure No. 1

Splitting testimony into fragments

The testimony is split into fragments in a process known as narrative structuring in order to optimize the process of analyzing testimony and identifying ***structural text-forming regularities***.

After studying the testimony, the expert splits it up into fragments that are distinctly different from one another and reflect the succession of the event and its social arrangement.

General logic of narrative structuring:

Fragment No. 1. Preamble of the event (introduction, etc.) (F. 1);

Fragment No. 2. Events and actions that happen right before the key event[5] (F. 2);

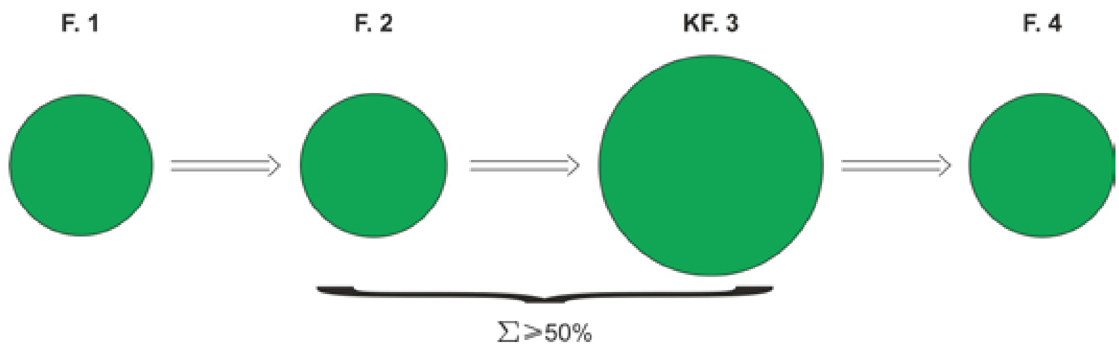
Fragment No. 3. Key event fragment (KF. 3);

Fragment No. 4. Actions that follow the key event (F. 4).

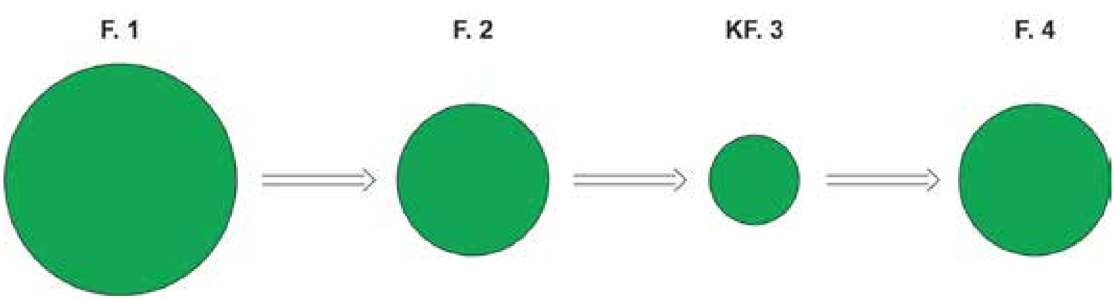
5 **Key event** is an act of criminal nature.

Structural regularities are regularities of unfolding of the text content and the logical sequence in the development of the temporal and social arrangement of the testimony, and also the correlation of the informative content of event fragments (Procedure No. 1).

a) ***manifestation in truthful testimony***: If testimony is given not immediately after the event but after a certain period of time, once the disorganizing influence of the affective state has diminished, such testimony normally shows a clear succession of the events in space and time, and individual parts of the narrative do not contradict each another. Another regularity is that in truthful testimony Fragment No.2 "Events and actions that immediately precede the key event" and Fragment No.3 "Key event" collectively have a greater number of characters (words) than any other individual fragment.



b) ***manifestation in fabricated testimony***: If the number of characters in Fragment No. 3 "Key event" is less than in any other fragment, this may indicate a concealment or distortion of testimony relating to the key fragment. Other possible indications of this are distortions in the temporal sequence of the narrative or logical contradictions among separate parts of the narrative. In some cases, Fragments 1,2,4 can be artificially exaggerated by a person, at the expense of accumulation of secondary details in order to create an illusion of cooperation, thereby hiding the key details and cause-effect relationship of the event.



This procedure makes it possible to determine the aggregation of the psychological indications of reliability or unreliability in each specific fragment, and perform a comparative analysis of the fragments of testimony recorded at different times.

Procedure No. 2

Determining the aggregation of the psychological indications of reliability or unreliability in testimony of the person in question based on content analysis

This procedure makes it possible to identify ***semantic, cognitive, and behavioral*** regularities and consists of four separate actions.

Action 2.1

Recording the psychological indications of reliability

To formalize this process, a list of psychological indications of reliability (43 attributes)[6] was prepared, and a procedure for recording them in a special worksheet (so-called "coding") developed (Worksheet 2.1.1).

Fig. 2. Sample worksheet for recording psychological indications of reliability (Worksheet 2.1.1)

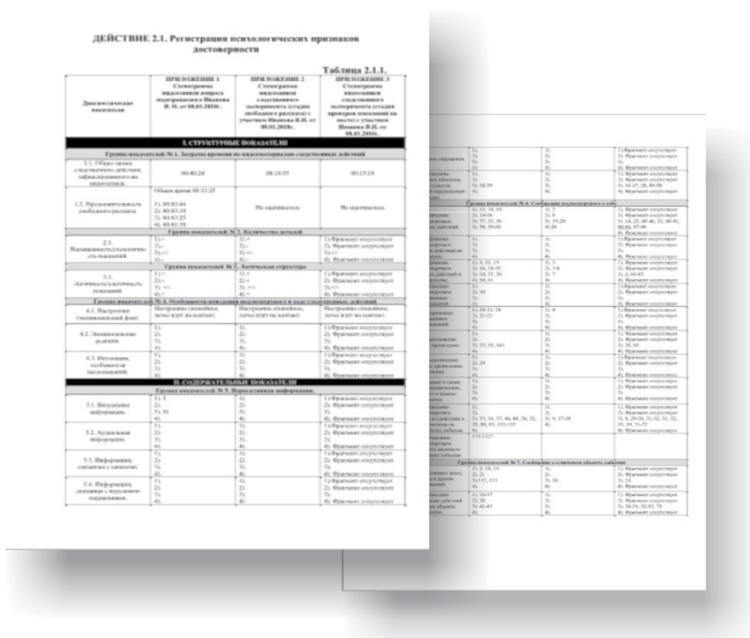

ДЕЙСТВИЕ 2.1. Регистрация психологических признаков достоверности

Таблица 2.1.1.

I. СТРУКТУРНЫЕ ПОКАЗАТЕЛИ

II. СОДЕРЖАТЕЛЬНЫЕ ПОКАЗАТЕЛИ

Psychological indications of reliability belong to ***<u>semantic text-forming regularities</u>***.

6 Some content criteria were borrowed from the techniques: Reality monitoring (Johnson M. K., Raye C. L. 1981), Scientific Content Analysis (Sapir A., 1987), Criteria-based statement analysis (Steller M., Köhnken G., 1989).

Semantic regularities are regularities of unfolding details during the narrative of an event that occurred. There regularities are reflected in Undeutsch's hypothesis asserting that statements derived from the memory of real-life experiences differ considerably from fabricated accounts (Procedure 2, Actions 2.1, 2.2).

a) <u>manifestation in truthful testimony:</u> a verbal description of real-life experience and the complex of sensations imprinted in the long-term memory as a result of the event that was experienced, and of a general understanding of the context of the situation, and its cause-and-effect relationships. Memories of the real-life experience are based on perceptive processes; therefore, they normally include perceptive information (visual information, sounds, smells, taste, and sensation), contextual information (details regarding the location and the time of the event), information about actions / interactions and affective information (description of emotions experienced during the event) (Undeutsch, 1967), (Johnson, Raye, 1981), (Sapir, 1987), (Steller, Köhnken, 1989). These memories are usually clear, colorful, and vivid, and are not normally difficult to verbalize. Such testimonies are rich in details that reflect the events and the surrounding physical world (the presence of the required set of psychological indications of reliability).

b) <u>manifestation in fabricated testimony:</u> sketchiness and scant details (this particularly applies to secondary details), inability to accurately describe the actions of other event participants, and to put them in a spatial and temporal context (the lack of the required set of psychological indications of reliability) (Undeutsch, 1967), (Johnson, Raye, 1981), (Sapir, 1987), (Steller, Köhnken, 1989).

There is an example of a completed worksheet 2.1.2 below where counting of psychological signs of reliability is performed that lets a specialist analyze their fragmentation. The quantity of psychological signs must be the highest in the key fragment (a criminal stage of a crime) in a true-life story.

Worksheet 2.1.2 The calculation of psychological signs of *reliability* in the investigation action and their key fragmentation.

Testimony Fragment	Annex 1 Video recording transcript of interrogation	Annex 2 Video recording transcript of confidential investigative actions	Annex 3 Video recording transcript of crime reenactment	The total amount of psychological signs in the fragment.
No. 1 "Going to E.I. Petrova"	14	5	3	22
No. 2. "Interaction with E.I. Petrova that preceded the act of violence"	12	4	30	46
No. 3. "Act of violence against E.I. Petrova"	21	19	28	68
No. 4. "Leaving the crime scene"	6	7	7	22
Total	53	35	68	

As the above calculation demonstrates, more psychological signs of reliability are concentrated in the ***No. 3.*** Fragment (***"Act of violence against E.I. Petrova")*** at all stages of investigative actions which is particularly the case when dealing with truthful testimony.

Action 2.2

Populating the matrix of psychological indications of reliability

The process of identifying psychological indications of reliability also includes populating the matrix of psychological indications encountered in a free narrative. One investigative activity that most fully represents the testimony of the person under investigation is selected for the purposes of populating the matrix. The interrogation normally fits this requirement.

This activity is used to calculate the recorded psychological indications of reliability falling into two categories. The matrix was developed especially for the purposes of monitoring the required set of psychological indications of reliability[7] (Shapovalov, 2013) in a free narrative.

Fig. 3. Sample Matrix of Psychological indications of Reliability

Матрица психологических признаков достоверности в свободном рассказе

[illegible]	Психологических признаков достоверности	Свободный рассказ			
		Ф.1	Ф.2	Ф.3	Ф.4
1 категория (обязательные)	(5.1., 5.2.) Логичность				
	6.1. Описание подэкспертным собственных действий. 6.3. Привязка подэкспертным собственных действий в пространстве			/	
	6.4. Общее описание подэкспертным собственных высказываний. 6.5. Цитирование собственных высказываний *(по отношению к ключевому объекту события)**			/	
	6.9. Описание подэкспертным физического воздействия на ключевой объект события. Описание результата воздействия.				
	7.2. Описание подэкспертным действий ключевого объекта события. 7.4. Привязка действий ключевого объекта события в пространстве			/	
	7.5. Описание подэкспертным физического воздействия на него, со стороны ключевого объекта события. Описание результата воздействия.				
	7.6. Общее описание подэкспертным высказываний ключевого объекта события. 7.7. Цитирование подэкспертным высказываний ключевого объекта события *(по отношению к подэкспертному)**			/	
	9.2. Внесение уточнений по собственной инициативе. Начало рассказа после длительной паузы по собственной инициативе. Возврат к главной темы после отвлечения				
	9.6. Иллюстративные жесты				
	9.7. Мимико-пантомимическое сопровождение речи				
	$\sum_1$ – сумма балов по категории 1	=			
2 категория (факультативные)	2.2. Интонация, особенности высказываний				
	5.1. Визуальная информация				
	5.2. Аудиальная информация				
	5.3. Информация, связанная с запахами				
	5.4. Информация, связанная с вкусовыми ощущениями				
	5.5. Физические ощущения				
	5.6. Описание материальных объектов, живых существ, особенностей окружающей среды				
	6.2. Привязка подэкспертным собственных действий во времени				
	6.5. Цитирование собственных высказываний *(по отношению к второстепенным участникам события)**				
	6.6. Эмоциональное состояние, проявления				
	6.7. Физиологические потребности, проявления, состояние				
	6.8. Сообщение о своих мыслях, размышлениях, возникших в момент события				
	7.1. Использование имен, прозвищ и других обращений				
	7.3. Привязка действий ключевого объекта события во времени				
	7.7. Цитирование подэкспертным высказываний ключевого объекта события *(по отношению к второстепенным участникам события)**				
	7.8. Сообщение подэкспертным об эмоциональном состоянии ключевого объекта события				
	7.9. Сообщение подэкспертным о физиологических потребностях, проявлении, состоянии ключевого объекта события				
	7.10. Описание подэкспертным особенностей внешнего вида ключевого объекта события				
	8.1. Необычные подробности				
	8.2. Неожиданные трудности во время события				
	8.3. Использование в речи жаргонизмов, прилагательных, эпитетов, и уменьшительно-ласкательной формы для передачи свойства предмета				
	9.1. Контекстуальные вставки				

7 **An essential set of psychological indications of reliability**. A truthful narrative must earn at least 10 points in the category of "essential psychological indications of reliability" and at least 11 points in the category of "optional psychological indications of reliability". This makes it possible to conclude that the scope of details is rich, and that the person in question is familiar with the informational, spatial, static, dynamic, and temporal characteristics of the event under investigation.

Category 1 includes psychological indications that must be present in a truthful narrative according to the content regularities that have been identified (a total of 10 indications).

This category includes:

- the logic of the narration;
- the ability of the person under investigation to describe their own actions and those of the victim in a spatial context;
- the ability of the person under investigation to quote their own speech and that of the victim (if the criminal act involved a verbal interaction);
- making unsolicited corrections;
- the mimic and pantomimic manifestations that accompany speech, etc.

Category 2 includes optional indications (a total of 30). According to the content regularities that have been identified, at least 11 of these 30 indications must be present in a truthful narration.

This category includes:

- the visual information;
- the description of physical objects, living beings, and specific features of the environment;
- the description by the person in question of their own emotional state at the time of the event;
- the communication of their thoughts and reflections occurring at the time of the event;
- the reporting by the person in question of the physiological needs, manifestations, and states of the key object of the event, etc.

Action 2.3

Recording psychological indications of unreliability

To formalize this process, a list of psychological indications of unreliability (25 indications) was prepared and a procedure for their recording in a special worksheet developed (Worksheet 2.3.1). The list includes such indications as:

- sketchiness of testimony;
- alogism (illogical nature) of testimony;
- rationalizations, generalizations;
- cognitive operations;
- probabilistic wording in response to a meaningful question or topic;
- reduced number of self-references;
- reduced locomotion (illustrator gestures, head, finger, wrist, foot movements) in response to a meaningful question or topic, in contrast with other fragments of testimony;

Psychological indications of unreliability are manifested in ***cognitive, behavioral, and mnestic*** text-forming regularities and in the incongruent content of statements made and emotions experienced.

Cognitive regularities are manifested in the process of mental construction of a certain event and its subsequent recounting in testimony, and also during alteration and deliberate distortion of the circumstances of a past event.

a) <u>manifestation in truthful testimony:</u> studies indicate that cognitive regularities are not normally present in truthful testimony, because a person only needs to describe their memories of a past event when asked to do so (Johnson, Raye, 1981).

b) <u>manifestation in fabricated testimony:</u> the process of constructing fictitious testimony affects their semantic component (sketchiness, scant details) and specifics of recounting (long pauses, repetition of the same statement, and other psychological indications of unreliability).

There is an example of a completed

worksheet 2.3.2 below where counting of psychological signs of unreliability is performed which helps to identify in which testimony fragments their most concentration is shown. The greatest amount of psychological signs of unreliability in a certain fragment may be an indication of its unreliability, that is, it is highly likely the case of deliberate distortion or concealment of valuable information on the part of investigated person associated with this fragment.

Worksheet 2.3.2 The calculation of psychological signs of *unreliability* in the investigation action and their key fragmentation.

Testimony Fragment	Annex 1 Video recording transcript of interrogation	Annex 2 Video recording transcript of confidential investigative actions	Annex 3 Video recording transcript of crime reenactment	The total amount of psychological signs in the fragment.
No. 1 "Going to E.I. Petrova"	2	1	1	4
No. 2. "Interaction with E.I. Petrova that preceded the act of violence"	1	0	1	2
No. 3. "Act of violence against E.I. Petrova"	0	1	1	2
No. 4. "Leaving the crime scene"	1	3	2	6
Total	4	5	5	

As it is seen from a performed calculation, the highest number of psychological signs of unreliability is concentrated in the 4th fragment, which may indicate the attempt of an investigated person either hide or modify valuable information associated with this fragment.

Action 2.4

Qualitative measurement of psychological indications of unreliability

Unlike psychological indications of reliability, psychological indications of unreliability are less conclusive and sooner point to critical areas in testimony that need to be examined thoroughly using not only psychological assessment but also other forensic techniques, which is an area of activities for other experts. Manifestation of nonvocal indications does not necessarily mean that the person in question is lying; they could be merely a reflection of stress experienced by him as he recounts certain events.

It is the job of the expert psychologist to identify and record the psychological indications of unreliability and scientifically substantiate their nature and possible causes.

Procedure No. 3

Comparative analysis of testimonies

The procedure of comparative analysis of testimonies is a verification procedure designed to identify ***mnestic (memory)*** text-forming regularities.

Mnestic (memory) regularities are associated with the regularities of short-term and long-term memory processes.

a) ***manifestation in truthful testimony:*** when a person experiences a certain event, their memory is saving a complex of sensations (visual images, sounds, smells, tactile sensations, emotions, etc.) and an understanding of the general context of the situation with its cause-and-effect relationships and motives for actions/interactions. Normally, this complex of information about the event is spontaneously, without a conscious effort to remember or memorize, committed to long-term memory, especially if the event itself caused the person to experience intense emotions, and in the majority of cases can be easily verbalized by the person on multiple occasions without changes in the key aspects of the testimony.

b) ***manifestation in fabricated testimony:*** fictitious events and actions are normally based only on cognitive operations (i.e. on fiction) and are not backed up by a complex of sensations (experience), which significantly reduces the chances of a transition of previously recounted fictitious information from short-term to long-term memory, and affects the accuracy with which it is recounted during another round of questioning. Thus, if the testimony of the person in question is not backed up by a complex of sensations (experience) but is based on fiction, a diagnostically significant variability will be traceable during subsequent interrogations (especially if the event under investigation consists of multiple episodes).

Action 3.1

Formalized comparative analysis of testimonies given by the person in question at different times

This action can have the desired diagnostic effect when at least two video recordings are available of the person in question giving testimony regarding the event being investigated. In some cases, the comparative analysis may be performed using transcripts of interrogations that were not video recorded. The comparative analysis is performed using a special worksheet in keeping with formal requirements (Worksheet 3.1).

The findings of a comparative analysis of testimony given by the person in question during different phases of investigative activity are used to determine the consistency or inconsistency of the semantic/content component of statements about the details and circumstances of the event.

Fig. 4. Sample worksheet 3.1 for performing the comparative analysis of testimony

Fragment	APPENDIX 1 Transcript of the video recorded interrogation of V.V. Ivanov on October 10, 2012	APPENDIX 2 Transcript of the video recorded crime reenactment (free narration phase) with V.V. Ivanov on October 12, 2012	APPENDIX 3 Transcript of the video recorded crime reenactment (onsite testimony verification phase) with V.V. Ivanov on October 12, 2012	Variability / stability
No. 2 "Consumption of alcohol together with A.A. Sidorov"	**a) Details of the surroundings in the backyard of A.A. Sidorov.**			
	17-18 He invited me to come in for a drink. I did, and we had a drink. He had a small bottle [00:07:31 illustrator gesture] that was not quite full, maybe 50 grams **(1)**. We drank it.	5-6 We went to his place (Where?) in the backyard (In what street?) General P-a Street. We came into his place, and he had a small bottle that was not quite full **(1)**. He and I drank it.	6-8 [00:02:05] (the couch was on this side) yes, there was a stool here [illustrator gesture]; he and I were seated on the couch. I sat here [illustrator gesture] and he sat there (Where did it stand? Here?) Yes, right in the middle there was a small bottle on it, not quite full **(1)**. We sat and had a drink.	**(1)+**
	b) V.V. Ivanov buys more moonshine			
	27-31 Then I said "Let me go get some more". He said "I don't mind but got no money", to which I said that I would borrow some. I went to Tanya Tiutiunnik and borrowed a bottle from her (What kind of bottle?). Half-liter bottle (Full?) Yes. I brought it to him. I went right back to him. This took about 15 minutes **(1)** [00:09:34] There and back. I came back to him and we sat down to drink.	6-9 We decided to get more (Where were you drinking?) In the backyard under the awning. He has a vineyard there. I went to borrow some and I was away for 15 minutes. I came back with a half-liter bottle **(1)**. He and I sat down and we continued drinking.	9 [00:02:48] We were sitting and drinking. We finished the small bottle, and I went to get another one **(1)**.	**(1)+**
	c) An unidentified man enters			
	32-36 We each had a shot **(1)**... Then a guy came to him, and they both walked off in the direction of the ... sauna **(2)** [00:10:01 illustrative gesture] (What was the young man wearing?) Light colored shorts and some other light colored piece of clothing. I did not look at him much. I was already, you know. They spoke for about 10 minutes **(3)**. I did not hear what they were talking about. This guy left without sitting down at the table.	9-12 We each had a shot **(1)**. A guy came to him. They walked off and talked **(2)**. I did not hear what they were talking about. They talked for about 10 minutes **(3)** (Do you know that young man or not?) No, I don't know him. When they were done talking, that guy left. He wouldn't drink with us or anything. He just took off.	9-11 We had some drinks from the bottle **(1)**. A guy came to him. They stood [00:03:01 illustrator gesture] there and talked (At the corner?) Yes, they went over there **(2)**.	**(1)+** **(2)+** **(3)+**

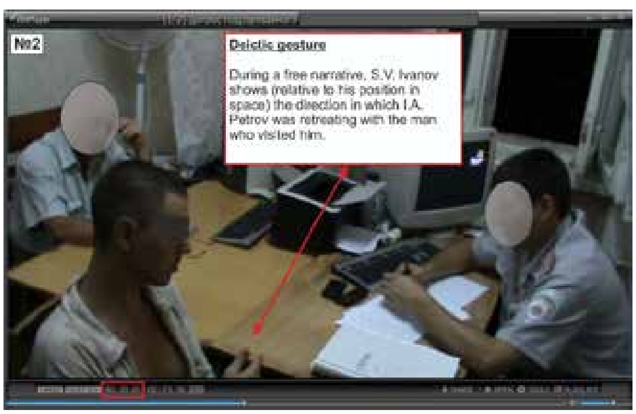
Action 3.2

The comparative analysis of mimic and pantomimic manifestations of the person in question recorded at different times and relating to one and the same testimonial context (freeze frame analysis)

The purpose of the freeze frame analysis is to:

- determine the consistency or inconsistency of illustrator gestures and mimic and pantomimic manifestations that accompany speech, with the statements content;
- determine the consistency or inconsistency (repeatability) of illustrator gestures relating to the same act (point in space, object, action or interaction, etc.) recorded during different investigative activities while describing one and the same event;
- record nonvocal characteristics from the list of psychological indications of unreliability and correlate them with the content of communications;
- make an illustrative presentation of freeze fames in combination with the semantic component of statements.

Example No. 1. Consistency of the deictic gesture



In this example, the consistency of the so-called deictic gesture is illustrated. On the first freeze frame, the person under investigation, during his interrogation on his own initiative, points with his right hand, in which direction his acquaintance I.A. Petrov walked with the man who came to him. Then he repeats the same gesture during the investigation experiment. Due to the fact that non-verbal manifestations, as a rule, are not controlled by consciousness, especially taking into account their repeatability and consistency, it can be said with a high degree of probability that an untried person reports information about a really experienced event. These features correspond to the laws of the mental process of reproduction.

Example No. 2. Inconsistency of the deictic gesture

In this example, the mismatch of a deictic gesture is illustrated. On the stop frame No. 1, the person under investigation during the stage of the free story, points with his left hand to the side where he saw the victim as soon as he left the car. However, during the verification of the testimony on the spot, leaving the car, V.V. Sidorov could not see the victim in that place about himself, which he indicated during the stage of free story. On the stop-shot number 2 V.V. Sidorov indicates where the victim was in relation to him, when he got out of the car. The registered feature of non-verbal behavior of Sidorov VV may indicate an attempt to conceal his share of participation in the commission of criminal acts against the victim.

Thus, the production of this procedure makes it possible to clearly illustrate whether the pantomimic manifestations of the researcher are consistent with the semantic component of the utterances. The consistency of the verbal and non-verbal behavior of the subject is a sign of possession of spatial and dynamic characteristics of the event under investigation.

Behavioral regularities that accompany text-forming. Regularities of manifestation of nonverbal communicative means depending on the content of statements (description of a real-life event vs. a fictitious event).

a) ***manifestation in truthful testimony:*** the events described are freely accompanied by illustrator gestures that demonstrate the size of objects, the spatial position of objects (deictic gestures), actions/interactions, and mimic-pantomimic manifestations. Mimic-pantomimic manifestations reflect the nature of the emotional undertone of the event. Due to the fact that gesture production is a largely subconscious process, an equally important indicator of the testimony reliability is the consistency and repeatability of illustrator gestures when a person repeats their narrative of the same event. The consistency of illustrator gestures may serve as an indirect proof that testimony is based on memories of real-life experiences;

b) ***manifestation in fabricated testimony:*** a general freezing of locomotion when recounting a key fragment of testimony with a possible manifestation of adaptor gestures. Also, when the person under investigation has consciously altered a certain fragment of the event or the share of their involvement or has inverted their own role (from the role of an assailant to that of a person defending himself), inconsistency and contradictions between deictic gestures and gestures describing actions/interactions can manifest themselves during subsequent interrogations. These contradictions can manifest themselves particularly clearly when the pantomimic manifestations of the person under investigation during an interrogation are compared to those during a crime reenactment.

Consistency of the content of statements and emotions experienced by the person.

a) ***<u>manifestation in truthful testimony:</u>*** consistency of emotions experienced and mimic expressions accompanying them, and the context of the event being described. For example, when describing a sudden attack the person in question can unconsciously show surprise or fear on their face. In some cases, when emotions are particularly intense, a person can begin to cry (this is more common for children and women).

b) ***<u>manifestation in fabricated testimony:</u>*** inconsistency, conflict between the context of the event being described and the emotion being experienced. For example, the mother is outwardly calm when reporting the abduction or death of her child.

Procedure No. 4

Analysis of communicative interaction between the person in question and participants of investigative activities

During an analysis of video recorded communicative interaction between the person in question and the participants of investigative activities (primarily, the investigator), account is taken of:

- the specifics of questions wording by the investigator and other participants of investigative activities (the law prohibits asking questions that are worded to include an answer, a portion of an answer, or tips in the form of leading questions);

- whether or not the person in question was given an opportunity to present their testimony in the form of a free narrative;

- whether or not the person in question is giving meaningful answers to the investigator's questions and supplements them with secondary details;

- whether or not the person in question is refining the testimony without prompting or initiating by the investigator.

The findings of the analysis of the aforementioned specifics of communicative interaction between the person under investigation and the participants of investigative activities are used to prepare an opinion on

Fig. 5. Sample worksheet for calculating the types of questions asked of the person in question during an investigative activity.

Testimony Fragment	Open-ended questions	Alternative Questions	Suggestive (Leading) Questions	Total
No. 1 "Going to E.I. Petrova"	0	0	0	0
No. 2. "Interaction with E.I. Petrova that preceded the act of violence"	0	0	0	0
No. 3. "Act of violence against E.I. Petrova"	5	3	0	8
No. 4. "Leaving the crime scene"	0	0	0	0
Total	5	3	0	8

whether or not there are signs of undue psychological pressure exerted on the person under investigation during the course of the investigative activity.

A motivation to present testimony as a regularity of text-forming. At the core of this regularity is the primary motivation of the person in question to present testimony and the effect that basic emotions have on the process of speech production.

a) ***manifestation in truthful testimony:*** a feeling of guilt over their deed can motivate the person under investigation to confess, which in turn is a means of obtaining psychological relief and a way to expiate guilt. This enables the person to "get things off their chest" and reduce the destructive impact of the negative emotion being experienced. When the person under investigation decides to confess in order to relieve the burden of this emotion, usually the *position of the leader* is assumed, and not the follower: their testimony is filled with details and their behavior is full of illustrator gestures; they are also inclined to add unsolicited information.

b) ***manifestation in fabricated testimony:*** the fear of being exposed as the perpetrator of the crime; the desire to avoid a sense of shame or an otherwise humiliating situation. This emotion freezes the gestures and has a disorganizing effect on oral speech; as a result, the person in question does not demonstrate a motivation to present testimony, assumes the *position of the follower*, and gives curt answers to questions without volunteering additional descriptions. The same characteristics can also be manifested in speech and behavior when the person has been forced to make false confession (compulsive lie) when the testimony of the person under investigation was obtained with the use of undue methods of influence.

Synthesis section (integration of investigation results)

Results obtained during the aforementioned procedures are integrated in the synthesis section and supplemented with the expert's commentary. Based on the results of the completed investigation and the aggregation of available data, the expert draws a conclusion on whether or not any psychological indications of reliability or unreliability (or their individual fragments) are present in testimony given by the person in question during the course of investigative activities.

The expert psychologists will use the synthesis section to describe the psychological indications by each fragment of testimony. If necessary, for illustration purposes the expert can include the most diagnostically relevant statements of the person in question, while making sure to include references to annexes and lines from which they are quoted.

Formulation and presentation of the conclusions based on the completed investigation

The expert psychologist will make brief meaningful summary of the investigation results in the context of the questions addressed during the investigation. The conclusions must be highly specific and pertain to semantic and not structural parameters of the process. The synthesis section serves as a basis for drawing the conclusions.

Some examples of conclusions are presented below. However, they should not be viewed as templates, as conclusions must be drawn on a case-by-case basis and necessarily derived from the synthesis section:

- The testimony of the person in question, which as a whole relates to the event under investigation, has been found to contain a set of psychological indications of reliability, which show that the subject is quite familiar with the informational, spatial, static, dynamic, and temporal characteristics of the event under investigation;
- The recorded specifics of presenta-

tion of the testimony by the person in question and the communicative interaction with the participants of the investigative activities indicate that during the investigative activities the subject recounted the details of the event independently and was in a position of the leader, not that of the follower;

- The analysis of the structure of the video recorded communicative interaction between the participants of the investigative activities and the person in question has not revealed any signs of undue psychological pressure;

- The testimony given by the person in question concerning the intermediate stage of the physical confrontation has been found to contain a number of contradictions known as psychological indications of unreliability. These contradictions may be attributable to the natural process of forgetting. However, it cannot be ruled out that the person in question has an ulterior motive to conceal or deliberately distort their involvement in this crime. The process of deliberate concealment or distortion of information requires cognitive efforts and consideration of spatial and dynamic characteristics of the event, which may result in this kind of errors and inaccuracies;

- The video recorded testimony by the person in question shows a lack of logic, a poorly structured testimony, a testimony that is inconsistent, contradictory, and variable, and lacks the required set of psychological indications of reliability, which is a sign that the person in question is experiencing considerable difficulty in recounting the details associated with the event under investigation, while the aggregation of facts presented cast doubts on their familiarity with the informational, spatial, static, dynamic, and temporal characteristics of the events recounted;

- A comprehensive analysis of the structure of the video recorded communicative interaction between the participants of the investigative activities and the person in question as an interrogated person makes it possible to trace the following factors: an interrogation is structured in the “question-answer” format; the investigator asks questions that contain possible answer alternatives as well as leading (suggestive) questions; a free narrative stage as a basis of the confession is absent at all phases of investigative activities. The aggregation of these factors indicates that the speech activity of the person in question was fully subordinated to answering questions from the investigator, which is a sign of undue psychological pressure on the person under investigation during the course of investigative activities.

References

Johnson M. K., & Raye C. L. (1981) Reality monitoring, *Psychological Review*, Vol. 88, 1, 67—85.

Shapovalov V. (2015). *Technique for psychological assessment of the testimony reliability in legal practice.*, Kiev : Osvita Ukrainy, 60 pp.

Smith N. *Reading between the Lines: An evaluation of the Scientific Content Analysis technique (SCAN).* London, 2001.

Steller, M. & Köhnken, G. (1989) Criteria-based statement analysis. In D.C. Raskin, D. (Ed.) *Psychological Methods in Criminal Investigation and Evidence.* New York: Springer Publishing Company, 217-245.

Undeutsch U. (1989) The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility Assessment*, Dordrecht : Kluwer, pps. 101—119.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Zeugenaussagen. [Assessing validity of testimonial statements]. In U. Undeutsch (Ed.), l-Iandbuch der Psychologie. Vol. II: Forensische PsycholoKie (pp. 26-181). Gottingen: Verlag filr Psychologie. (Japanese translation by fl. Uemura (1973), Shogen no shinri. Tokyo: University of Tokyo Press.]

Фрай О. (2005) Детекция лжи и обмана. – СПб.: ПРАЙМ-ЕВРОЗНАК, 2005 - 320с.

Шаповалов В.А.(2013) Методика определения психологических признаков достоверности / недостоверности в юридической практике : методическое пособие / В. А. Шаповалов. – 2-е изд-е, перераб. и дополн. — К. : Кафедра, 2013. — 108 с.

Шаповалов В.А. (2016) Методы психологической оценки достоверности сообщаемой информации : методическое пособие / В. А. Шаповалов. — Киев : Освита Украины, 2016. — 168 с.

Экман П., Фризен У. (2010) Узнай лжеца по выражению лица / П. Экман, У. Фризен ; пер. с англ. В. Кузина. — СПб. : Питер, 2010. — 272 с.

How to Calculate the Expected Agreement and the Combined Accuracy of Two Test Results

Raymond Nelson[1,4], John Kircher[2,4] and Mark Handler[3,4]

Abstract

We describe how to obtain the combined accuracy of the results of two independent tests and the expected rate of concordance or agreement between them. For non-independent tests – those in which the results of one test may have some shared source of variance with the other test – we show a table of results using the phi-correlation coefficient as the measure of correlation (dependence) between the tests. We manipulate the level of dependency between the tests (phi-coefficient) from 0 (independent) to 1 (perfectly correlated), and we manipulate the range of decision accuracy from .5 to 1 for both Test-1 and Test-2. We found that when the two tests agree on the outcome, decision accuracy increases to the extent that the two tests are independent (phi approaches 0). Decision accuracy for two tests can actually decrease under some circumstances, and this may be more likely when the phi correlation coefficient is high. Decision accuracy increases when the accuracy of individual tests increases and the outcomes for the two tests are less covariant.

We describe how to obtain the combined accuracy of the results of two tests beginning with the expected rate of concordance or agreement between two tests for which the results are independent. For non-independent tests – those in which the results of one test may have some shared source of variance with the other test – we show a table of results using the phi-correlation coefficient for two dichotomous outcomes. The phi-correlation (or Mean Square Contingency) coefficient (Mangal, 2010; "phi-coefficient," n.d.) is a measure of association between two binary variables (i.e. pass/fail, black/white, agree/disagree).

Agreement between two independent test results

The expected concordance rate for two tests is a function of the accuracy estimates for the two tests. The expected rate of agreement is the sum of the expected rate the two tests will be *correct and agree*, and the expected rate the two tests will be *incorrect and agree*. Step 1 shows the calculation of the rate that two independent tests will be correct and agree when both estimated to have an accuracy level of 80%. Step 2 shows the calculation of the expected rate that two independent tests with an estimated accuracy level of 80% will be incorrect and agree. Step 3 shows the sum of the expected agreement when the two tests are correct and incorrect.

1 Raymond Nelson is a psychotherapist and polygraph examiner who works as a research specialist with Lafayette Instrument Company, a company that develops and markets polygraph technologies. Inquiries can be sent to raymond.nelson@gmail.com.

2 John Kircher is a professor of educational psychology at the University of Utah, and is associated with Converus Inc. a company that develops and markets the EyeDetect ocular motor credibility assessment test.

3 Mark Handler is an experienced law enforcement polygraph examiner, the Editor in Chief for the American Polygraph Association, and works for Converus Inc. a company that develops and markets the EyeDetect ocular motor credibility assessment test.

4 There are no proprietary interests associated with the content of this publication.

```
Step 1: expected rate that two tests with accuracy = 80% will be correct
and agree.

.80 * .80 = .64

Step 2: expected rate that two tests with accuracy = 80% (20% incorrect)
will be incorrect and agree.

.20 * .20 = .04

Step 3: sum of expected agreement for correct and incorrect results.

.64 + .04 = .68
```

The combined rate of agreement for two independent tests for which the accuracy estimate of each is 80% will be 68%. Like all statistical calculations, this estimate is premised on certain assumptions. In this case, an important assumption is that the accuracy rate for each test can be characterized as a random variable. That is, that the results can take any of a range of values (from 0 to 1). Another important assumption is that the results from the two tests are independent of each other. Independence means that the two test results do not covary. For independent tests, whatever influences the outcome of one test will have no effect on the outcome of the other - independent outcomes have no source of shared variance.

Estimated accuracy when two independent tests agree

Combined test accuracy is different than test agreement or concordance. When results from two independent tests are in agreement the *aggregated accuracy* of the two results can be estimated as the ratio of expected agreement when correct to the total rate expected agreement when correct and incorrect. Example 1 shows the combined accuracy rate for two independent tests when the results are concordant using the values from Steps 1-3 above.

The aggregated accuracy rate of two tests for which the estimated accuracy of each is 80% is 94%. Of course, if the tests agree on 68% of occasions, they will disagree on 32% of occasions. Disagreements might be considered indeterminate. Once again, a combined accuracy rate of 94% rests on some important assumptions – that testing errors are random events, and that the tests results are independent. Under ideal circumstances the two tests would be completely independent – with no shared source of variance other than the criterion of interest. In real-life circumstances, perfect independence is difficult to achieve, just as perfect covariance is not likely to occur. More commonly there is some degree of correlation or covariance between two tests intended to quantify the same phenomena.

When two test results are correlated – as can be expected when the tests use the same methodologies, or when the administration of one test is influenced by knowledge of the other test result – it is possible that whatever caused the occurrence of an error at the first test might also cause the occurrence of an error at the second test. For this reason, these formula in steps 1, 2 and 3 cannot be taken as an expression of the expected accuracy when two tests results agree if the test are non-independent (or covariant).

```
Example 1: aggregated accuracy rate for concordant test outcomes.

.64 / .68 = .94
```

Combined accuracy of two non-independent (covariant) tests

Combined accuracy of two test results that are non-independent will be influenced by the degree of covariance between the two tests. Correlation of two binary variables (i.e., correct or incorrect outcomes from two tests) is described using the phi-coefficient.

One method to calculate the phi statistic is to take the Pearson correlation coefficient for all binary results of the sample of cases for which each case was evaluated with both tests. This is convenient because commonly available spreadsheet applications today include a variety of functions to calculate such mathematical statistics. A second way to calculate the phi-coefficient is to use the following formula (Formula 1.).

```
Formula 1: calculation of phi from observed frequencies.

phi = (A * D − B * C) / sqrt((A + B) * (C + D) * (A + C) * (B + D))
```

Values for A, B, C and D are taken from a 2x2 contingency table as shown in Table 1. Cell A is the proportion of cases where Test-1 and Test-2 are both correct. Cell B is the proportion of cases where Test-1 is correct and Test-2 is incorrect. Cell C is the proportion of cases where Test-1 is incorrect and Test-2 is correct. Cell D is the proportion of cases where both Test-1 and Test-2 are incorrect.

The marginal values A+B indicates the cases for which Test-1 was correct, while C+D indicates cases for which Test-1 was not correct. Similarly, the marginal values A+C indicates cases where Test2 was correct while B+D indicates cases for which Test-2 was incorrect. Values for A, B, C and D can be calculated from the three inputs: 1) A+B or the Test-1 cases that are correct, 2) A+C or the Test-2 cases that are correct, and 3) Cell A or the cases where both Test-1 and Test-2 are correct.

Table 1. 2x2 Contingency table for binary results of Test-1 and Test-2 for a sample of cases.

		Test-2		
		Correct	Wrong	
Test-1	Correct	A	B	A+B
	Wrong	C	D	C+D
		A+C	B+D	A+B+C+D

Values for cells B, C, and D can be obtained with subtraction. Thus, given the marginal

proportions .8, .2, .8, and .2, B = (A+B) – A = .80 - .71 = .09, C = (A+C) - A = .80 - .71 = .09 and D =

C+D – C = .20 - .09 = .11. The 2x2 contingency is shown in Table 2.

Table 2. 2x2 Contingency table for binary results of Test-1 and Test-2 for a sample of cases. (input values are underlined)

		Test-2		
		Correct	Wrong	
Test-1	Correct	**.71**	.09	**.80**
	Wrong	.09	.11	.20
		.80	.20	A+B+C+D = 1

After A, B, C and D are obtained, the total agreement between the two tests will be equal to A+D. Accuracy when the two tests agree will be equal to the total A / (A+D). The phi correlation between the two tests can be calculated using the formula shown earlier. The increase in accuracy is the difference between the input value A+B (Test-1 accuracy) and the accuracy of the two test results when they agree. Formula 2 shows the increase in test accuracy for the two tests in Table 2, compared to the accuracy of just the first test (input A+B).

Formula 2: increase in accuracy for Test-1 and Test-2. Increased

accuracy = (A / (A + D)) – (A + B)

= (.71 / (.71 + .11)) - (.71 + .09)

= (.71 / .82) – (.80)

= .87 - .80

= .07

Results from formula 2 show a combined accuracy rate of .87 for the data for two tests shown in Table 2. This is an increase in accuracy of 7 percentage points compared to a single test with an accuracy rate of 80%. Formula 3 shows the calculation of the phi-coefficient for the data in Table 2.

Formula 3: calculation of phi for Table 2. phi

= (.71 * .11 – .09 * .09) /

sqrt((.71 + .09) * (.09 + .11) * (.71 + .09) * (.09 + .11))

= .0781 – .0081) / sqrt(.80 * .20 * .80 * .20)

= .07 / .16

= .44

The phi-coefficient for Table 2 is .44. Like the correlation coefficient, the phi statistic for 2x2 tables will give a signed decimal value between -1 and +1, with the value zero indicating no relationship or complete independence between the two test results, which can also be thought of as a random relationship. Perfect independence is not expected, but values closer to zero signify greater independence for the results of the two tests.

It is possible to calculate the test agreement and accuracy for a range of possible values for cells A, B, C and D. The results can be displayed in a table format with the calculations for phi, agreement, accuracy when the two test results agree, and the difference between the accuracy of two tests and one test (shown as Accuracy Increase). To do this it is necessary to calculate possible permutations of A, B, C and D for the range of input values for Test-1 accuracy (A+B), Test-2 accuracy (A+C) and the agreement between Test-1 and Test-2 (cell A). Appendix A shows an R function to accomplish the permutation and calculations of the Table values.

Table 3 shows the reduced output from the R function (R Core Team, 2016) in Appendix A while varying the results of Test-1 and Test-2 from .5 to .99 and also varying the proportion of agreement between Test-1 and Test-2 from .01 to .99. Results shown in Table 3 are sorted for the accuracy increase (Accy. Increase column) when Test-1 and Test-2 agree.

Table 3. Reduced output for agreement and accuracy for 2 tests for a range of phi correlation values.

Ipput A+B	Input A+C	Input A	B	C	D	phi	Agreement	Accuracy when 2 tests agree	Accy. Increase
0.7	0.5	0.41	0.29	0.09	0.21	0.26	0.62	0.66	-0.04
0.6	0.5	0.41	0.19	0.09	0.31	0.45	0.72	0.57	-0.03
0.8	0.5	0.41	0.39	0.09	0.11	0.05	0.52	0.79	-0.01
0.5	0.5	0.31	0.19	0.19	0.31	0.24	0.62	0.5	0
0.5	0.5	0.41	0.09	0.09	0.41	0.64	0.82	0.5	0
0.6	0.5	0.31	0.29	0.19	0.21	0.04	0.52	0.6	0
0.7	0.6	0.51	0.19	0.09	0.21	0.4	0.72	0.71	0.01
0.6	0.6	0.51	0.09	0.09	0.31	0.62	0.82	0.62	0.02
0.8	0.6	0.51	0.29	0.09	0.11	0.15	0.62	0.82	0.02
0.7	0.7	0.61	0.09	0.09	0.21	0.57	0.82	0.74	0.04
0.8	0.7	0.61	0.19	0.09	0.11	0.27	0.72	0.85	0.05
0.6	0.6	0.41	0.19	0.19	0.21	0.21	0.62	0.66	0.06
0.5	0.6	0.41	0.09	0.19	0.31	0.45	0.72	0.57	0.07
0.8	0.8	0.71	0.09	0.09	0.11	0.44	0.82	0.87	0.07
0.9	0.9	0.81	0.09	0.09	0.01	0	0.82	0.99	0.09
0.5	0.6	0.31	0.19	0.29	0.21	0.04	0.52	0.6	0.1
0.6	0.7	0.51	0.09	0.19	0.21	0.4	0.72	0.71	0.11
0.7	0.7	0.51	0.19	0.19	0.11	0.1	0.62	0.82	0.12
0.7	0.8	0.61	0.09	0.19	0.11	0.27	0.72	0.85	0.15
0.5	0.7	0.41	0.09	0.29	0.21	0.26	0.62	0.66	0.16
0.6	0.8	0.51	0.09	0.29	0.11	0.15	0.62	0.82	0.22
0.5	0.8	0.41	0.09	0.39	0.11	0.05	0.52	0.79	0.29

* Table 3 contained over 27,000 rows before reduction for this example.

Of great interest is the fact that under some circumstances accuracy when two tests agree is lower than accuracy for a single test. (An example of this is the first row where the accuracy of Test1 is 0.7 and the accuracy when the two tests agree is 0.66.) Inspection of the details of Table 3 suggests that lower combined test accuracy may be related to conditions where the phi-coefficient is high and the difference in accuracy for Test-1 and Test-2 (margins A+B and A+C) is greater. For example, in the first row, the chance performance of Test-2 reduces the accuracy that would have been achieved if only Test-1 had been administered.

Discussion

Under the ideal circumstance that two test results are independent -- they have no shared source of variance -- calculation of the expected rate of agreement between the two tests is simple and straightforward. Another reason this estimate is imperfect is that it regards test errors as if they are completely random events. In reality, testing errors may occur due to systematic causes. Systematic causes of error may be may be due to individual or group differences physiology, psychology, or level of development. They may also occur in the form of deliberate strategies intended to alter a test result. Systematic causes of error may be especially problematic when the two tests are not independent, such as when they use similar methodologies or when the tests are administered under non-blind conditions (wherein an evaluator knows the outcome of the other test).

Conclusion

The correlation between binary outcomes of two tests can be calculated using the phi-coefficient which can be interpreted in a manner similar to the Pearson correlation coefficient. Under circumstances where the phi-coefficient approaches the value 1 – where there is very high correlation between outcomes – there is no increase in accuracy when the two tests agree on the outcome as compared to the accuracy achieved by a single test.

Increases in accuracy when two test results agree, are associated with lower phi-coefficients (i.e., when there is greater independence between the two tests). When the two test results agree, decreases in test accuracy are associated with lower independence and greater covariance between the tests. A practical implication of this is that the strategic and effective use of multiple testing strategies will require some understanding of the concepts and principles of science, testing, statistical classification, and basic probability theory. Effective use of multiple testing strategies may also depend in part on the independence of the two tests.

References

Mangal, S. K. (2010). *Statistics in Psychology and Education.* New Dehli: Phi Learning Private Limited.

Phi coefficient. (n.d.). In *Wikipedia*. Retrieved September 22, 2017 from https://en.wikipedia.org/wiki/Phi_coefficient.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Appendix A.

R Function to Permute the Table of 2x2 Matrices for Two Tests

```
permutePHI <- function(AB=seq(from=0, to=1, by=.01),
                       AC=seq(from=0, to=1, by=.01),
                       A=seq(from=0, to=1, by=.01),
                       rem.neg=FALSE ) {
  # R function to permute a 2x2 matrix, phi and accuracy for the binary outcome of two tests
  # 9/18/2017 Raymond Nelson
  ####
  # input can be vectorized
  # input AB is the marginal sum of the criterion for cellA and cellB
  # input AC is the marginal sum of the criterion for cellA and cellC
  # input A is the proportion of the criterion agreement between test1 and test2
  # criterion is whatever condition satisfies cellA. for example: correct decisions
  # this may also work for other kinds of dichotomous outcome such as truth or deception
  # rem.neg will remove rows where phi is < 0
  # output is a data frame with cols for 2x2 cells and marginal values
  ####
  # permutation
  colAB <- rep(AB, each=(length(AC)*length(A))) # use "each" to repeat each item
  colAC <- rep(rep(AC, each=length(AB), times=length(A)))
  colA <- rep(A, times=(length(AB)*length(AC))) # use "times" to repeate the vector
  # construct a data frame from the permuted vectors
  DAT <- cbind.data.frame(AB=colAB, AC=colAC, A=colA)
  # remove permutations where A is greater than AB or AC
  removeRows <- which(DAT$A > DAT$AB | DAT$A > DAT$AC)
  if(length(removeRows > 0)) {
    DAT <- DAT[-removeRows,]
  }
  # calculate the cells
  # cellA is the proportion of criterion1 events where test1 and test2 agree
  # cellB is the proportion of criterion1 events for test1 but not test2
  # cellC is the proportion of criterion1 events for test2 but not test1
  # cellD is the proportion test1 and test2 events that fail criterion1
  DAT$cellA <- DAT$A
  DAT$cellB <- ifelse(DAT$AB == 0, 0, signif(DAT$AB * (1 - DAT$A / DAT$AB),2))
  DAT$cellC <- ifelse(DAT$AC == 0, 0, signif(DAT$AC * (1 - DAT$A / DAT$AC),2))
  removeRows <- which((DAT$cellA + DAT$cellB + DAT$cellC) > 1)
  if(length(removeRows > 0)) {
    DAT <- DAT[-removeRows,]
  }
  DAT$cellD <- round(1 - (DAT$cellA + DAT$cellB + DAT$cellC),2)
  # calculate the agreement between the two tests
  DAT$agreement <- DAT$cellA + DAT$cellD
  # calculate the phi coefficient
  DAT$phi <- round(((DAT$cellA * DAT$cellD) - (DAT$cellB * DAT$cellC)) /
                     sqrt((DAT$cellA + DAT$cellB) *
                            (DAT$cellC + DAT$cellD) *
                            (DAT$cellA + DAT$cellC) *
                            (DAT$cellB + DAT$cellD) ),2)
  # remove rows where phi is non numeric due to div/0
  if(length(which(is.na(DAT$phi)) > 0)) {
    DAT <- DAT[-which(is.na(DAT$phi)),]
  }
  # calculate the accuracy when the two tests agree
  DAT$accuracy <- signif(DAT$cellA / DAT$agreement,2
  if(length(which(is.na(DAT$accuracy)) > 0)) {
    DAT <- DAT[-which(is.na(DAT$accuracy)),]
  }
  # determine the increase over the margin A+B
  DAT$accyIncrease <- round(DAT$accuracy - DAT$AB,2)
  # remove rows where the phi coefficient is negative
  if(isTRUE(rem.neg)) {
    if(length(which(DAT$phi < 0) > 0)) {
      DAT <- DAT[-which(DAT$phi < 0),]
    }
  }
  return(DAT)
} # end permutePHI()
```

The Perceived Validity of the Polygraph in Polish Criminal Courts

Kamil Niziołek[1]

Faculty of Christian Philosophy, Cardinal Stefan Wyszynski University in Warsaw

Ewa Gołębiewska

Faculty of Christian Philosophy, Cardinal Stefan Wyszynski University in Warsaw

Abstract

The main aim of this study is to determine Polish judges' perceptions of polygraph validity in criminal cases. One hundred and thirty one (131) Polish criminal judges took part in the study. They were asked to provide a rating of how valid, in their opinion, polygraph examinations are in comparison to other types of forensic methods available. DNA analysis, Dactyloscopy and Balistics, received the highest ratings of all while especially low ratings have been found for Palinology, Presentation and Osmology. Polygraph received the lowest ratings of all. Participants also answered open-ended questions to allow them to qualify and explain the basis for their ratings of polygraphy. The majority of comments were negative and reflected the perception that polygraphy was a low validity forensic method, that it had a low level of legal utility, and that examiners were lacking in professionalism. The other comments were categorized as mixed or positive. The mixed comments focused on legal limitations of polygraph examinations, while the positive ones emphasized the perceived high level of professionalism among examiners and utility of this method.

Keywords: polygraph, validity, judges, perception, opinions, Poland

Introduction

In over a century of polygraph examinations history, there has been a great deal of research into its validity. Criterion validity refers to how well a test or model predicts the outcome of interest. In Physiological Detection of Deception (PDD) research, criterion validity is allegorical to test accuracy. Usually accuracy of test is being reported as True-Positive Index (TPI; the proportion of true positive results to all positive results), True-Negative Index (TNI; the proportion of true negative results to all negative results) and Total Index (the proportion of true positive and true negative results to all results). The only reliable way to determine accuracy has been to follow up on real cases where results of the tests were confirmed by independent evidences (American Polygraph Association, 1997).

The most accurate technique of polygraph examinations is the Comparison Question Technique (CQT). The roots of validated CQT can be found in Reid (1947). His research suggested that comparison questions, intend-

1 Correspondence concerning this article should be addressed to Kamil Niziołek, Institute of Psychology, Wóycickiego 1/3, 01-938 Warszawa, bud. 14, Poland

Contact: kamil.niziolek@outlook.com

ed to cause a response in truthful examinees, can improve test validity and reduce probability of false-positive errors. That finding was later refined by Backster in the Backster Zone Comparison Technique (1963). The most significant improvements concerned the procedures and rules for target selection, numerical scoring and question formulation.

An important moment in discussion on polygraph validity was a debate between Lykken and Raskin (Lykken, 1979; Podlesny & Raskin, 1977; Raskin & Podlesny, 1979). While it was generally related to data interpretation, it also dealt with a theoretical viability of polygraph examinations. Lykken argued that neither empirical nor theoretical aspect of polygraph examinations had shown their value. The contrary was claimed by Raskin, who showed that in some fields polygraph studies, validity can reach levels of about 90%.

In order to investigate an accuracy of Physiological Deception Detection (PDD) Abrams (1973) reviewed studies dating back to the early years of the 20th century, and reported an average accuracy rate of 98%. In later reviews, he gave slightly lower but still very satisfactory values, documenting the total validity of polygraph between 88% and 91% (Abrams, 1977, 1989).

Ansley (1983) reported the results of 1,964 laboratory cases and 1,113 field cases and described a validity rate of 96%, (inconclusive results were excluded). In later studies (Ansley, 1990), he summarized the results of 10 field studies, involving 2,042 criminal cases since 1980. Based on these studies, he reported a validity rate of 98% for deceptive cases and 97% for truthful cases, using the decisions of the original examiners.

Significant effort to investigate polygraph examinations validity was made by US government (U. S. Congress. OTA, 1983a, 1983b; U.S. Department of Defense, 1984). It is noteworthy that the Office of Technology Assessment (U. S. Congress. OTA, 1983b) reported that much of the research in this field was methodologically flawed. For their study of the validity of polygraph at the request of the United States Congress, only ten studies met their criteria for inclusion into their research. Seven of these studies dealt with the blind scoring of polygraph tests. Their research indicated a rather large range of accuracy for the various studies. On average, deceptive cases were correctly qualified about 86% correct, truthful cases 76%.

However, to estimate polygraph validity especially field studies should be considered. The author feels only the real cases where examinee's deceptions has been confirmed from external source can give us reliable validity data. Studies held before 1980s often involved instruments without amplifiers, non-standard examination formats, and poorly trained examiners. Also, those studies took inconclusive decisions into account. Such an approach was incorrect, because inconclusiveness is related to polygraph utility not validity. The polygraph average accuracy conducted in studies after 1980s has been reported at 98% (American Polygraph Association, 1997). Despite empirical evidence, documenting validity of polygraph examination, this method still remained questionable in late 1990s, even in the country of its origin (USA; Iacono & Lykken, 1997). At the beginning of new millennium Raskin and Honts (2002) reported the validity of the polygraph as lower, but still exceeding 90%.

Further confirmation of validity of polygraph examination came from a complex scientific review, which was completed by the National Research Council (2003) This study used the receiver operating characteristic (ROC) curve and considered the area under the curve (AUC) as an indicator of validity. It showed an average AUC of .860 for laboratory studies and .890 for field studies.

The validity of polygraph examination remains an important research topic. Recently, meta-analysis was used to calculate the effect size of validated polygraph techniques (American Polygraph Association, 2011). Results were summarized for 45 different samples from experiments and surveys, including scored results from 295 scorers who provided 11,737 scored results of 3,723 examinations, including 6,109 scores of 2,015 confirmed deceptive examinations, 5,628 scores of 1,708 confirmed truthful exams. Fourteen different PDD techniques were supported by a minimum of two published studies each that satisfied the qualitative and quantitative requirements for inclusion in the meta-analysis.

Comparison Question Techniques intended for event-specific (single issue) diagnostic testing, in which the criterion variance of multiple relevant questions is assumed to be non-independent, produced an aggregated decision accuracy rate of 89%. Comparison Question Techniques which were interpreted with the assumption of independence of the criterion variance of multiple relevant questions (multiple-issue and multiple -facet) and produced an aggregated decision accuracy rate of 85%. The combination of all validated polygraph techniques, produced a decision accuracy of 86.9% (American Polygraph Association, 2011).

Because this paper is focused only on judicial cases, it is worth to notice that according to APA standards of practice (American Polygraph Association, 2012) only tests that have accuracy over 90% and inconclusive rate below 20% can be used in judicial cases. The most common test employed in that field by professional examiners is Utah CQT with accuracy reported at 93% and inconclusive rate 10.7% (Handler & Nelson, 2009).

Despite confirmed accuracy of polygraph examinations, its application in court differs considerably worldwide. In Europe it ranges from substantial role in judicial system in Belgium to, for instance, to Spain where polygraph was never involved in court (Meijer & van Koppen, 2017). Application of polygraph examinations in Poland is a very specific issue, so a brief summary is needed.

Polygraph examinations in Poland have been developed since 1950s when Polish secret services included this method in espionage and internal affairs cases. Despite this well-known use of the polygraph for intelligence purposes, there was no polygraph examination in judicial cases until Horoszowski carried out an examination in 1963. From this time polygraph remains a controversial topic of many debates between court, politicians, scientists and examiners. After 1990 almost all of Polish security agencies and national intelligence applied polygraph examinations in recruitment process and internal investigations (Krzyscin, 2000; J. Widacki, 1975, 2007).

Despite this, in 2007 most of Polish judges did not find polygraph to be a reliable method, and reported that it was rather rarely employed in court (Pasko-Porys, 2007). That statement was later confirmed by Moszczyński (2011), who investigated accuracy of criminology methods as perceived by judges in Warsaw. His study, carried out in 2008-2009, showed that judges rated validity of polygraph as very low (2.9 in range from 1 to 10). A study conducted by Widacki (M. Widacki, 2013) showed that polygraph examinations in Poland are contaminated by both juridical authorities and examiners. On the one hand we can see that examinations are contracted too late (after interrogations and investigative experiments) and questions to examiners are defined inappropriately. On the other hand we can see a lot of methodological mistakes. The most common are: non-professional pretest interviews, incorrect structure of tests, inappropriate formulation of comparison and relevant questions and mistakes in data analysis.

From this time, there were at least two important events that may have influenced the state of affairs. Firstly, in November 2013 Polish Society for Polygraph Examinations (Polskie Towarzystwo Badań Poligraficznych – PTBP) was founded. The Society accepts the APA standards of practice and is ready to cooperate on promoting validated techniques and the highest ethical standards within the polygraph profession (Gołaszewski, 2014). Secondly, in 2015 Polish Supreme Court has definitively clarified problem of polygraph admissibility, stating its full admissibility either *in rem* (before indictment) or *in personam* (after indictment; Sygn. akt I KZP 25/14).

At this time, polygraph is hardly ever employed in criminal cases. Most recent data come from 2011, when on average, in 6,705 criminal cases only one polygraph examination was conducted (Widacki, 2012). However, this proportion seems to be higher year by year. One of the main causes of that state is low number of examiners in Poland. There are approximately 65 examiners, but only half of them are suitably qualified. About 45 examiners are members of local polygraph associations, which are Polish Society for Polygraph Examinations (Polskie Towarzystwo Badań Poligraficznych, PTBP) – leading association – and Polish Polygraphers Association (Stowarzyszenie Poligraferów Polskich, STP). About 15 of these examiners are American

Polygraph Association members. In Poland, polygraph examinations in criminal cases are conducted as forensic expert tests. In order to be allowed to conduct examinations an examiner needs to enjoy the status of expert witness (in Polish legal system – a person who has a special knowledge in certain area). For this purpose one needs to present certificates or any other relevant documents confirming his skills and knowledge. Unfortunately, this matter in not respected properly. Consequently, non-professional examiners are still present in Polish courts. They often lack knowledge about new research on polygraph, do not participate in any expert organizations and do not respect standards of practice. The low quality of such expert's comprehension undermines the role of polygraph in legal processes.

The advances in state of the art polygraph technologies and methods, along with the growing frequency of court cases in which polygraph plays a role in Poland, impose an obligation to investigate opinions of Polish judges on the validity of polygraph. It is necessary to establish whether polygraph enjoys a growing appreciation in the Polish juridical context. The aim of this study was to determine Polish judges' opinions about what the polygraph may contribute to the criminal justice system. In particular, it is to examine how valid, in their opinion, polygraph examinations are in comparison to other types of forensic methods, such as DNA analysis or dactyloscopy. It will be also verified, whether any prior experience with polygraph examination is related to a higher evaluation of validity of this method.

Materials and methods

Participants

One hundred and thirty one (131) Polish criminal judges from three levels of judicial districts (Rejonowy, Okręgowy Apelacyjny) took part in the study, 70 males (53.4%) and 61 females (46.6%). Work experience of subjects measured in years ranged from 2 to 50 (M = 18.23, SD = 8.58). All Polish judicial districts were involved in this study. Detailed information about participants is presented in Table 2 in the appendix.

Procedure

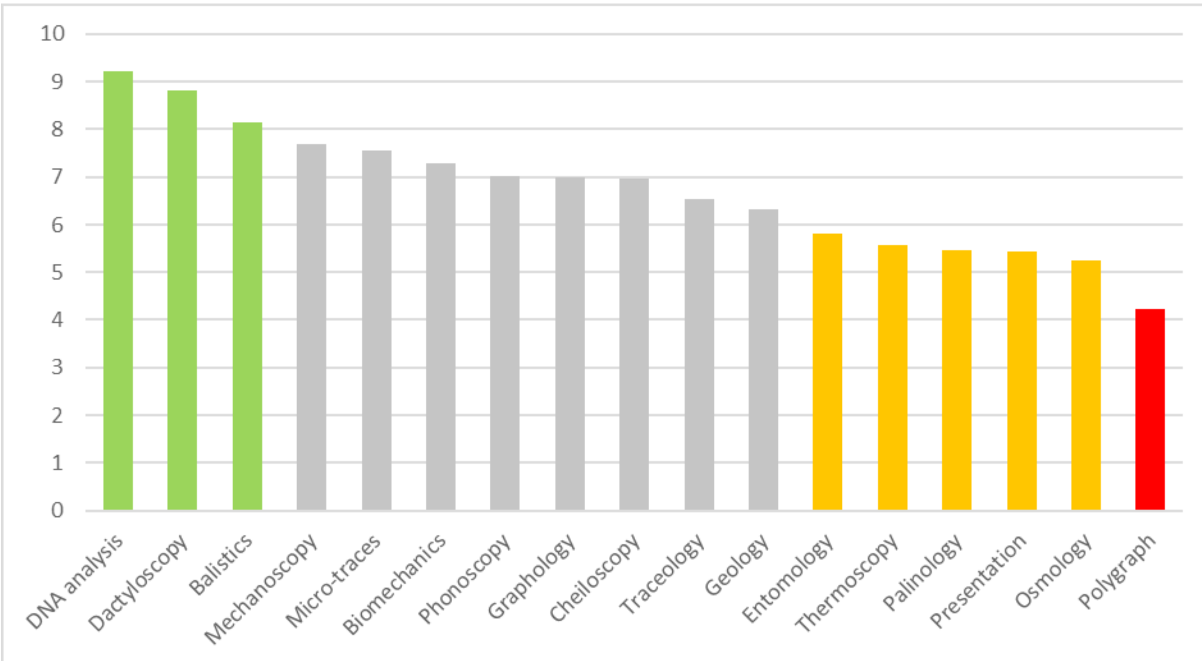
The participants took part in a Computer Assisted Web Interview (Śmiłowski, 2008). The survey concerned their opinions on validity of different forensic methods of investigation. In the beginning participants were informed about a specific definition of validity. To be precise they were given the following definitions: 'validity is the accuracy of examination's measurement', 'validity refers to how well an examination measures what it claims to measure', 'validity refers to how well an examination predicts the outcome of interest'.

Next, they were instructed to rate validity of all methods of criminal examinations on a scale from 0 (*not valid at all*) to 10 (*extremely valid*). The participants were also asked whether they had any prior experience with polygraph examinations in criminal cases (*yes/no*). This question will allow for describing judges' familiarity with the polygraph examination and for verification whether this familiarity is related to different opinions on the method. Mann-Whitney U test will be used to this end. Finally, the participants had an opportunity to give comments in an open-ended non-obligatory question regarding polygraph examinations "*What is your opinion about polygraph examinations?*". Three independent researchers coded these comments to establish whether they were positive, negative or mixed. A short summary of the content of these comments will also be provided.

Results

The mean ratings of validity for all methods of criminal examinations presented in the study are depicted in Figure 1. More details on distribution of judges' opinions can be found in Table 3 in the appendix. The highest ratings have been found in relation to DNA analysis (M = 9.21; SD = 0.99), Dactyloscopy (M = 8.82; SD = 1.21) and Balistics (M = 8.14; SD = 1.42), while especially low ratings have been found for Palinology (M = 5.47; SD = 2.19), Presentation (M = 5.44; SD = 1.83) and Osmology (M = 5.24; SD = 2.23). Polygraph received the lowest ratings of all considered methods (M = 4.24; SD = 2.23).

Figure 1. Mean validity of forensic methods as perceived by Polish criminal judges



The majority of judges in the study had no experience with polygraph. As many as 86 of them (65.5%) gave a negative answer to the question "Have you ever had an opportunity to involve polygraph examination in your work?". 45 participants (36.4%) declared that they had some experience with the method.

Judges unfamiliar with polygraph scored its validity rather low (M = 4.02; SD = 2.40), similar, yet slightly higher results have been found in group of judges familiar with this method (M = 4.64; SD = 2.17). Difference between the groups was evaluated using a non-parametrical Mann-Whitney test, since the assumptions for the parametrical t-test were not met (big differences in sample sizes, not normal distribution of variable). Mann-Whitney test indicated that the scoring of polygraph validity was not significantly higher for judges familiar with polygraph examinations (Mdn = 5) than for judges who had no experience with this method (Mdn = 4.5), U = 1631, p = 0.137.

Only 19 participants answered the last, open-ended question on polygraph validity. It is noteworthy that almost all comments were received from judges familiar with this method. Eleven of comments (57.8%) were negative. These statements focused on low level of legal utility of polygraph, its general low level of validity and on a low level of professionalism of examinations. Four mixed or neutral comments (21.1%) focused on high validity of polygraph examinations at *in rem* stage (before indictment) and on legal limitations of polygraph examinations in regard to *in personam* stage (after indictment). Four positive comments (21.1%) considered high level of professionalism of examiners and utility of this method.

Table 1. Comments regarding polygraph

Type of a comment	N (%)	Example
Positive	4 (21.1%)	"Outstanding professionalism of examiner."
Mixed	4 (21.1%)	"Because this forensic method is used to limit number of suspects, it is the most useful in the preparatory stage, so in legal proceedings it is rare in practice."
Mixed	4 (21.1%)	"Polygraph examination has operational (criminalistic) value, although it has no proceeding value. In Polish legal system there are no expert witnesses carrying out their own practice and there are no advisory/consultative institutions. There are some examiners who do not meet conditions of legal practice, that is do not have status of expert witness."
Negative	11 (57.8%)	"I am strongly against this way of determining facts, because there is a possibility that an examinee will interfere with results of the examination."
Negative	11 (57.8%)	"I am very sceptic about polygraph examinations; certainly I would not base sentence only on that evidence."

Discussion

Our results show high similarity to the findings of Moszczyński (2011), who investigated accuracy of criminology methods as perceived by judges in Warsaw. In both studies DNA analysis and dactyloscopy received the highest score while polygraph – the lowest. Although polygraph perceived validity seems to be higher than it used to be in the last decade, for Polish judges, it still remains the most questionable forensic method.

Our findings are also coherent with results of Widacki (2013). Both studies showed that the role of polygraph in legal processes in Poland is additionally undermined by presence of non-professional polygraph examiners. They often lack knowledge about new research on polygraph, do not participate in any expert organizations and do not always respect standards of practice. That statement is additionally supported by comments of the judges. Important parts of these comments addressed the examiner rather that method itself.

Just before writing this paper a new regulation of polygraph practice in Poland came into force - the Ministry of Development and Finances decree dated 10 April 2017 concerning psychophysiological examination, physical fitness test and psychological examination of Tax and Customs Board Officers (Dz. U. z 2017r., poz. 805). According to this regulation, a polygraph examiner requires the following; "graduated with a Master's Degree, accomplished professional course and obtained certificate recognized by American Polygraph Association (APA) or accomplished polygraph course at least elementary level or have appropriate skills of carrying out polygraph examination and results interpretation recognized by Polish polygraph organizations, including Polskie Towarzystwo Badań Poligraficznych (PTBP)". We hope that this legislation act will help to provide high levels of polygraph examinations, compatible with standard of practice.

Importantly, although Polish Supreme Court in 2015 has definitively clarified problem of polygraph admissibility, stating its full admissibility either *in rem* (before indictment) or *in personam* (after indictment; Sygn. akt I KZP 25/14), only few Polish judges involve this method, or even know about that possibility, in advanced phases of court cases.

Limitations and Future Research

The current study has some limitations. First, it was conducted on a relatively low number of participants. For some judicial districts only one judge participated in our survey. Furthermore, more in-depth qualitative research is needed to provide a wider picture of judges' opinions on validity of polygraph examination.

Future research should monitor opinions of Polish judges on the validity of polygraph. It is necessary to establish whether polygraph enjoys a growing appreciation in the Polish juridical context. Moreover, it is appropriate to investigate how polygraph validity is rated among other law executives, like prosecutors or lawyers.

Conclusion

To sum up, for Polish judicial instances polygraph examination is not a reliable method compared with others, for instance, DNA analysis. Despite empirical evidence polygraph is still being employed in judicial cases rather reluctantly. Opinion regarding polygraph validity of Polish judges is probably affected by a lack of knowledge about scientific research, the low number of professionals and unfamiliarity with polygraph standards of practice. One can only hope that debate about the use of polygraph in Poland will continue and will devote more attention to new scientific articles and the growing popularity polygraph acceptability elsewhere

References

Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Science*, *18*(4), 313–326.

Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.

Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington Books.

American Polygraph Association. (1997). The Validity and Reliability of Polygraph Testing. *Polygraph*, *26*(4), 215–239.

American Polygraph Association. (2011). Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph Techniques. *Polygraph*, *40*(4), 194–305.

American Polygraph Association. (2012). By-laws: American Polygraph Association, effective 1-1-2012. Retrieved April 21, 2017, from http://www.polygraph.org/

Ansley, N. (1983). A compendium on polygraph validity. *Polygraph*, *12*, 53–61.

Ansley, N. (1990). The Validity and Reliability of Polygraph Decisions in Real Cases. *Polygraph*, *19*(3), 169–181.

Backster, C. (1963). *Standardized polygraph notepack and technique guide: Backster zone comparison technique*. New York: Cleve Backster.

Gołaszewski, M. (2014). Polish Society for Polygraph Examinations was founded. *European Polygraph*, *8*(1), 16–17. https://doi.org/10.2478/ep-2014-0004

Handler, M., & Nelson, R. (2009). Utah Approach to Comparison Question Polygraph Testing. *Polygraph*, *38*(1), 15–30.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology*, *82*(3), 426.

Krzyscin, A. (2000). The Debate Over Polygraph in Poland. *Polygraph*, *29*(3), 226–236.

Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin*, *86*, 47–53.

Meijer, E. H., & van Koppen, P. J. (2017). Lie Detectors and the Law: The Use of the Polygraph in Europe. In D. Canter & R. Zukauskiene (Eds.), *Psychology and Law: Bridging the Gap.* (pp. 31–50). Aldershot: Ashgate.

Moszczyński, J. (2011). *Subiektywizm w badaniach kryminalistycznych*. Olsztyn: Wydawnictwo UWM.

National Research Council. (2003). *The Polygraph and Lie Detection*. Washington D.C.: National Academy of Sciences.

Pasko-Porys, W. (2007). Current Legal Status of Polygraph and Level of Practice in Poland. *Polygraph*, *36*(3), 165–176.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological Measures and the Detection of Deception. *Psychological bulletin*, *84*, 782–799. https://doi.org/10.1037/0033-2909.84.4.782

Raskin, D. C., & Honts, C. R. (2002). Handbook of Polygraph Testing. W M. Kleiner (Red.), *Handbook of Polygraph Testing*. San Diego, CA: Academic Press.

Raskin, D. C., & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin*, *86*(1), 54–59. https://doi.org/10.1037/0033-2909.86.1.54

Reid, J. E. (1947). A revised questioning technique in lie-detection tests. *The Journal of Criminal Law and Criminology*, *37*(6), 542–547.

Śmiłowski, E. (2008). Tendencje w technikach badań surveyowych-rozwój CATI, CAPI, CAWI. In K. Mazurek-Łopacińska (Ed.), *Badania Marketingowe. Metody, nowe technologie, obszary aplikacji* (pp. 43–54). Warszawa: PWE.

U. S. Congress. OTA. (1983a). *Review and Analysis of Polygraph Analog Studies. Scientific Validity of Polygraph Testing : A Research Review and Evaluation.* Washington D.C.: U.S. Government Printing Office.

U. S. Congress. OTA. (1983b). *Scientific Validity of Polygraph Testing (OTA-TM-H-15).* Washington, DC: U.S. Government Printing Office.

U.S. Department of Defense. (1984). *The accuracy and utility of polygraph testing.* Washington, DC: Author.

Widacki, J. (1975). The Polygraph in Poland. *Polygraph*, *4*(1), 33–38.

Widacki, J. (2007). Polygraph Examinations in Poland: An Historical Outline. *Polygraph*, *36*(3).

Widacki, J. (2012). Polygraph Examination in Criminal Cases. Current Polish Practice. A Critical Study. *Polygraph*, 6(4), 249–256

Widacki, M. (2013). Najczęściej popełniane błędy w wykonywanych w Polsce ekspertyzach kryminalistycznych z zakresu badań poligraficznych. In M. Gołaszewski (Ed.), Współczesne Standardy Badań Poligraficznych (pp. 101–109). Emów: Agencja Bezpieczeństwa Wewnętrznego..

Appendix A

Judicial districts frequencies

Judicial District	Frequency	Percent	Judicial District	Frequency	Percent
Kaliski	1	.8	Elbląski	3	2.3
Koniński	1	.8	Gdański	3	2.3
Krośnieński	1	.8	Opolski	3	2.3
Legnicki	1	.8	Suwalski	3	2.3
Nowosądecki	1	.8	Tarnobrzeski	3	2.3
Piotrkowski	1	.8	Toruński	3	2.3
Poznański	1	.8	Koszaliński	4	3.1
Przemyski	1	.8	Łomżyński	4	3.1
Siedlecki	1	.8	Olsztyński	4	3.1
Sieradzki	1	.8	Słupski	4	3.1
Świdnicki	1	.8	Gorzowski	5	3.8
Warszawski	1	.8	Krakowski	5	3.8
Włocławski	1	.8	Radomski	5	3.8
Bydgoski	2	1.5	Katowicki	6	4.6
Częstochowski	2	1.5	Płocki	6	4.6
Jeleniogórski	2	1.5	Białostocki	7	5.3
Kielecki	2	1.5	Lubelski	7	5.3
Łódzki	2	1.5	Warszawsko-praski	7	5.3
Ostrołęcki	2	1.5	Gliwicki	8	6.1
Rzeszowski	2	1.5	Tarnowski	9	6.9
Szczeciński	2	1.5			
Bielski	3	2.3	**Total**	**131**	**100.0**

Appendix B

Mean validity ratings of criminal examinations

Method	M	Median	SD	SEM
DNA analysis	9.21	9.00	0.99	0.09
Dactyloscopy	8.82	9.00	1.21	0.11
Balistics	8.14	8.00	1.42	0.12
Mechanoscopy	7.69	8.00	1.45	0.13
Micro-traces	7.56	8.00	1.65	0.14
Biomechanic	7.30	8.00	1.43	0.12
Phonoscopy	7.02	7.00	1.57	0.14
Graphology	7.00	7.00	1.99	0.17
Cheiloscopy	6.97	7.00	2.29	0.20
Traceology	6.54	7.00	1.84	0.16
Geology	6.33	7.00	2.04	0.18
Entomology	5.82	6.00	2.40	0.21
Thermoscopy	5.56	6.00	2.00	0.10
Palinology	5.47	6.00	2.19	0.19
Presentation	5.44	5.00	1.83	0.16
Osmology	5.24	5.00	2.23	0.19
Polygraph	4.24	5.00	2.33	0.20

The Difference Between the Manual and Automatic Settings for the Electrodermal Channel and a Potential Effect on Manual Scoring[1]

Alexander Kalafati[2] and Donald J. Krapohl[3]

Currently the Empirical Scoring System (ESS) is often used by polygraph examiners all over the world. It is scientifically validated[4], it is simple to remember and easy to implement. One of its unique scoring rules is the assignment of +2, 0, or -2 to differences in electrodermal response (EDR) amplitudes while all the other polygraph data channels are scored with only +1, 0, or -1. The weighting of the electrodermal data in this way has been repeatedly shown to improve decision accuracy. The weighting also has another consequence: A mistake in evaluating an EDR can cause a shift of 2 to 4 points. Scoring mistakes in other channels risk only half as many points. Moreover, examiners tend to assign scores to the electrodermal channel more often than they do to other channels, thereby increasing the impact of erroneous EDR scores. For some techniques, such as the mixed-issue Air Force MGQT, a shift of only 4 points in the spot score of a single test question could change a polygraph result of truthfulness to one of deception, or the reverse.

Settings for the Electrodermal Channel

Almost all polygraphs have more than one setting for their electrodermal channel. The two most common are often called "manual" and "automatic" (sometimes "auto-centering"). As will be taken up later in this paper, the instrument processes the electrodermal signal differently for each of these settings, with predictable effects on the resultant display. What has not yet been established is whether the setting for the electrodermal activity (EDA) channel can differentially affect the response amplitudes that examiners score. Said another way, there exists no studies to give confidence that the relative amplitudes of EDRs between relevant and comparison questions are consistent irrespective of whether one chooses the manual or automatic setting. This gap in our knowledge requires attention.

This paper does not directly answer the question regarding EDA settings and scores. Rather, it is an informational article to explain

1 Editorial note: The APA editor would like to thank Thomas Kuczek, Ph.D., Professor of Statistics, Purdue University for his comments, critiques and editorial suggestions to earlier versions of this manuscript.

2 Mr. Kalafati graduated from the Ministry of Defense Military University, specialist in psychology. He retired from the Russian military in 2011. Since then he has been a private polygraph examiner in Moscow, Russia. In 2015 he started development of a new Russian polygraph system called Triumph which is manufactured and marketed in Russia, where he is the head of the development team. Questions and comments to this article can be directed to Mr. Kalafati at psy.akalafati@gmail.com

3 Mr. Krapohl is a Past President of the American Polygraph Association (2006-2007), its former Editor, and regular contributor to this publication. Mr. Krapohl is a long-time polygraph practitioner and has conducted extensive research of polygraph and other credibility assessment technologies. He retired in 2015 as Deputy Director for the National Center for Credibility Assessment and entered private practice as the Director of Educational Services with the Capital Center for Credibility Assessment (C3A). He can be reached at apakrapohl@gmail.com.

4 ESS was conceived by Mark Handler and Raymond Nelson, who credit Donald Krapohl for first showing the potential value of weighted EDA scores, and the US Dept of Defense for first making use of a 3-position scoring method as an alternative to the traditional 7-position system. The ESS first appeared in print in 2008 (Nelson, Krapohl & Handler.) Subsequently there have been more than a dozen published studies regarding ESS with various polygraph techniques. Nelson's 2015 update of the APA's 2011 Meta-analysis is the most recent and complete record of the relevant research citations. Nelson's list of studies regarding ESS can be found in Appendix B of *Fundamentals of Polygraph Practice* (2015) by Krapohl and Shaw.

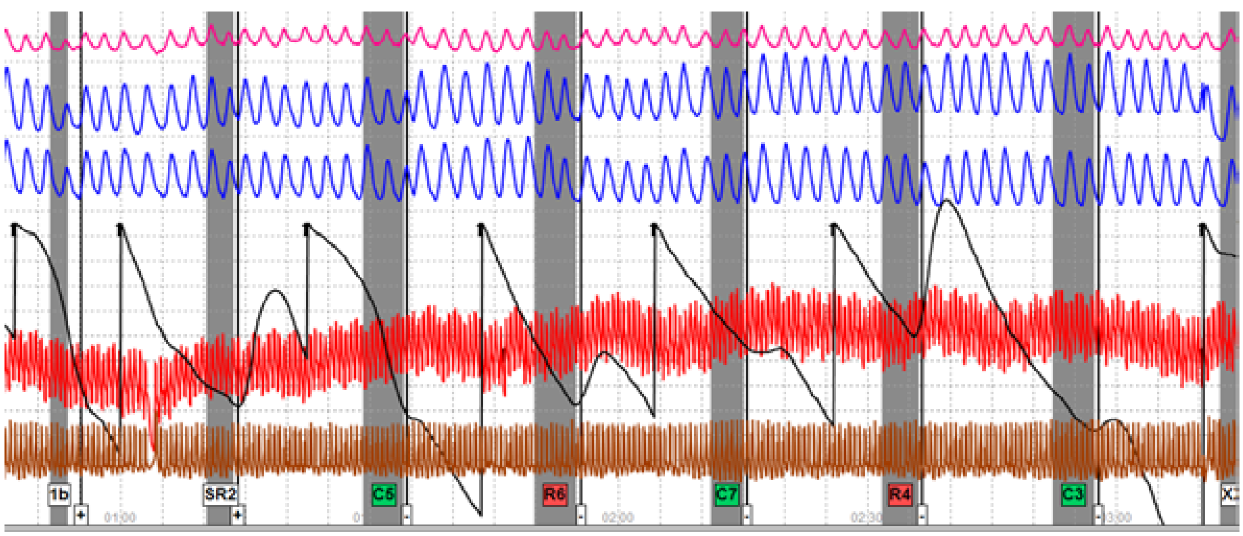
why and how automatic and manual EDA settings differ from one another. Our ultimate goal is to give polygraph practitioners a fundamental understanding of the signal processing behind the EDA tracing so that they may be better informed as they decide which is better for their examinations.

Manual EDA

The manual EDA setting is designed to represent changes in EDA with a minimum of filtering. When the tracing goes up it means that skin resistance is decreasing, and when the tracing goes down skin resistance is increasing. Polygraph manufacturers have long represented EDA data in this manner because humans find it easier to compare "peaks" than "holes" based on our daily experience[5].

The manual EDA setting has a disadvantage with certain examinees, however (see Figure 1). The EDA tracing may rapidly drop or rise due to a continuous shift in the examinee's tonic resistance or conductance, which may require frequent recentering and complicating the task of analyzing the data.

Figure 1. Persistent drop in the manual EDA baseline due to longer-term increases in the examinee's tonic electrical resistance. Actual EDA recorded on a commercial computer polygraph.



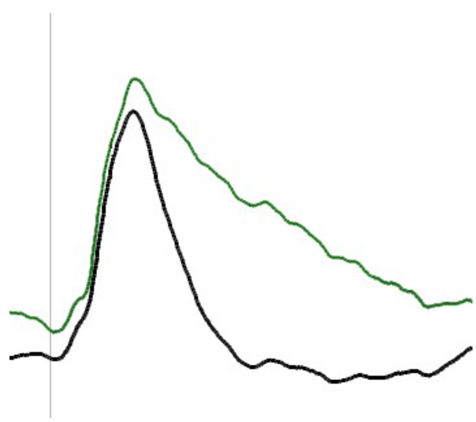
5 Every day we evaluate heights of mountains, buildings, people etc.

Automatic EDA

All commercial polygraphs use filters with EDA data because it removes noise from the signal. Filtering the EDA is not just to make it "pretty." Filtering is necessary to help reveal useful information in the tracing. However, not all manufacturers approach filtering in the same way. At least one commercial polygraph offers no options for its EDA tracing, providing only one highly conditioned signal. Some manufacturers have been completely transparent with their customers as to what their EDA filters do, and others have been far less so. Whether these filters affect manual scoring is an open question, and beyond the scope of this introduction to EDA filtering.

On all computer polygraphs an automatic EDA channel is offered to address the problem of tonic drift. With this setting the pure EDA signal is conditioned using a form of high pass filter (HPF).[6] A HPF allows the passage of fast changes in the signal and stops slow changes. How much data is removed depends on how strong the filter is, from minimal to aggressive. With EDA data the HPF will attenuate a portion of tonic changes of the signal, thereby removing a degree of the signal drift. Slow or constant changes tend to be removed. Figure 2 shows how the automatic setting tends to delete slow changes in the signal.

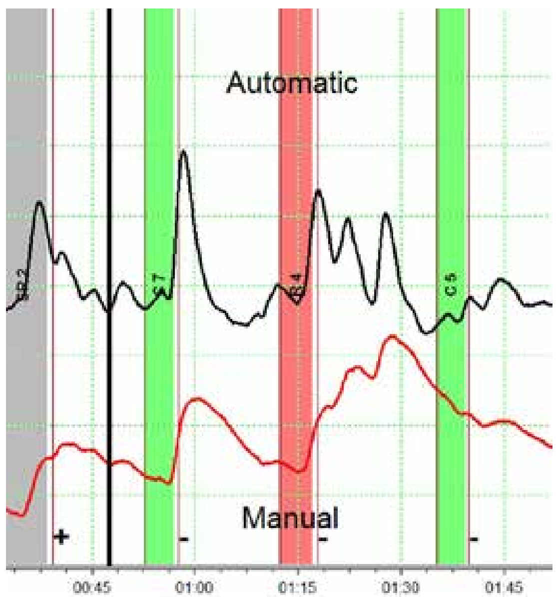
Figure 2. The upper tracing represents EDA data in the manual setting. The lower tracing shows the same data in the automatic setting. Note the differences in the recovery after the response. Actual electrodermal response recorded on a commercial computer polygraph.



6 https://en.wikipedia.org/wiki/High-pass_filter. High pass filtering can be done electronically with a chip or in program software. Nevertheless the principles will be the same.

Figure 3 shows the same EDA signal displayed simultaneously in manual and automatic mode. The upper tracing is the automatic mode and the lower the manual mode. The automatic setting characterizes the complexity differently from the way it is displayed in the manual mode. Moreover, because there is a different rate of change in the manual EDA tracing between the phasic reaction at C7 and R4, the relative amplitudes of the tracings have been affected: In the manual mode the bigger reaction is on R4 but in the automatic mode it is at C7.

Figure 3. Comparison of a complex reactions with simple ones.



The shift in relative EDR amplitudes happens because of the nature of the HPF. The HPF works similarly to the mathematical function called a derivative. We will use the concept of the derivative to explain the principles of how the HPF functions, but we caution readers that our purpose here is only illustrative and that the HPF and the derivative are not identical concepts.

We can implement the self-centering function with help of HPF or computer algorithms or even mechanically, nevertheless it will work the same way. In geometry a derivative represents the speed of change of the signal[7] (Crowell, 2017; Kudryavtsev, 1981). If the change is slow or constant the derivative value is near 0 and represented on the graph as a flat line. If the speed of the reaction is rapid the value of the derivative is greater than 0, and the tracing will rise. It is possible to measure the speed of the reaction in manual EDA. We measure the angle of the reaction, as it is shown in Figure 4.

In Figure 4, points A, B, and C denote

7 https://en.wikipedia.org/wiki/Derivative

places where a change in the angle of the reaction takes place. From the start of the reaction until point A the angle of the reaction is approximately 45° and the automatic EDA also starts to rise, but from point A to point B the angle lowers to approximately 30° but it is still rising, while the automatic EDA starts to fall. This occurred because the speed of the reaction starts to fade. From point B to point C the reaction changes its speed for the third time. In this segment the manual tracing starts to rise with an angle of approximately 60° which causes the rise in the tracing of the automatic EDA. The automatic tracing lost the rise in the manual tracing from point A to point B. If the angle of the reaction in the manual tracing from point B to point C equaled 45°, the amplitude of the automatic EDA would be at the same level as it was at point A, and the rise in the manual tracing between points A and C would have been lost entirely in the automatic mode. Conceptually, this is what took place between the manual and automatic tracings in Figures 3 and 5.

Figure 4. Simplified illustration of manual (upper tracing) and automatic (lower tracing) modes for the same EDA data using a commercial polygraph available in Russia. Note how changes in the rate of rise for the manual tracing is recharacterized in the automatic tracing.

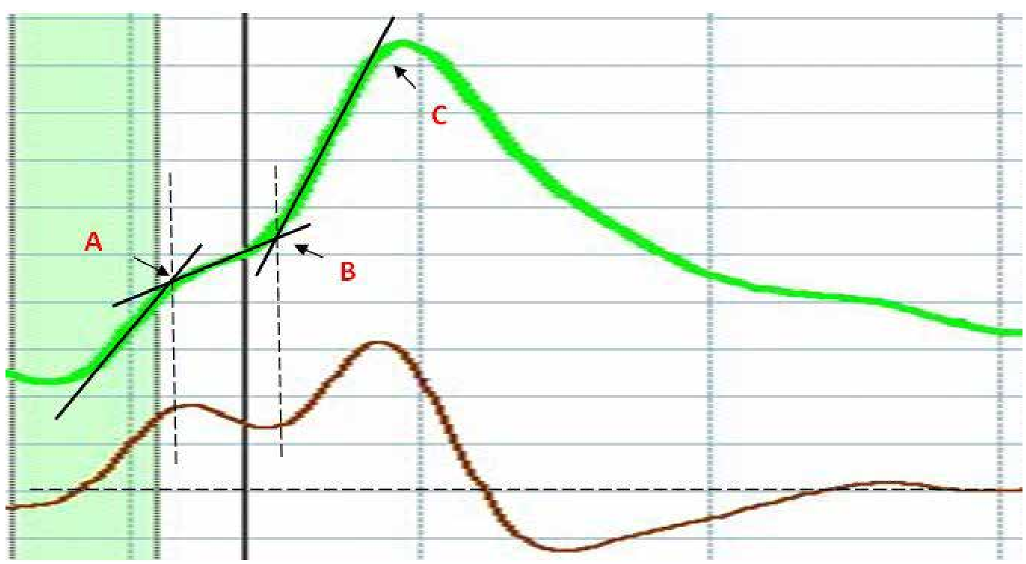
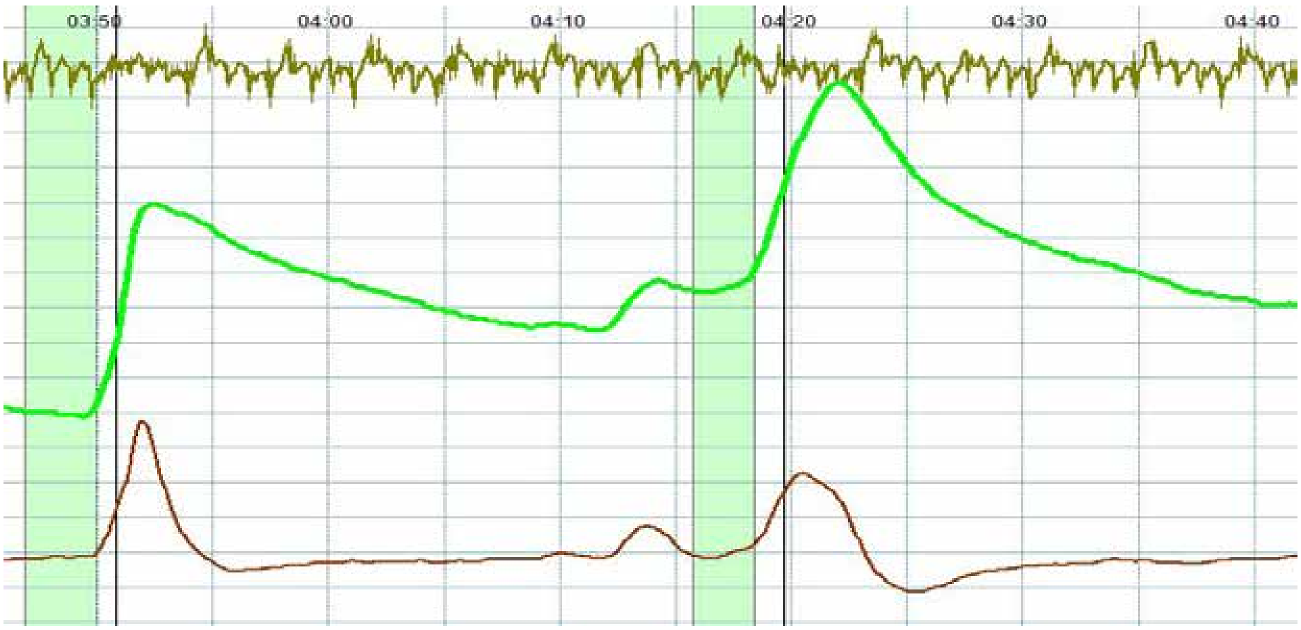


Figure 5. Two responses in the manual mode (upper EDA tracing) with the same amplitude but different rates of change. Note that in the automatic mode (lower EDA tracing) the two responses no longer have the same amplitudes. The topmost tracing is for the motion sensor.



As we mentioned previously, the automatic EDA mode was designed to re-center the tracing without manual intervention from the user. As a consequence, a portion of the reaction duration is lost (see Figures 2-5). This may be why Kircher, Kristjansson, Gardner, and Webb (2004) failed to find any significant value in the EDA reaction duration, while the Kircher and Raskin (1988) study results showed it as a valuable feature. The former study used charts with automatic EDA and the latter study used the manual EDA mode.

Finally, we cannot use the information in the automatic mode to transform the signal back to the manual mode. The HPF irretrievably filters out some information. Certain HPFs can also create features which have nothing to do with physiology, such as when the EDA tracing falls below the baseline. Because the electrodermal response has no parasympathetic component, the drop below baseline following an electrodermal response is considered an artifact created by the filter.

Since the 1960s most EDA scoring systems have used the ratio of the amplitude of electrodermal responses to relevant and comparison questions for assignment of scores to the EDA channel. Those ratios do not change according to the type of EDA filtering a manufacturer offers in its computer polygraphs. There has been no published research that offers guidance on whether there are meaningful differences in EDA scores between the two modes of display. In 2015 the second author of this article conducted an unpublished government study to look at score assignment in the EDA for manual and automatic modes for the two most popular computer polygraphs. It was found that the EDA scores differed 10%-15% per case between the automatic and manual modes. None of the differences contributed to decision errors, however.

Conclusion

Our understanding of the automatic EDA filtering suggests:

- The automatic EDA mode is a secondary, filtered channel of the manual EDA. It is very sensitive to changes though it does not always represent the entire phasic response.

- The automatic EDA does not show slow tonic changes in skin resistance activity.

- The automatic EDA does not show the real length of the reaction.

- The manual EDA tracing shows *changes* in skin resistance while the automatic EDA tracing represents the *speed* of changes in skin resistance. In other words, automatic EDA shows the *speed of phasic changes*, while manual EDA shows the *phasic and tonic changes* of the EDA.

- Because ESS is intentionally designed to weight the EDA channel, ESS scores may be especially vulnerable to larger shifts in EDA scores between the automatic and manual mode than are other scoring systems. This question calls for more research.

Final Thoughts

The aim of this article is to show, in principle, how EDA data appear differently when changing from manual to automatic modes. The central point is that the EDA tracing in the automatic mode is not the same as the tracing from the less-filtered manual mode. The two channels do not behave exactly the same way, and depending on the speed of the phasic response, the choice of EDA mode can alter manual scores. The second author has been cataloging the incidence of opposite scores between EDA modes he encounters in conducting quality control reviews of field cases. While there have been no cases where an opposite score between EDA modes has caused an opposite polygraph decision, the choice of EDA modes has made the difference between inconclusive and non-inconclusive results on rare occasions. In his ongoing data collection project, EDA scores have been affected by the selected EDA mode in about one in eight cases. A full report will be the subject of a future paper.

As noted in the earlier sections of this paper, some polygraph manufacturers have been very open about the manner in which they filter the EDA data on their computer polygraphs, and they are to be congratulated. It warrants comment, though, that no published research by any researcher or research organization has considered the impact of changes in EDA filtering on scoring systems. Ever. Filtering is surely different between analog and digital polygraphs, between the various makers of computer polygraphs, and sometimes between models and even software updates from the same manufacturer. This may not be a serious problem, but no one yet knows.

We do not take lightly the larger implications attendant to the statements in our paper: Scoring and algorithmic approaches based on data from one EDA mode may not generalize perfectly if another EDA mode was used during a given examination. Whether the effect is significant or trivial is not yet known for any of the computer polygraphs, and we do not speculate further on the question. We do risk the suggestion, however, that examiners may wish to use the less-filtered manual mode for hand scoring whenever possible until published research offers a scientific basis for choosing an EDA mode with more filtering.

Most polygraph examiners do not typically think about how filtering affects the data upon which they rely to perform their services for the courts, law enforcement, national security, and the general public. Indeed, the polygraph literature infers there is a great trust in the polygraph manufacturers for ensuring the fidelity of our instrumentation. Everyone should agree it is important to have confidence that the tracings we score truly represent the examinee's physiological activity. In the case of the EDA, the variety of filtering approaches across and within instrument manufacturers suggests no one yet knows what the best filtering method should be. This is unfortunate. The authors view data filtering not just as an engineering question, but one of public trust. What would be the best way(s) to filter the EDA to maximize scoring accuracy? When that question is answered it seems reasonable that the profession move to such a filter. Until then, the authors propose that our current state of knowledge points to a filtering solution involving the least amount necessary to reduce noise without distorting or removing diagnostic information.

References

Crowell Benjamin (2017), *Fundamentals of Calculus*. Fullerton, California. Open source book available at http://www.lightandmatter.com/.

Kudryavtsev L.D. (1981). Course of the mathematical analysis. Moscow, 687.

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

Kircher, J.C. and Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*(2), 291-302.

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

Lafayette Instrument Company Response to Kalafati and Krapohl

Raymond Nelson, Research Specialist

Brent Smitley, Engineering Manager

Electrodermal activity (EDA) is a complex phenomenon that is of great interest to polygraph examiners and the polygraph profession. For this reason, there is some degree of both satisfaction and caution surrounding this opportunity to respond to the Kalafati and Krapohl (2018) publication. A sense of caution is useful here mainly due to the risk of misunderstanding and attendant risk of conflict among individuals who undoubtedly deserve respect as both friends and colleagues. With this reservation, this response is intended to clarify some details and to provide additional important information on the matter of EDA signal processing. Hopefully this additional information, together with other publications, will be of use to field practitioners, trainers, researchers and others involved in the formulation of standards and guidelines for polygraph field practice, and for those interested in understanding the science and engineering aspects of EDA as it applies to the polygraph test.

There are two main areas of discussion that are necessary in response to the Kalafati and Krapohl publication. The first is the stated purpose of the article, which is to convey information on how and why manual EDA scoring and automatic EDA scoring differ from each other. The second, larger issue is how these differences impact the outcome of a polygraph exam. Kalafati and Krapohl state that the article is not intended to directly address the EDA settings and scoring, but the implications to the scoring cannot be avoided. The article begins with comments on the ESS scoring algorithm and ends with a recommendation for a preferred EDA mode based on the potential for scoring differences, therefore the scoring aspect is inescapable and needs further discussion.

With regard to the technical discussion on how and why differences occur between manual and automatic EDA modes, the information presented in the article has a degree of validity but is incomplete. High pass filters are typically described in terms of filter cutoffs and frequency response. A filter, by definition, attenuates certain frequencies in a particular waveform while allowing other frequencies to pass through. By attenuating very low frequencies, the filter produces a “return to baseline” feature in the EDA trace. It is this return to baseline function that keeps the automatic trace centered and removes the need to constantly manage a manual EDA trace that has an unstable tonic baseline.

A satisfactory discussion of EDA filtering should help a polygraph user understand the component frequencies that constitute the EDA measurement, how the filters affect those frequencies, and how the attenuation of those frequencies ultimately affects the trace that is viewed on the screen. The notion that a high pass filter is similar to a derivative tends to move the discussion away from a correct understanding of EDA filtering and into an area that may be poorly understood. Even though it is emphasized that a high pass filter and a derivative are not equivalent, introducing the notion that filtering can be understood in terms of a derivative, excludes a robust discussion of what the filtering is actually doing.

A second technical aspect that needs to be noted is the automatic trace filter specifications. Kalafati and Krapohl state several times that filter specifications differ from system to system, with some systems having known specifications and some being unknown. However, a subjective assessment of the Kalafati and Krapohl article is that this issue is given *secondary* consideration in the discussion.

Understanding how the filter specifications affect the automatic trace is of *primary* importance and is critical to a correct understanding of how and why manual EDA is different from automatic EDA. The filter cutoffs and frequency response will have a direct effect on the return to baseline time, the reaction duration, the amplitude attenuation and how the filter processes slope changes in the manual trace. It is possible to configure a filter that has measurably improved performance over the filter used to illustrate the EDA mode differences in the article. Kalafati and Krapohl's assertion that scoring differences will be possible regardless of filter settings will still be true, but optimizing the filter performance will change the context of the discussion and will minimize those differences. The filter specifications are something that should be detailed in a robust discussion on EDA modes and not something that should be minimized.

Beyond the technical issues that have been noted, the larger, more important issue is that of scoring. Ultimately, if the two different EDA modes did not, on occasion, result in different scores, there would be no need for this discussion. It is in this area that a reasoned challenge can be made to the underlying premise that comes through in the Kalafati and Krapohl article. The article starts with the notion of greater potential for "scoring mistakes" if the EDA is affected by the high pass filter and the ESS scoring algorithm is used. Near the end, it also poses the question if any manufacturers know the best filtering configuration to "maximize scoring accuracy". In between is the implication that manually–centered EDA is correct or more correct and where the manual and automatic modes differ, the automatic mode is less correct or less accurate. The underlying premise is that there is an objective accurate score for any given EDA trace and the mode that best matches that objective score is the most accurate. This simplistic thinking must be rejected. In the discussion of differences between manual and automatic modes, the scores for each mode are not being compared to some objective "accurate" score; rather two subjective scores are being compared to each other.

The overarching purpose, or the criterion, for polygraph is to discern deception from truth telling. However, any single EDA score, whether produced via manually-centered EDA or auto-centered EDA is an insufficient basis for reliable and accurate conclusions about deception and truth-telling. Both EDA modes can be expected to produce some scores that are inconsistent with the criterion of deception or truth-telling, along with a usable volume of scores that are consistent with differences between deception and truth-telling and sufficient to achieve a strong and usable diagnostic coefficient. This is because there is inherent variability in EDA data and EDA scores. To put it another way, neither manually processed EDA nor automatic EDA correlates perfectly to deception; both are an approximation. Similarly, EDA scores alone, without the scores from the other recording sensors, are an insufficient basis of information for a scientific credibility assessment test. However, multiple EDA scores combined effectively with scores from other recording sensors have proven to be a reliable and effective basis for scientific conclusions about deception and truth-telling. A general purpose of scientific tests is to quantify things that cannot be subject to physical measurement, and for this reason scientific tests are not expected to be infallible. All test data is a proxy or substitute for a phenomenon that we wish to measure but cannot physically measure – in this case deception and truth-telling.

The question then emerges of which EDA signal processing mode will maximize the diagnostic value of the EDA scores and correlate closer to the criterion of interest, which is the discernment of truth and deception. In other words, which EDA mode will provide better reliability and validity and will be correct more often and incorrect less often for most people most of the time?

All measurements contain both a signal of interest and noise. The EDA measurement is no exception. The goal of any filtering solution is to remove as much noise or useless information as possible while preserving as much diagnostic information as possible. The idea of filtering immediately destroys any premise that two waveforms with different filters can be identical to each other. There will inevitably be differences. At the same time, it opens the possibility that a filtered solution can be more useful if the filter is optimized to remove detrimental information while preserving all of the diagnostic information.

The Kalafati and Krapohl article does an adequate job of demonstrating that scoring differences are possible, but the conclusion that manually scored EDA is preferable is based on incomplete analysis. The main unanswered question is how often is the automatic processing solution preferable and more useful than the manually scored solution? It is well known that using a manual EDA mode requires managing of the trace which takes attention away from other aspects of the polygraph exam. It is also well known that manual EDA with a strong tonic baseline instability is difficult to score. Additionally, analysis has shown that tonic changes in the EDA waveform have no diagnostic value and can be removed without affecting the ability of the exam to achieve the overarching purpose of discerning truth and deception.

The question of whether EDA scores concur with the criterion state of deception or truth-telling is an empirical matter for which the selection of a signal processing solution will be best guided by studying the correlation of EDA scores with confirmed case data. This will depend in part on the feature extraction method, the details of how a decision model will make use of the EDA scores combined with the other sources of recorded data and the signal processing method. Early analysis using this approach suggests that a properly optimized automatic mode has a stronger correlation to truth and deception than does the manual mode. Speculatively, the reason for this is that a well-designed automatic filter does a good job of preserving the diagnostic data and that the tonic data that is removed is not only useless to the diagnostic value of the waveform, but detrimental to it.

The question of concern to polygraph examiners is this: which EDA mode has the greater correlation with deception and truth-telling? It is easy to agree with the statements of Kalafati and Krapohl regarding the importance of the fidelity of our polygraph instrumentation and the need for both public and professional trust in the design specifications of instruments in use in field practice settings. Our analysis has shown that while some auto-centering EDA solutions will under perform relative to manually-centered EDA data, it is also possible to design auto-centering EDA signal processing models that have more optimal signal to noise ratios that outperform the manually-centered EDA. Our confidence on this matter is such that we already have taken the important steps of documenting and publishing the design specifications and performance characteristics of our EDA signal processing methods – putting these models into context in the published scientific literature.

It is our position that any field practice recommendations or policy decisions about which EDA signal processing methods are better should be based on evidence from studying those models. We at Lafayette Instrument Company are confident in our analysis of these issues, and we are confident in asserting that the EDA signal processing solutions included in our instruments and software products are the optimal solutions that can be made available at this time.

For more information, please email info@lafayetteinstrument.com.

Limestone Technologies Response to APA EDA Publication

In the recent APA publication by Kalafati and Krapohl titled "The Difference Between the Manual and Automatic Settings for the Electrodermal Channel and a Potential Effect on Manual Scoring", a gap in knowledge has been identified as to the unknown impact different settings for electrodermal activity (i.e. manual versus automatic) has on response amplitudes that examiners score.

Automatic EDA historically has been the setting of choice among examiners and researchers as the automatic EDA allows examiners to conveniently observe an examinee's immediate EDA response to stimuli from a known baseline. Unlike manual EDA, the automatic EDA mode best meets this objective as it maintains the consistent baselines which are essential to quality polygraph. However, scoring variances between examiners that switch between manual and automatic EDA have been widely observed among various polygraph instrumentation. These variances - while not surprising considering the different input conditioning between the two - have consequently introduced an area of discussion as to whether examiners should refer to the automatic or manual EDA data.

At Limestone Technologies we see the value in both measurements. Automatic EDA solves the problem of tonic drift, making the data easier to interpret. On the other hand, manual EDA more closely reflects all the gathered data (including the tonic drift). It is for these reasons why our polygraph software, by default, always collects and stores the manual EDA data. An examiner using our instrumentation always has access to the manual EDA information, regardless of the setting that was used. This is done for ongoing research purposes and to provide examiners the opportunity to compare both the automatic and manual EDA. It is the recommendation of Limestone Technologies that our clients make comparisons post exam so that differences can be immediately identified and accounted for.

When a variance is observed the question remains as to whether the automatic or manual data should be used for the reading. As the article stated, there is not sufficient evidence to provide a clear path forward. Rather, the decision should be based on the examiners/administrations comfort and experience interpreting automatic or raw data.

Limestone Technologies Inc.

Stoelting Co. Response to Article, "The Difference Between the Manual and Automatic Settings for the Electrodermal Channel and a Potential Effect on Manual Scoring"

Michael A. Cochran and Ricardo C. Fuentes*

Introduction

The assertions and cautions of the article written by Kalafati & Krapohl (2018) for the American Polygraph Association's polygraph journal do not apply to Stoelting's computerized polygraph systems, to include our CPSpro Polygraph System. More specifically, Kalafati & Krapohl stated in their article that their aim was "to show the possibility of opposite EDA scores when changing from manual to automatic modes" (p.7). Although this may be true for other polygraph equipment manufacturers, this does not and cannot occur using a Stoelting computerized polygraph system.

The reason is simple. Stoelting always uses the original, unaltered, raw data obtained during the chart recording process.

Use of Filters, Modes, and Other Software Options

With Stoelting computerized polygraph systems, there is nothing the computer can do to alter the data that is collected during chart recording, regardless of settings set by the examiner. There are no modes, filters, or other options that will affect what is collected and saved in the Stoelting's polygraph system software. When chart recordings are redisplayed, after collecting a chart, the system uses only the original, unaltered, raw data; regardless of how it was displayed for viewing convenience during or after chart recordings.

Stoelting's current polygraph system (CPSpro) is truly a computerized polygraph system; not a computerized analog polygraph system.

This cannot be over emphasized; when using optional user-preference settings such as electrodermal (ED) filter, auto-center, detrend, etc., in the CPSpro Fusion polygraph software, the original data remains unchanged. While these options provide a means to make data collection easier and more convenient for examiners, they are only a visual representation of the raw data. All analyses are based on the original raw, unfiltered data.

Electrodermal Filtering

When the ED filter option is enabled, the CPSpro Fusion software implements a high pass filter which helps stabilize the baseline for viewing by the examiner in the chart recording mode. It does not affect the original data that is always recorded and stored in the disabled mode. The ED filtering allows the CPSpro Fusion software to display more stable tracings while recording charts with the filter enabled, but, by default, the software only displays the unfiltered (raw) data in chart redisplay mode. The filter has no effect on any of

* Authors' Note:

Michael A. Cochran is the Chief Executive Officer and Chief Financial Officer of Stoelting Co.

Ricardo C. Fuentes is the Polygraph Division Director of Stoelting Co.

Comments concerning this article should be addressed to Ricardo Fuentes, Polygraph Division Director, 620 Wheat Lane, Wood Dale, IL 60191. Contact: ricardo@stoeltingco.com or (630) 708-7659.

the calculations used for displaying response strength or computer analysis algorithms.

Even though the ED filter does not affect the outcomes of manual scoring methods conducted by the examiner, Stoelting recommends polygraph charts not be scored while they are being recorded.

Demonstration of Effects of Chart Recording and Redisplay Modes

To further demonstrate how the ED filter does not affect (when enabled) the Stoelting computerized polygraph systems in chart redisplay mode, two chart images are illustrated in chart recording mode and redisplay mode. In Figure 1, a chart recording is displayed with the ED filter enabled. A reaction was caused when a 100K resistor was attached to the ED channel during the chart recording mode. This caused an accurate, constant 10 micro Siemen increase in the tracing. With this change, the channel tracing would be expected to remain flat at the peak reaction, since it is a static change; however, the ED filter gradually allowed the tracing to return to baseline, even though the 10 micro Siemen value remained unchanged.[1]

Chart Recording Mode - Figure 1

Figure 1. While enabled, the electrodermal filter allowed tracing to return to baseline during reaction in chart recording mode.

In Figure 2, as expected, in the redisplayed mode, the CPSpro Fusion software displayed (in raw data) an increase in reaction while the 100K resistor was attached. The reaction remained stable and constant and then returned sharply to baseline after the 100K resistor was removed.[2]

1 This feature allows the CPSpro Fusion software feature minimizes the adjustments needed by the examiner during the chart recording mode.

2 This Stoelting feature allows the CPSpro Fusion software to display all tonic changes, in raw form, noted during the entire chart recording mode.

Chart Redisplay Mode – Figure 2

Figure 2. The ED tracing (in redisplay mode) indicated a constant, stable reaction until the 100K resistor was removed.

Conclusions

Stoelting's conclusions to our response to the article:

- There is nothing the examiner (using a Stoelting CPSpro system) can do to alter the data that is collected during chart recording, regardless of preferences set in the CPSpro software.
- When polygraph charts are redisplayed, the CPSpro software uses only unaltered, raw data obtained in the chart recording mode.
- The ED filtering allows the CPSpro Fusion software to display more stable tracings while recording charts with the filter enabled.
- The ED filter has no effect on any of the calculations used for displaying response strength or computer analysis algorithms.

All features listed above have been incorporated in Stoelting's computerized polygraph systems and will continue to be in all future systems.

References

Kalafati, A. & Krapohl, D. J (2018). The Difference Between the Manual and Automatic Settings for the Electrodermal Channel and a Potential Effect on Manual Scoring. *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice*

Electrodermal Signal Processing:
A Correlation Study of Auto-Centered EDA and Manually-Centered EDA with the Criterion State of Deception and Truth-telling

Raymond Nelson[1]

Abstract

Effectiveness of different EDA signal processing solutions was studied using a sample of N=36 confirmed criminal investigation polygraphs, consisting of 18 truthful and 18 deceptive cases. Examples of EDA data are shown in the time-series domain and after Fourier transformation to the frequency domain. Criterion correlation coefficients of logged relevant/comparison polygraph scores using optimized auto-centered EDA filter specifications are shown to equal or exceed that of manually-centered EDA data. A second sample of N=47 charts from unconfirmed polygraph examinations was used to evaluate the rank correlations of responses extracted from manually-centered EDA data with the rank scores of several auto-centered EDA solutions that are described in extant publications.

Introduction

Electrodermal Activity (EDA) is a general term that subsumes a variety of electrical phenomena in the skin (Johnson and Lubin, 1966). The older term galvanic-skin-response (GSR) is now thought to be outdated and inadequate (Boucsein, 2012) because galvanic processes, involving the production of electrical current as a function of chemical reaction, has little, if anything, to do with EDA reactions during polygraph testing.

EDA has been discussed at length in the psychophysiology literature (Boucsein, 2012; Fowles, 1986; and Venables & Christie, 1973). EDA during polygraph testing is typically recorded using exosomatic methods involving the introduction of a small direct current (DC) to the skin. Current density is limited by convention to a maximum of 10 micro-amperes per square centimeter (10uA/cm2; Boucsein, 2012, Edelberg, 1967). Ohms' law (Ohm, 1827) states that Voltage (V) equals Current (I) times Resistance (R), represented mathematically as $V = I * R$. If the voltage is held constant, then changes in EDA can be observed and recorded as changes in current. Conversely, if the current is held constant then changes in EDA can be recorded as changes in voltage.

Both skin resistance and skin conductance circuits have been described for use in field polygraph instruments (Boucsein & Hoffman, 1979; Honts & Barger, 1990; Lykken & Venables, 1971), and both SR and SC can be calculated from both constant-voltage and constant current circuit designs. Boucsein (2012) was unconvinced that either resistance

1 Raymond Nelson is a polygraph examiner and psychotherapist and has published numerous research studies and instructional articles on all aspects of the polygraph. He is one of the developers of the OSS-3 and ESS scoring algorithms. Mr. Nelson is an elected member of the APA Board of Directors and is a past-President of the APA. Mr. Nelson has contributed to numerous areas of polygraph field practice policy, and lectures often at national and international conferences and accredited polygraph training programs world-wide and throughout the US. Mr. Nelson is a research specialist for Lafayette Instrument Company, a company that manufactures polygraph software and instrumentation and has proprietary interest in some of the technologies discussed in this manuscript.

or conductance units are superior – as their relationship is mathematically defined by Ohm's law such that if one value is known the other is also known. More importantly, discussions of the unit of measure should not be confused with discussions of the method of recording. Equally important is that all EDA data – and virtually all data – requires signal processing to be of any practical or analytic value.

All data is a combination of signal and noise. Ideally the amount of useful signal is great, and the amount of noise is small. One of the goals of effective signal processing is to optimize the signal to noise ratio. EDA signal processing in computerized polygraph systems begins with the analog-to-digital conversion of the information captured by the sensor and can include a combination of hardware and software filters. Practically all electronic devices in existence today, if they are powered by AC current, will include some form of filtering to remove unwanted 50Hz or 60Hz noise resulting from the use of AC current sources in residential and commercial buildings. Other filters can be used to condition the EDA signal for usability, display, and feature extraction.

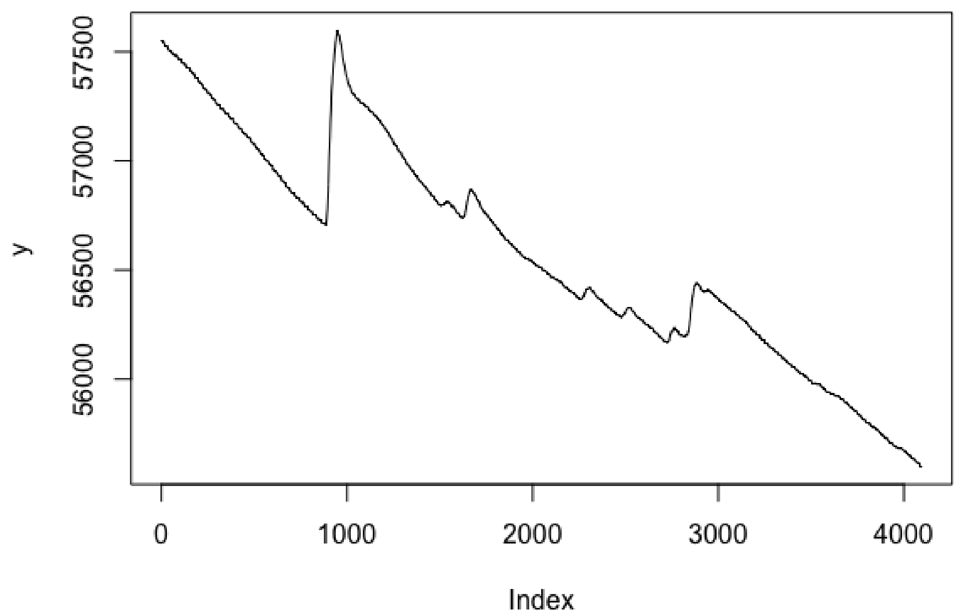
EDA filters can be implemented in hardware or software with the same result, though there are advantages to software solutions when upgraded filter designs become available. Another advantage of software filter solutions is that digitized signals can be stored in a pre-filtered state for which improved signal processing models can later be implemented. In contrast, hardware filters are likely to result in recorded data that is permanently fixed to the signal processing method at the time of recording.

Regardless of whether implemented in hardware or software, most EDA signal processing methods involve passive filters, though active filters – created with amplifiers – can also be used. EDA data can be thought of as consisting of a range or spectrum of frequencies, including lower frequency tonic activity and comparatively higher frequency phasic responses. Both tonic and phasic EDA are continuously changing in response to the examinee's internal – psychological and physiological – response to external or environmental stimuli.

Environmental stimuli during polygraph testing are reduced to a minimum so that the test stimuli can maximally influence the examinee's recorded physiological data. Because physiological signals are inherently noisy, a sufficient volume of data must be obtained for analysis. Because polygraph tests – and scientific tests of all types – are used to quantify phenomena that cannot be measured physically, analysis consists of numerically and statistically quantifying the probabilistic values associated with categorical conclusions. The analytic theory of the polygraph holds that greater changes in physiological activity are loaded at different types of test stimuli as a function of deception or truth-telling in response to the investigation target stimuli. Interpretation of comparison question polygraph data involves only the use of phasic EDA responses. A known phenomenon is that some individuals will, for a variety of reasons, have unstable tonic EDA that can substantially increase the difficulty of extracting and interpreting phasic EDA responses. For this reason, EDA filtering options have been included in field and laboratory polygraph systems for several decades.

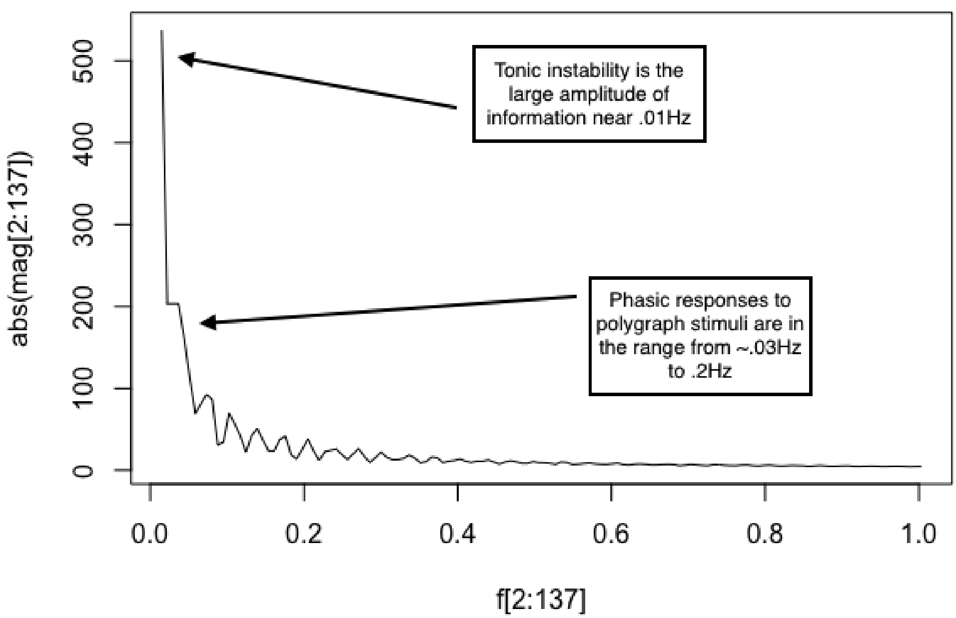
EDA data is commonly displayed in the time-series domain. The time-series domain is plotted graphically so that changes in physiology are shown vertically on the geometric Y-axis. The time-period of data acquisition and recording is plotted on the X-axis so that the series of changes in physiology can be visualized. Figure 1 shows a segment of unstable EDA data in the time-series domain. Phasic EDA responses can be observed as the upward segments that occurred in response to polygraph stimulus questions.

Figure 1. EDA data shown in the time-series domain.



EDA data can also be transformed mathematically, using the Fourier transform (Brigham, 2002 ; Cooley, 1987), to the frequency domain. Frequency-domain plots permit visualization of the relative strength of the different frequencies across the spectrum of frequencies contained in the recorded data. Frequency domain plots show the spectrum of frequencies on the X-axis, with the relative strength of those frequencies shown on the Y-axis. Figure 2 shows the data from Figure 1 after Fourier transformation. Because Fourier transformation does not itself involve the filtering or amplification of information, there is no loss of information when data are transformed to the frequency domain. Data in the frequency domain can be transformed mathematically back to the time-series domain, using an inverse Fourier transform, without loss of information.

Figure 2. EDA data from Figure 1, shown in the frequency domain after Fourier transformation.



Data shown in Figures 1 and 2 exhibit the characteristic tonic instability that led to the introduction of high-pass filtered and auto-centering EDA modes. Tonic instability can be seen in Figure 1 as the descending trend in the EDA data throughout the recorded time period. Tonic instability can be seen in Figure 2 in the relatively large amplitude of frequencies in the range near .01Hz. Tonic instability is a well-known phenomenon in the psychophysiology literature. Phasic responses to polygraph stimulus questions can be observed in Figure 2 as increased amplitudes of the frequencies from ~.03Hz to .2Hz, compared to the amplitudes of frequencies over .4Hz

Given the choice of either manually-centered EDA or auto-centered EDA modes, Krapohl and Shaw (2015) offered a recommendation to use the manually-centered EDA mode. Raskin, Honts and Kircher (2014) also suggested the use of the manually-centered EDA mode, though they also wrote that a high-pass filter with a time constant of 10 seconds would not damage the information of interest to polygraph examiners. The time

2 One time constant is the time it takes for data to return 63.2% of the distance to baseline. So, 1 - .632 = .368 and .368^5 = .007. In five times the time constant the data will return 99.3% of the distance to baseline, which is essentially a complete return to baseline.

constant refers to the length of time that the data can be characteristically expected to return 63.2% of the distance to the onset value. By convention, when using an auto-centering filter of this type the data will have returned completely to the response onset or baseline level in five time constants[2], so that a time constant of 10 seconds will result in a signal that returns to baseline within 50 seconds. A time-constant of 10 seconds calculates mathematically to a high-pass corner frequency of 0.0159Hz. Figure 3 shows the same data after processing the signal with an auto-centering (high-pass) EDA mode using a first-order Butterworth-type (Butterworth, 1930) high-pass filter with a corner frequency of 0.0159Hz, following the recommendation of Raskin, Honts and Kircher (2014). Butterworth-type filters can be implemented in hardware and can also be implemented using digital signal processing (DSP), as was done in this project.

Figure 3. EDA data using a high-pass auto-centering EDA filter with a corner frequency of 0.0159Hz.

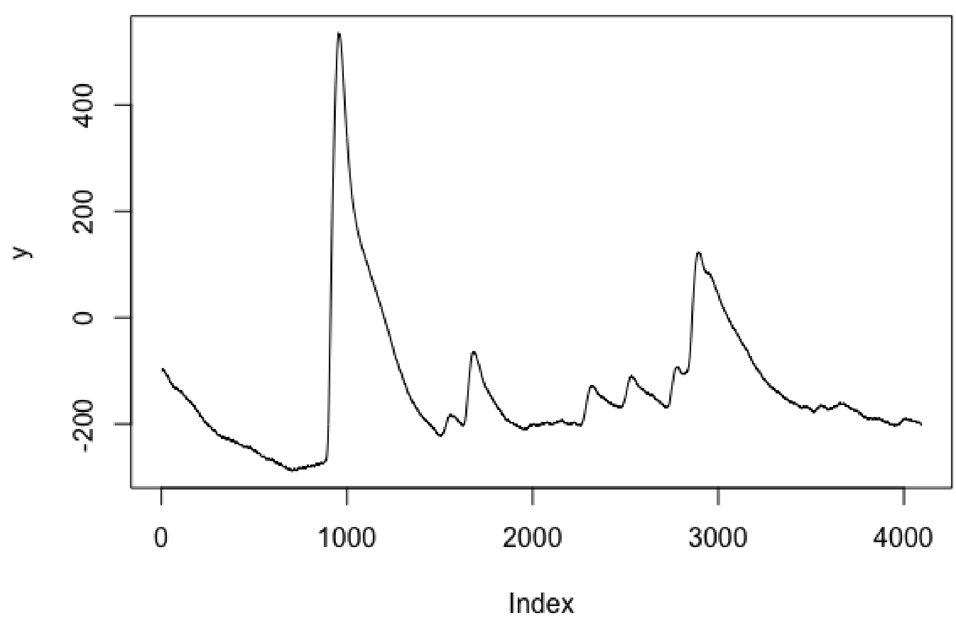
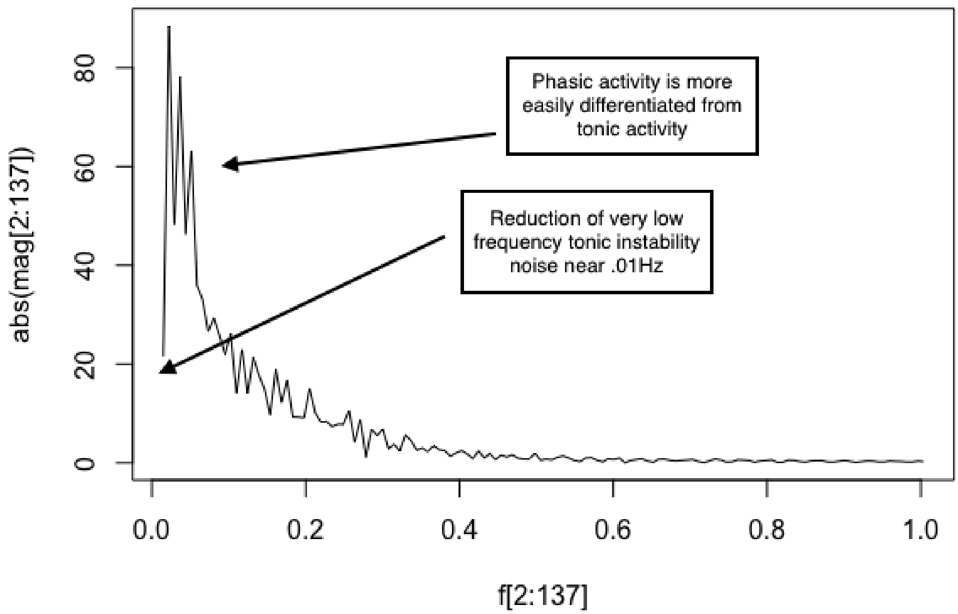


Figure 4. Shows the auto-centered EDA data in the frequency domain.



Method

Diagnostic coefficients

Diagnostic effectiveness of different EDA signal processing models was evaluated using data from confirmed field examinations. Data consisted of N=36 cases, including n=18 innocent cases and n=18 confirmed guilty cases. Confirmation of guilt or innocence was established by a combination of confession and extra-polygraphic evidence. Demographic information, including age, gender, education, occupation, ethnicity and health were not provided. All cases were conducted using an event-specific comparison question test (CQT) format with two relevant questions and a question sequence repeated three times. Examination targets included theft, assault, forcible sexual offense, burglary and illegal drugs. All exams were conducted using the Lafayette LX4000 polygraph system for which recorded data are stored in an un-filtered digital format that is amenable to the study of different filter solutions. All cases included an acquaintance test that was not included in the analysis. Data were inspected for compliance with correct examination protocols and were subject to automated feature extraction.

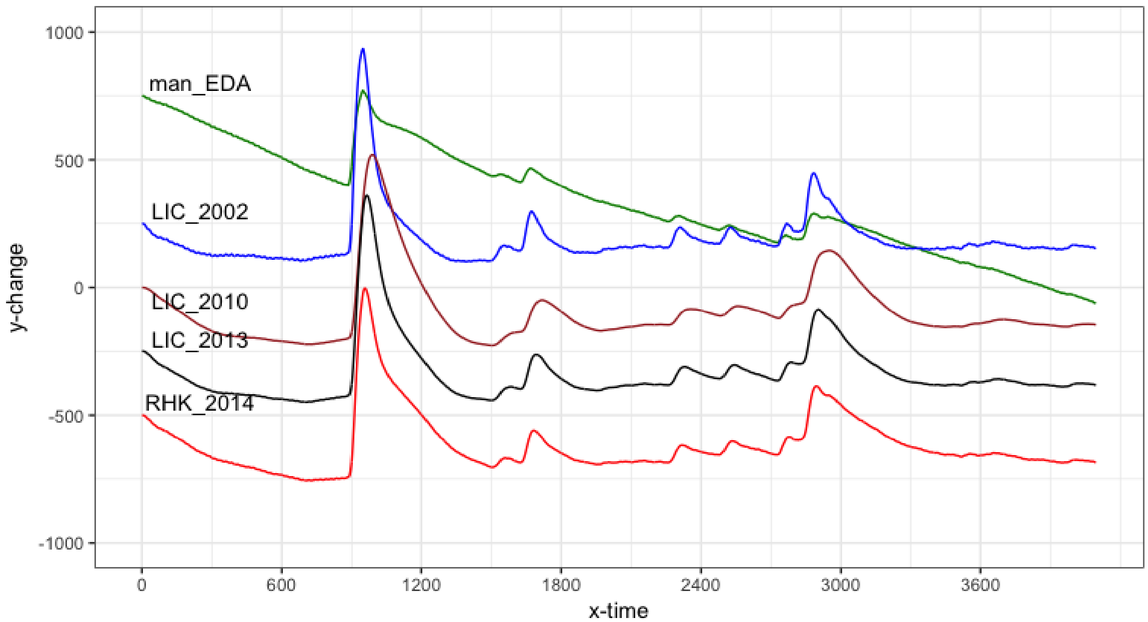
Recorded data were exported to the NCCA ASCII format and imported to the R statistical computing language environment (R Core Team, 2017) where all analysis was completed. Changes in activity were extracted for EDA amplitude change that occurred in response to all relevant and comparison test stimuli. The response feature was the maximum distance from the onset of a positive slope segment to a subsequent peak of a positive slope segment, as described by Kircher and Raskin (1988), using an evaluation window of 15 seconds from stimulus onset and a .5 second latency requirement.

The relative change in activity was calculated as the R/C ratio for each relevant stimulus, for which the change in the activity at each relevant stimulus was compared to the change in activity using the greater change in physiological activity at either the preceding or subsequent comparison stimulus. R/C ratios were subject to log-transformation so that the distribution of numerical results was symmetrical and centered at zero. This transformation permitted the calculation of point-biserial correlation coefficients for the criterion state (guilt or innocence) and the logged R/C scores using manually-centered EDA and different auto-centering EDA solutions.

Limited information is available in the published literature on EDA signal processing design specifications in computerized polygraph systems. Lafayette Instrument Company (LIC; 2013), has made filter design specifications available, including the auto-centering EDA filter design in use in computerized polygraph systems from 2002 to 2007, and changes to the auto-centering EDA filter design in 2010 and 2013. The LIC 2002 auto-centering EDA specification used a high-pass filter with a corner frequency of .04Hz along with a low-pass smoothing filter in the form of a moving average period of .5 seconds, which corresponds to a corner frequency of .883Hz. The high-pass-filter in 2010 was developed heuristically through user experience with the goal of developing an auto-centering EDA that provided a satisfying experience under a wide variety of field testing conditions, and employed a high-pass corner frequency of .05Hz coupled with a low-pass smoothing component also at .05Hz. Because all DSP filters include a gain coefficient, the result was an EDA signal that was smooth, stable and easy to use and interpret. In 2013, Fourier analysis (Shown in Figures 2 and 4) and optimization experiments led to the development of an auto-centering EDA filter with a high-pass corner frequency of .03Hz along with a smoothing filter at .2Hz. DSP versions of each of these auto-centering EDA filter specification were calculated and coded in the R statistical computing language (R Core Team, 2017) for this study.

Published descriptions of auto-centering EDA filter designs were not located for most commercially available polygraph instruments, though it can be reasonably assumed that important differences exist in the auto-centering EDA filter specifications for different polygraph instrument manufacturers. Raskin, Honts and Kircher (2014) wrote that a high-pass filter with a time constant of 10 seconds would not damage the signal of interest to field polygraph examiners. For this study, an auto-centered EDA filter was optimized using a high-pass filter with a corner frequency of .0159Hz coupled with a low-pass smoothing filter with a corner frequency of .443Hz, and compared with the other auto-centering EDA filter specifications and with manually centered EDA data. Figure 5 shows a plot of these EDA filter specifications

Figure 5. Composite plot of different EDA filter specifications.



Rank correlation of Auto-centered EDA and Manually-centered EDA modes

Rank order scores were determined for all measured responses to relevant and comparison stimuli using the manually-centered EDA and all auto-centered EDA modes for which published descriptions exist for the filter specifications. Rank correlations, including Pearson's Rho and Kendall's Tau, were then calculated for all auto-centered EDA solutions relative to rank scores using the manually-centered EDA mode. Data for the rank correlation was N=12 unconfirmed screening exams for which the target issues involved corruption, unauthorized information disclosure, receiving unlawful benefits, and contact with organized criminal groups. Confirmed case data was not necessary for the rank correlation analysis. Examination formats involved a mix of two, three and four question multiple issue screening formats. Demographic information such as age, gender, education, and ethnicity were not included in the data that was available for study. Rank scores were calculated for each of 47 test charts for the 12 examinations.

Results

Table 1 shows the point-biserial correlation of logged R/C scores calculated using manually-centered EDA data along with the point-biserial correlations for logged R/C scores of four different auto-centered EDA solutions, including the LIC 2002 filter specification, LIC 2010 specification, and the LIC 2013 filter specification. Also shown in Table 1 is an auto-centering EDA filter specification that was optimized using the recommendation of Raskin, Honts and Kircher (2014). The coefficient of determination can provide a more easily intuitive estimate of the proportion of information in the criterion state that is explained by or can be attributed to the logged R/C ratios. The auto-centering EDA filter specification optimized from the recommendation by Raskin, Honts and Kircher produced the strongest correlation with the criterion (r = .70), marginally exceeding most of the other filter specifications and equaling the filter specification from LIC (2013).

Table 1. Point-biserial correlations and (coefficients of determination) for EDA filters.

	Manual EDA	LIC2002 Auto EDA	LIC2010 Auto EDA	LIC2013 Auto EDA	RHK2014 Auto EDA
r_{pb}	.69	.68	.58	.70	.70
r^2	.48	.46	.33	.49	.49

Table 2 shows the rank correlations and coefficients of determination for rank scores of each of the auto-centering EDA specifications with the rank scores using manually-centered EDA data. Spearman's Rho is a common form of rank correlation, calculated in the same manner as the Pearson correlation coefficient, but is thought to provide an optimistic estimate of rank order relationships. For this reason, Kendall's Tau coefficients are also shown in Table 2.

Table 2. Correlations and (coefficients of determination) for rank scores of different auto-centered EDA specifications with ranks scores of manually-centered EDA data.

	LIC2002 Auto EDA	LIC2010 Auto EDA	LIC2013 Auto EDA	RHK2014 Auto EDA
Spearman's Rho	.96 (.92)	.92 (.84)	.95 (.90)	.96. (.92)
Kendall's Tau	.90 (.81)	.84 (.71)	.89 (.79)	.92 (.84)

As expected, Kendall's Tau correlations provide a more conservative estimate than Spearman's Rho. The auto-centering EDA filter specification that was optimized following the recommendation of Raskin, Honts and Kircher (2014) again produced the strongest rank correlation with rank scores calculated from manually-centered EDA (τ = .96), marginally exceeding that of the other filter specifications.

Summary and Conclusion

This project involved the study of the strength of relationship between the criterion state of deception or truth-telling and polygraph scores calculated from the change in phasic EDA extracted from responses to relevant and comparison stimuli using CQT data from confirmed field cases. EDA responses were calculated using several different DSP solutions for auto-centering EDA specifications as described in the available published literature on polygraph instrumentation. An additional analysis focused in the strength of relationship between rank order scores of manually-centered EDA data and rank scores using several auto-centered EDA filter specifications.

Two issues of practical concern will determine the effectiveness of any EDA signal processing solution for field examiners who are interested in maximizing the discrimination of truth and deception for practical purposes, and for researchers who may be interested in studying and understanding the effectiveness of the polygraph test. The first of these concerns involves the strength of relationship between the criterion of deception or truth-telling and the numerical scores based on phasic EDA responses extracted from recorded EDA data. Signal processing designs that maximize the correlation between the extracted data and the criterion should be preferred over those for which the correlation with the criterion state is weaker.

A secondary concern will be the degree to which the scores of an auto-centering EDA solution may differ from the scores of manually-centered EDA data. It should not be overlooked that this concern should be secondary to the diagnostic coefficient. Nevertheless, it will be preferable to minimize the occurrence of score differences between different EDA signal processing solutions to the extent possible. It is, however, inevitable that different signal processing designs and different filter specifications will, under some circumstances, result in different scores – and this may cause great consternation to persons who harbor unrealistic expectations that the different signal processing models should always perform equivalently or uniformly.

In the extreme, some people may wish

to avoid the sensation of uncertainty or ambiguity that comes from having to choose from different possible signal processing solutions by requesting that the options be restricted. The fallacy herein is that the reduction of signal processing options for avoidant reasons would require a parochial approach to this decision, to maintain the false perception that no such choice exists. In fact, EDA signal processing options and choices exist, and the selection of a signal process method must still be made by someone – including, unfortunately, the option of ignoring any potential for improved EDA signal processing. Developing awareness of these options and choices, and then becoming aware of the underlying signal-processing issues creates opportunity and potential for continued improvement.

All data is a combination of signal and noise. Data will ideally consist of only signal information with little or no noise, and this would result in criterion correlations that are very high or perhaps uniform with the criterion of interest. More realistically, all data contain both signal and noise. Any suggestion that a manually-centered EDA signal processing mode is *purer* than auto-centered EDA data would be misguided and misinformed. An optimal signal processing model – for any data analytic context – will maximize the availability of useful signal information (phasic EDA responses) and reduce the volume of interfering noise (tonic EDA instability) within the data. Properly constructed EDA filter designs, whether implemented in hardware or software, will optimize the signal to noise ratio in a manner that can be investigated and verified with regard to the criterion of deception or truth-telling. It can therefore be argued that a correctly designed auto-centered EDA mode, because it can reduce unwanted tonic noise, can provide EDA data that is more *pure* in terms of available and useful signal information than a manually-centered EDA mode.

All data require signal processing to be of any use. For this reason, it should be obvious, but is worth stating explicitly, that manually-centered EDA data is not synonymous with *raw* EDA data. Indeed, the notion of *raw* EDA data is a misnomer and misunderstanding when referring to data that has been acquired as a function of an electronic circuit coupled with an analog-to-digital-converter before packeting and transmission to a computer where information is further transformed from digital information to human-readable numerical sequence that can scaled and be plotted graphically. All EDA data has been processed if it available for any form of interpretation.

Just as there is no form of raw EDA data that can be made available for use by polygraph field examiners, it would be a false suggestion that older style analog polygraph instruments provide a more pure or raw form of EDA data than modern computerized polygraph instruments. EDA data for analog polygraph instruments of the past was recorded via liquid ink that was drawn via capillary action through a tubular pen that was controlled by an electric motor that moves physically back and forth as a function of an electronic circuit that almost certainly required electronic amplification (a form of active filter) to make use of the tiny voltages applied to human polygraph examinees. Higher frequency noise from analog EDA data was smoothed by the physical friction of the pen moving across the chart paper, and by the damping action of the moving mechanical parts that were needed to move the pen. Low-frequency tonic instability was well-known to examiners who used analog polygraph instruments and is without doubt the reason that engineers began several decades ago to include an auto-centering EDA mode in analog polygraph instrument designs.

Virtually all computerized polygraph systems today provide examiners with the option to use manually-centered EDA signal processing designs or auto-centering EDA signal processing solutions. Both EDA modes involve signal processing which inherently involves some form of passive or active filtration in addition to other operations. In years past, prior to any opportunity to study the effectiveness of different auto-centered EDA filter designs, it would have perhaps been a wise and understandably cautious position to advise the use of manually-centered EDA data whenever possible. This would have avoided the quandary of unknown information about the adequacy of the EDA filter specifications. Today – having studied and analyzed various EDA signal processing designs and filter specifications – such a recommendation would seem unnecessarily short-sighted.

An overly-simplistic and under-informed decision process would attempt to view the choice between manually-centered EDA data and auto-centered EDA data in terms such as "*do-i-want-all-of-the-data-or-do-i-want-some-of-it-to-be-filtered-out?*" Simplistically we might want all of the data; but in reality, we will get all of the data together with all of the noise. More realistically we will want the data with as little noise as is reasonably achievable – similar to focusing the lens of a camera, for which even unfocused images contain all of the image data in terms of the light waves that contributed to the image, though an unfocused image is confused by excessive noise. Unfiltered data in many contexts is often of little or no practical value. What is important is that signal processing filters are designed correctly. Virtually no field of professional activity that involves the use of data today will forgo the use of well-constructed signal processing methods – including high-pass, low-pass and other filter designs – to improve the signal-to-noise ratio and usability of data that will be subject to analysis.

In summary, numerical scores that were extracted from auto-centered EDA filter specifications that were developed through statistical analysis and optimization produced point-biserial criterion correlations that equaled or exceed that of scores calculated from responses extracted from manually-centered EDA data. In addition to optimizing the diagnostic value of the recorded data, properly designed auto-centering EDA modes can improve the ease of use, interpretation, and manageability of the data. Not surprisingly, auto-centered EDA rank scores that were developed through non-statistical heuristic methods did not exceed the correlation of manually-centered EDA rank scores. This may have an important bearing on field practice recommendations, for which the principles of evidence-based practices dictate that standards and guidelines should be based on more than mere conjecture whenever possible. It may also become increasingly important for polygraph professionals and others to insist that all polygraph instrument manufacturers and technology providers should account for their signal processing specifications – or at the very least they should account for their performance characteristics. All professions faced with an option of different technology solutions will need to be correctly informed and educated about technology and performance issues or they will be at risk of becoming confused or misled in their decision-making about technology and field practices. It is hoped that the publication of this analysis will be of some use to field practitioners and researchers interested in contemporary polygraph methods.

References

Boucsein, W. (2012). *Electrodermal activity (2nd ed.).* New York: Springer.

Boucsein, W., & Hoffmann, G. (1979). A direct comparison method of the skin conductance and skin resistance methods. *Psychophysiology, 16(1)*, 66-70.

Brigham, E. O. (2002). *The Fast Fourier Transform.* New York, USA: Prentice-Hall.

Butterworth, S. (1930). On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer 7,* 536-541.

Cooley, J. W. (1987). The re-discovery of the Fast Fourier Transform algorithm. *Mikrochimica Acta, III,* 34-45.

Edelberg, R. (1967). Electrical properties of skin. In C. C. Brown (Ed.) *Methods in Psychophysiology* (1-53). Baltimore, MD: Williams & Wilkins.

Fowles, D. C. (1986). The eccrine system and electrodermal activity. In M. G. H. Coles, E. Donchin, & S. W. Porges, (Eds) *Psychophysiology: Systems, processes, and applications,* (51-96). New York, NY: Guilford Press.

Honts, C. R., & Barger, S. D. (1990). Comparison of the relative utility of skin conductance and skin resistance couplers for the measurement of electrodermal activity. *Polygraph, 19,* 199-209.

Johnson, L. C., & Lubin, A. (1966). Spontaneous electrodermal activity during waking and sleeping. *Psychophysiology, 3(1),* 8-17.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291-302.

Krapohl, D. J. & Shaw, P, K. (2015). *Fundamentals of Polygraph Practice.* Elsevier.

Lafayette Instrument Company (2013). *LX4000 and LX5000 EDA Circuit Design and Evolution.* [Retrieved from: http://lafayettepolygraph.com/eda-info.pdf on 1/27/2018].

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology, 8,* 656-672.

Ohm, G. S. (1827). *Die galvanische Kette, mathematisch bearbeitet.* Berlin: T. H. Riemann.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raskin, D. C., Honts, C. R., & Kircher, J. C. (2014). *Credibility Assessment.* Elsevier.

Venables, P. H., & Christie, M. J. (1973). Mechanisms, instrumentation, recording techniques, and quantification of responses. In W. F. Prokasy & D. C. Raskin (Eds.), *Electrodermal activity in psychophysiological research* (1-124). New York, NY: Academic Press.

Dedicated to Memory A.S. PODSHIBIAKIN (15.07.1946-31.12.2016)

Criminalistics Investigation Using a Polygraph: Russian Federation Experience

Alexander S. Podshibiakin, Yury I. Kholodny and Donald J. Krapohl[1]

Abstract

Luria (1928) first suggested focusing on the examination of traces in human memory of those who are suspected of crimes. This article outlines the theory behind the forensic polygraph examination technique aimed at searching for memory traces of the past events. The theory supports the application of polygraph techniques in the course of forensic psychophysiological examinations, which results may be admissible in Russian courts. Legal rules enacted in Russia recently have created a favorable environment for a wider use of the polygraph in criminal procedure practice. The *Typical technique of forensic psychophysiological expertise using polygraph* was created to advance this area of practice for the Investigative Committee of the Russian Federation.

Keywords: forensic polygraph examination, memory, traces in human memory, criminal procedure practice.

Criminal acts almost inevitably produce evidentiary traces that can conceptually fall into one of two categories. The first type is tangible or physical: fingerprints, explosives traces, tool marks, bullet striations, etc. The second one is the memory of the event imprinted in mind of the criminal (hereafter referred to as "ideal traces").

In 1920s, when the well-known Russian psychologist A.R. Luria urged investigators to pay "serious attention to the investigation of the traces of crime that persist in the criminal itself, in his mind", he believed "these

1 D.Sc. (Law) Alexander Podshibiakin (deceased) was a professor with the Moscow State Institute of International Relation (MGIMO) under the Russian Ministry of Foreign Affairs. He is also the Honored Worker of Science of the Russian Federation.

D.Sc. (Law), Ph.D. (Psychology) Yury Kholodny is a professor with the Bauman Moscow State Technical University, and previous author with this publication.

Donald J. Krapohl is an APA Past President and Editor, coauthor with Pamela Shaw of the textbook *Fundamentals of Polygraph Practice*, and a Division Director for the Capital Center for Credibility Assessment, the largest provider of certified polygraph examiners to the US Government.

Authors' Note: This article was originally written in the Russian language by the first two authors, and subsequently translated into English. The third author strived imperfectly to make the paper approachable to native English speakers in the polygraph community, but some Russian concepts and terms may still not be immediately clear to readers unfamiliar with Russian science in general and Russian polygraphy in particular. In future articles we will expand upon the idea of memory traces in forensic and polygraph science to acquaint this concept to a larger professional readership. Questions and comments regarding this article should be sent to Dr. Kholodny at kholodny@yandex.ru.

2 In the 1930s the USSR made assessments of the application of the polygraph for law enforcement purposes solely on ideology grounds, stating that "no unscientific and unethical methods, as the lie detector and the like, frequently used in policing and court proceedings of imperialistic countries, may be used in the criminal proceeding" (Panteleev I. F., Selivanov N. A. (1984). *Criminalistics*. Moscow: Legal literature, p. 30).

traces are as feasible and objective as any other external traces" (Luria, 1928). For many decades, however, the former Soviet forensic science field totally dismissed the possibility of applying any device, polygraph in particular, to inquire into the ideal traces. Even now amateurs mislabel the polygraph as a "lie detector"[2].

The fundamental changes in the life of Russian society since the beginning of the 1990s has revised the long-standing view of the impossibility of using the polygraph for law enforcement purposes. In 1993, according to federal law on Operational Search Activity (OSA), the Russian Ministry of Justice allowed interrogation using polygraph (IUP) to be implemented. A year later in Russia the first forensic science department was established to apply polygraph for law enforcement purposes: in this way IUP was incorporated into national forensic science (or, in Russian terminology – criminalistics). During the 1990s experience was accumulated for applying the polygraph to OSA investigations. Concurrently, considerable research was carried out to provide a theoretical footing for lawful use of this kind of interrogation in crime detection and investigation, which further promoted the incorporation of polygraph into crime investigation.

Finally, by the early 21th century a new branch was developed in the Russian forensic sciences: *criminalistic study with the polygraph for testing for traces in human memory* (*criminalistics polygraphology*). This branch of criminalistics deals with practical application of methods and programmed devices to diagnose the presence of traces of past events in human memory, and consequently extracting the information from the person which he conceals.

Theoretical basis of forensic research using polygraph

When conducting criminalistic studies on the use of the polygraph to uncover traces in human memory of a particular individual, it is his memory as well as documentary case information prompting such investigation that is the object of research.

An individual perceives the events in the outside world by means of various senses, and stores some internal representations of the sensations as "ideal traces", subject to the sensory channel in which the information was received – eye, ear or otherwise. Noteworthy occurrences (including details of a crime) induce human memory traces. However, different individuals will unavoidably have variations in their memory traces regarding the same event due to a range of subjective factors: the conditions under which the event was perceived by the individual, his physical and emotional condition, cultural and national background, etc.

When reproducing or communicating past information from his memory an individual must remember the past and have his memory traces materialized by means of motor activity – for example, oral or written communication, drawing scenes, pictures, use of gestures or methods of communication. During recall some of the above mentioned primary subjective factors come to have a repeated and unpredictable effect for the owner of memory traces, and subsequently, for the outside observer (e.g. investigator). The explanation for this is simple: "Remembering the past experience can't be absolute. The level of discrepancy between the memories and the past event depends on the personality features – his or her purposes, motives, aims, remoteness of the event remembered, as well as its significance to the individual" (Psychophysiology, 2012). Moreover, the ability to access memory traces can be affected by an objective factor – such as oblivion (the partial or complete loss of memory traces).

Finally, when reproducing memory traces, subjective intervening personal interests of an individual may come into play. It is a volitional act that demonstrates either the intentional distortion during the materializing process of memory traces (misrepresenting the memory trace), or mere denial to materialize such traces. The latter may occur in two ways: 1) "I will say nothing", or 2) "I will say nothing as I know nothing". With the exception of where there is no memory trace to materialize, all the above-mentioned cases involve concealment of the information.

It is well known that numerous life events (especially particular life circumstances) are not remembered by an individual. However, those events which are of high personal value are imprinted in the individual's memory automatically, quickly and create a "stable and long-lasting traces in emotional memory" (Danilova, 1999; p. 130) which are not deleted (unless special methods are applied), are preserved for all life (subject to healthy brain activity) and are not subject to oblivion (i.e. destruction).

At a particular time period some part of human memory activates and is accessible to an individual in his day-to-day activity. With current technology it is not possible to localize, describe and identify memory traces or "ideal traces" in the brain itself (in Russian psychology, psychophysiology and allied science it is characterized as an engram). Nevertheless, it is possible to diagnose the presence or absence of "ideal traces" of a particular event in human memory. This is because "the active memory is … an aggregation of active engrams….which operate within the electric system" of an interlinked human brain structure (Psychophysiology, 2012). When a personally significant "ideal trace" of the event is found in the individual's active memory, such as details to a crime, neural activity of the relevant brain structures increases. This activity can be externally detected as reactions displayed in some physiological systems of the human organism. The physiological reactions can be recorded with a polygraph for the purpose of the forensic diagnosis for the presence of "ideal traces".

Longstanding practice has produced strong evidence that successful detection of personally significant memory traces in a healthy individual can be revealed by reactions appearing in breathing, the cardiovascular system and by the electrical conductivity of the skin. Hence, the primary objective of criminalistics polygraphology is to diagnose the presence or absence of "ideal traces" of the past events in an individual's memory.

For convenience we use hereafter the acronym "CSTM" as shorthand for the more unwieldy "Criminalistics Studies on the use of polygraph Traces in human Memory.[3]" In the course of the CSTM, stimuli are used to elicit responses evoked by the presence of "ideal traces" of personally significant events:

- In the case of visual perception, tangible objects are used: things, photographs of people, photographs of objects, photographs of plots of land; maps, documents, printed words and statements etc.;
- In the case of audio perception, the CSTM uses semantic concepts (or ideal objects), e.g., verbally presented questions aimed at inquiring into a specific occurrence or a particular circumstance.

Stated another way, the stimuli in a CSTM can be objects (or images of objects) or spoken words, and their presentation during polygraph testing can induce physiological reactions that permit inference of the person's complicity in a crime. As Russian practice shows, ideal objects (i.e. vocal presentation of questions) are used in the CSTM in over 99% of the cases, whereas it is less than 1% for tangible things (images or objects). So CSTM involves the presentation of both tangible and ideal target objects to help discern the presence or absence of relevant target objects in human memory. Experts in criminalistics polygraphology (hereinafter – polygrapholog[4] [Russian terminology]) know well that target objects in the human memory during CSTM may be the acts that had been committed, the sequence of those acts, faces of the persons, their surnames and nick names, plots of land, addresses and names, maps and sketches, dates and time periods, etc.

Russian forensic science dealing with

3 In English this may be known as a polygraph examination.

4 In English this person would be called a polygraph examiner.

memory traces long ago proposed three distinctive features of "ideal traces" as compared to tangible (material) traces (Averjanova, Belkin, Korukhov & Rossinskaya, 1999):

1) "ideal traces", being intangible in themselves, are not accessible for direct investigation - they cannot be directly seen, touched, weighed, or measured;

2) investigating "ideal traces" is feasible only after their materialization (manifestation) in a form of motor activity;

3) materialized "ideal traces" can be as informative as tangible traces or evidence.

Investigative technology of the CSTM has discovered three other features of "ideal traces", not yet described in forensic science (Podshibiakin & Kholodny, 2001; 2002). Specifically, "ideal traces" can be:

1) willfully misrepresented by the owner of memory traces when materializing them;

2) discovered without materializing by the person (i.e. absent his volitional motor activity), e.g. through physiological recording;

3) lost by the person (i.e. forgotten) but is not subject to being deleted from memory purposefully

Using the CSTM imposes a number of requirements and restrictions. The crucial condition for using the polygraph is the voluntary consent of the examinee to conduct the CSTM, obtained in writing in a prescribed form. Voluntary participation is made necessary both by observance the constitutional rights and freedoms of the examinee, on the one hand, and specific CSTM technology on the other hand. A person cannot be forced to undergo testing: either he won't allow having sensors placed on him, or he won't follow the instructions given by the polygrapholog. Moreover, the polygrapholog is required to notify the person from the very beginning of his right to terminate the CSTM at any time.

The health and psychological condition of the person at the time of testing is crucial for successful diagnostics of the examinee's "ideal traces". Legal restrictions in applying CSTM were specified long ago, and are well known and strictly complied with by polygraphologs.

Preparing for the CSTM, the polygrapholog is bound to research the target event and select those facts and circumstances, "ideal traces" of which he is to determine whether they exist in the memory of the examinee. To have successful diagnostics and a fruitful CSTM, only those facts or circumstances should be offered for investigation which must be known to the person if he was involved in some event (reliably preserved by means of his emotional memory). Practical experience confirms the fact that "ideal traces" of personally significant past events imprinted in emotional memory can be preserved for many years. The principal authors' personal experience has been that a CSTM can successfully diagnose the examinee as having concealed information as many as 10-15 years after he had committed the crime.

Criminalistics study of human memory traces and the criminal procedure law.

Unquestionable efficiency in using the polygraph in the OSA prompted the investigators to look for the ways to implement it in actual investigative procedures. The first procedural application of the polygraph in Russia was made by Y.I. Kholodny in 1994. In the late 20th century there were many CSTMs conducted by the investigator, but using polygraph for investigative purposes in this way was reasonably criticized by procedural law experts and was soon discontinued.

The Federal Law on "Forensic Examination in Russian Federation" (enacted on May, 31, 2001, № 73-FL) was crucial in promoting polygraph into criminal procedural practice. That summer was the first time Russian specialists conducted a CSTM as an expert examination. It was called *forensic psychophysiological expert examination using polygraph* (hereafter – FPpE).

In general, *the central construct for the FPpE* is that human memory is an inherent part of the psyche. The *object for a specific FPpE* is the memory of the person submitted to expert examination on a particular case, as well as case materials. *The subject matter of a specific FPpE* is the information kept in the examinee's memory i.e. real data which are meaningful for the investigators or the trial in a criminal case and are formulated as questions, the answers to which must be given by the polygrapholog conducting the FPpE (Kholodny, 2002).

Since 2002 the number of FPpE in Russia had been steadily growing, which required specific methods to conduct. In 2006 a group of specialists proposed a "Technique for examination using the polygraph" (hereinafter – "Technique") which later was widely advertised for conducting FPpE (Komissarova & Khamzin, 2016). However, by 2008 it was obvious that "Technique" had a range of drawbacks and produced bad mistakes (Kholodny, 2008).

It should be noted that "Technique" misidentified the FPpE object, scope of expert-polygrapholog's authority and failed to show the list of tests which might be used, which resulted in erroneous conclusions, both procedurally and methodologically, based on the results of the expert examination (Orlov & Kholodny, 2012; 2013a). Therefore, results of an expert examination using a "Technique" could be easily dismissed by the court (Podshibiakin & Kholodny, 2013).

Moreover, a textbook on forensic science for the experts of the Investigative Committee of the Russian Federation (hereinafter – IC RF) stressed that "federal agencies and all polygraphologs were recommended to abstain from conducting FPpE using the "Technique" until a scientifically grounded counterpart was created" (Podshibiakin & Kholodny, 2014).

CSTM may be used for detecting crime as a IUP in OSA, or as FPpE in criminal procedure practice, but both are based on the uniform methodology guidelines. At the same time, despite the uniformity of those guidelines, the polygrapholog, when detecting the crime in OSA, has more discretion in choosing methods and tactics of the investigation. On the contrary, criminal procedure practice requirements impose substantial restrictions, and a lot of polygraphologs, who did IUP in OSA, are unaware about these restrictions, when conducting FPpE.

In Russia, upon terminating an expert examination a polygrapholog issues an "Expert Opinion" and submits FPpE's conclusion which is inevitably probable by virtue of the "ideal traces". The probabilistic method for detecting concealed "ideal traces" from a person's past (information he tries to conceal) fails to detract from the applied efficiency of such expert examination. And if when examining human memory applying proper methodology some reaction follows to the questions involving the event scrutinized, it shows "ideal traces" of that event in human memory, in which case the polygrapholog states his opinion in the following language: "the examinee's memory contains information that … (for example, "A" /the victim/ was stabbed in the back). Such fact has been established by the expert with 0.95 (95%) probability". Otherwise, the polygrapholog states his opinion in the following language: "there is no information in the examinee's memory that … (for example, "D" /the examinee/ stabbed in the back of "A"). Such fact has been established by the expert with 0.95 (95%) probability" (Orlov & Kholodny, 2009).

The Federal Law enacted on March 4, 2013 (№ 23-FL), amended the Russian Federation Criminal Code (hereinafter – RFCC) and provided that "when verifying the information about the crime committed pretrial, the inquiring officer, investigative authority, investigator, the chief of the investigative branch … are authorized to order a forensic expert examination" (Article 144, section 1). The RFCC specified that a "forensic expert examination may be ordered and held prior to bringing a criminal case" (Article 195, section 4). As a result, investigative authorities faced an absolutely different situation inasmuch as they could resort to using polygraph when necessary. Amendments to the RFCC, having enabled the conduct of an expert examination before bringing a criminal case, has in fact rendered FPpE an exclusive method of using the polygraph in investigations and preliminary inquisitions. That is why demand for a comprehensive scientifically-based FPeP methodology continues to grow.

The above-mentioned "Technique", actually being the sole document to conduct

FPpE, was subject to scientifically based criticism and found inappropriate to be implemented (Orlov & Kholodny, 2013b; Kholodny, 2014a). In 2014 the "Typical technique of forensic psychophysiological expertise using polygraph" (Kholodny & Orlov, 2014) was developed to replace the "Technique".

Unable to provide herein a detailed survey of the "Typical technique of forensic psychophysiological expertise using polygraph" (hereinafter – "Typical technique of FPpE"), some of its provisions shall be mentioned. For example, "Typical technique of FPpE" attracted attention to:

- the procedure of an implementing order to conduct the FPpE, with recommendations for the wording of the questions to ask in the course of the examination;
- the procedure for obtaining the examinee's consent to have the FPpE conducted;
- selection of tests (Kholodny, 2014a; 2014b) to be used in the polygraph examination, optimizing the number of those to provide proper solutions to the questions the FPpE had to answer;
- identifying the length of time for the expert to interrogate the examinee;
- procedural requirements to conduct the FPpE and the order of interaction between an expert and an investigator at the preparatory stage of organizing expert examination;
- the procedure for video recording the expert examination;
- restrictions imposed on the FPpE, and many other issues.

The "Typical technique of FPpE", among other issues, made corrections for the "Technique" faults, and, in particular, stated that "the FPpE object is the examinee's memory" (Kholodny & Orlov, 2014, p. 16). It should be mentioned that the FPpE technology, which are presented in the "Typical technique of FPpE", was actually developed in Russia early in 2003 (Orlov & Kholodny, 2015).

The "Typical technique of FPpE" is a work of authorship: i.e. it is not officially recognized, not compulsory, however it has been recommended by the authors for professional application for expert purposes. Should it be adopted and obtain practical recognition, a proposal might be submitted to have it formally approbated and adopted. Twenty experienced polygraphologs from the IC RF had been professionally trained in FPpE under "Typical technique of FPpE", from 2014 to 2015, to have it tested in practice. Moreover, the authors of such a technology "recommend to implement "Typical technique of FPpE" in other Federal Agencies and professional bodies dealing with FPpE" (Orlov & Kholodny, 2015; p. 257).

Overall, it may be stated that the new type of expert examination turned out to be in demand and tens (or even hundreds) of FPpE results have been admitted as evidence in various courts.

References

Averjanova, T.V., Belkin, R.S., Korukhov, Y.G. & Rossinskaya E.R. (1999). *Criminalistics.* Moscow: Group of Publishers Norma-Infra-M.

Danilova, N.N. (1999). *Psychophysiology.* Moscow, Aspect Press.

Kholodny, Y.I. (2002). Interrogation using polygraph and forensic psychophysiological expert examination. *The Volume of International scientific conference. (to commemorate Professor R.S. Belkin's 80th anniversary) "The role and significance of Professor R.S. Belkin's activity in the development of modern criminalistics".* Moscow, Academy of Management of Ministry of Internal Affairs, 422-426.

Kholodny, Y.I. (2008). Forensic psychophysiological expert examination using polygraph: the period of formation. *Bulletin of Criminalistics*, 1 (25), 25-33.

Kholodny, Y.I. (2014a). Methodical means of forensic psychophysiological expert examination using polygraph. *Legal Psychology*, 2, 26-33;

Kholodny, Y.I. (2014b) Some applied aspects of forensic psychophysiological expert examination using polygraph. *Bulletin of Criminalistics*, 2 (50), 27-36.

Kholodny, Y.I., Orlov Y.K. (2014). *Typical technique of forensic psychophysiological expertise using polygraph*. Moscow. Academy of Investigatory Committee of Russian Federation.

Komissarova, Ya. & Khamzin, S.R. (2016). Polygraph Abroad; Destination Russia. *APA Magazine*, 49, 1, 93-103.

Luria, A.R. (1928). Psychology in identifying of crimes traces. *Nauchnoe slovo [Scientific word],* 3, 79-104.

Orlov, Y.K. & Kholodny, Y.I. (2009). Forensic psychophysiological expert examination using polygraph: the problem of admissibility. *Bulletin of Academy of Economic Security of Ministry of Internal Affairs RF,* 12, 87.

Orlov, Y.K. & Kholodny, Y.I. (2012). Procedural aspects of using polygraph in criminal investigation. *Bulletin of Criminalistics*, 4 (44), 28-34.

Orlov, Y.K. & Kholodny, Y.I. (2013a). Procedural aspects of using polygraph in criminal investigation. *Criminal Process*, 3, 28-35.

Orlov, Y.K. & Kholodny, Y.I. (2013b). Using polygraph in criminal investigation: expert examination assessment recommendations. *Criminal Process*, 8, 64-73.

Orlov, Y.K. & Kholodny, Y.I. (2015). Polygraph studies as a forensic psychophysiological expert examination. *Bulletin of Academy of Investigatory Committee of RF*, 1, 251-257.

Panteleev, I.F. & Selivanov, N.A. (1984). *Criminalistics*. Moscow: Legal literature.

Podshibiakin, A.S. & Kholodny, Y.I. (2001). On specifying and adding targets of criminalistics diagnostics. *Russian legal doctrine in XXI century: problems and solutions: Scientific conference (October 3-4, 2001).* Saratov, p. 225-227.

Podshibiakin, A.S. & Kholodny, Y.I. (2002). Criminalistics polygraphology – a new trend in criminalistics aimed at researching “ideal traces”. *Criminalistics: urgent issues of theory and practice (July 20-21, 2002).* Rostov-on-Don. Rostov Law Institute of Ministry of Internal Affairs of RF, p. 293-297.

Podshibiakin, A.S. & Kholodny, Y.I. (2013). Quality assessment of criminalistics investigation and expert examination using polygraph. *Criminal Process*, 4, 60-66.

Podshibiakin, A.S. & Kholodny, Y.I. (2014). Use of polygraph for obtaining forensic information. In A.I. Bastrikin (Eds.), *Criminalistics. Textbook.* Moscow: Examination. Volume 1, 325-339. *Psychophysiology*. (2012). Sankt-Petersburg, Peter.

Technique for examination using the polygraph (2006). *Instrumental lie detection: realities and perspectives in its application against crime: The Volume of International Scientific Forum.* Saratov: Saratov Law Institute of Ministry of Internal Affairs of RF, 90-96;

The Evidence Recognition Sensor ™

Timothy J. Weber and R. Blake McConnell [1]

Abstract

The Evidence Recognition Sensor (ERS) was developed in order to provide criminal investigators with an affordable, easy to use interrogation tool requiring only a brief training course. This device is not a polygraph instrument as it consists solely of an electrodermal, respiration and seat motion sensor connected to a sensor box. It may only be employed in conjunction with the Concealed Information Test (CIT) and is thus utilized to detect evidence recognition rather than deception. Results indicate that in a group of 28 subjects, the accuracy of the ERS using the ICAS-4 scoring methodology is 86% for the guilty, 93% for the innocent and 90% overall with no tests for which a decision could not be made. The accuracy of the ERS at P < .05 is better than chance for both innocent and guilty subjects.

Introduction

The Evidence Recognition Sensor (ERS) was developed by International Credibility Assessment Service (ICAS) LLC and manufactured under an exclusive agreement by Axciton Systems Inc. to provide small and midsized law enforcement agencies with a means for their criminal investigators to detect evidence recognition without the expense of training a polygraph examiner and purchasing a polygraph instrument. The ERS is not designed to replace traditional polygraph instruments as it lacks the capability to detect deception though the conduct of comparison question specific issue tests or conduct applicant screening examinations. It is not a polygraph instrument, as it consists of only a seat motion sensor, single pneumograph tube, electrodermal finger plates and sensor box, and utilizes standard Axciton software. The ERS does not include a cardio component, relying exclusively on the electrodermal component to detect recognition. During the development of the ERS system this was not seen as problematic in that the eletrodermal component has been proven to be the most accurate in polygraph testing (Handler, Nelson, Krapohl and Honts, 2010). The ERS is designed to administer a Concealed Information Test (CIT), also known as the Guilty Knowledge Test (GKT), in criminal investigations to detect recognition of evidence discovered at the crime scene (Lykken, 1998). While a polygraph examiner could conduct a CIT in a criminal investigation, it is the investigator, who is first on the crime scene, who can best assess the questions to ask and control key information from leaking out to the public.

Other instruments capable of collecting psychophysiological data have been de-

1 Timothy J. Weber Ed. D. retired in 2013 after serving as an NSA special agent, Quality Assurance Program inspector, NCCA instructor and polygraph quality control chief. Dr. Weber is now a contract examiner for OGSystems Inc. and is the chief of research and education for ICAS LLC.

Retired FBI Supervisory Special Agent R. Blake McConnell, formerly a polygraph examiner assigned to the FBI Behavioral Analysis Unit, is currently president of ICAS LLC. He also serves as an adjunct professor at George Washington University's Forensic Psychology program and as a contract examiner for OGSystems Inc.

signed for non-polygraph examiners to conduct CIT's in support of criminal investigations and to detect deception in interviews of foreign nationals in military operations overseas. However, unlike the ERS, these instruments lack the ability to monitor chest movement associated with respiration patterns, known to be the most reliable method of detecting countermeasures during psychophysiological testing. ERS utilizes a single pneumograph tube placed at mid thorax and a seat motion sensor exclusively for the purpose of detecting countermeasures. Though increasing the ERS instrument's cost, the addition of this anti-countermeasure capability will significantly increase the reliability of ERS test results in that Ben-Shakhar (2016) noted that the CIT's lack of resiliency to countermeasures was well documented in the literature. The results of his previously conducted failed field tests were blamed on countermeasure use (Ben-Shakhar, 2016). Even as early as 1959, Lykken recognized this weakness.

The CIT

Unlike the traditional polygraph comparison question test (CQT) that tests for deception using direct, accusatory questions such as "Did you stab that man?" the CIT, is a technique designed to detect if a person is concealing guilty knowledge of a crime. The following scenario illustrates how the CIT is used in conjunction with an ERS in support of a homicide investigation:

The body of a victim who apparently died of stab wounds is discovered inside of a vacant house. Evidence from the crime scene is analyzed to identify items that are most likely to be memorable and salient (important) to the perpetrator. The following items are noted: a broken window used to gain entry, a bloody knife near the body that was forensically identified as the murder weapon, a towel used to gag the victim, duct tape used to bind the victim, and a sleeping bag that was laid over the victim's body. This is considered hold back information and is kept confidential by the investigating agency. A suspect is subsequently identified but denies having any knowledge of the crime. A CIT is constructed to determine if the suspect recognizes various elements from the crime scene that should only be known to the perpetrator (Lykken, 1998). The following questions may be asked during the ERS CIT examination:

1. Concerning the weapon used, was it a rope?
2. Was it a baseball bat?
3. Was it a handgun?
4. Was it a knife?
5. Was it a crowbar?

The key question about the actual murder weapon, the knife, is embedded among several control questions about items that were not tied to the crime scene (Krapohl, McCloughan and Senter, 2009).

The suspect is instructed to answer all of the questions on the test "No." The actual perpetrator of the crime is expected to have physiological reactions to the correct, or key, items that can be discriminated from the decoy, or control, items. It is exceptionally unlikely that someone having no involvement in the crime would consistently have their strongest reactions to the key items. With the inclusion of an increasing number of key questions it becomes all but mathematically impossible for an innocent person to randomly react to the majority of them. A numerical evaluation of the test data is conducted. The accuracy of the test is significantly affected by the number of key items tested. Thus the CIT should include as many key items as possible. Conclusive results of a CIT with a large number of key items can be relied upon with a high level of confidence (Elaad & Ben-Shakhar, 2003). The scenario supports several other key items such as, method of entry into the house, what was used to gag the victim, binding material, and item covering the victim.

The primary goal of key selection is to identify items that will be salient and memorable to the perpetrator but unknown to an innocent examinee. Key selection requires experience in evaluating crimes scenes from a behavioral perspective to ensure that items selected for testing have the best potential for being memorable and salient to the perpetrator (Elaad & Ben-Shakhar, 2003). A person's ability to recall details of a crime he or she committed can be influenced by emotions experienced while committing the crime. In a review of numerous studies involving emotional memory, Levine and Edelstein found that autonomic arousal during an emotional event is known to narrow a person's attentional focus toward central details of an event at the expense of attention to peripheral details (2009). This leads to enhanced encoding and consolidation of central information into long term memory, with an unremarkable or even impaired effect on memory for peripheral details. For example, an offender may have vivid memories of discovering valuable items while burglarizing a residence, but unable to recall items that were not worth taking. This phenomenon, known as memory narrowing, or tunnel memory, has also been shown to occur when negative or positive emotions are experienced in pursuit of a goal (Kaplan, Van Damme, and Levine, 2012; Levine and Edelstein, 2009). Emotions such as anger, fear, or frustration experienced when an incident is perceived as interfering with a goal have been shown to promote the narrowing of memory for details of that incident, just as positive emotions, such as hope, confidence, or excitement have promoted enhanced memory for details of incidents perceived as furthering goals. A thief who became angry when having difficulty breaching a door may have vivid memories of damage he caused to the door, just like a rapist overcome with confidence after discovering an unlocked bedroom window may have enhanced memory for details of the window. Once a goal is attained, however, a broader array of matters may demand a person's attention, decreasing the potential for memory narrowing. After successfully killing a rival gang member, an offender may be too concerned about the chain of events that will follow to experience memory narrowing while exiting the scene of the crime.

Due to the effects of memory narrowing, the level of autonomic arousal and range of goal related emotions an offender experiences during the commission of a crime can impact what he or she may later be able to accurately recall about details of the crime scene. This makes it important to consider what the offender's motivation may have been when evaluating a crime for potential test keys and subsequently constructing a CIT with the highest potential for providing accurate results.

Theoretical Foundation and Validity

Experimental research supports a theory that guilty CIT responses are consistent with the autonomic characteristics of an orienting response that occurs when a particular stimulus is perceived as being more significant than surrounding stimuli that lacks novelty (Verschuere, Crombez, De Clercq, and Koster, 2004). A key question is significant to a guilty examinee when it stimulates memory of something encountered during the commission of a crime while surrounding control questions do not. The primary physiological response associated with knowledge of a key item measured on a CIT is an increase in electrodermal activity (EDA), although other physiological reactions, such as respiratory suppression, heart rate deceleration, and decreases in normalized pulse volume, have also been found to have varying degrees of significance in detecting guilty knowledge (Elaad & Ben-Shakhar, 2003; Hira & Furumitsu, 2009; MacLaren, 2001; Verschuere, et al. 2004). The CIT has been described as a cognitive approach to psychophysiological detection (Ben-Shakhar & Freudy, 1990), although some have hypothesized that defensive reactions may be responsible for enhancing responses in high stakes situations (Verschuere, et al; 2004).

In 1998, Gaines (2013) conducted a study of the CIT that modified the Lykken scoring criteria. Gaines used a one chart for each of a minimum of three keys. His unique scoring system consisted of scoring the most significant electrodermal response as "3" the next most significant as a "2" and the third most significant as a "1." A "2" was added to the score if the most significant response was seen in the respiratory or cardio channel. In evaluating accuracy of the CIT with the polygraph, Gaines (2013) found that his "232 system" resulted in an accuracy rate of 82% for guilty and 87% for the innocent with an overall accuracy rate of 84% with an N of 89. The 232 in the system's name came from the weight assigned to each of the components in conducting the test data analysis.

The CIT has been the subject of more theoretical validation research than all other types of polygraph techniques combined (Krapohl, McLoughan, and Senter, 2009). Numerous studies have been conducted over the last three decades on the validity of the CIT and the various factors affecting it. Elaad and Ben-Shakhar, (2003) conducted a meta-analytic review of 80 such studies in order to estimate the validity of the CIT. Their analysis only included laboratory studies that used EDA measures alone as a basis for the reported results. They concluded that a properly constructed and administered CIT has an excellent potential for detecting information and accurately discriminating between individuals with and without knowledge of crime related events. Elaad and Ben-Shakar (2003) caution, however, that the relatively high level of validity demonstrated in their review may be overstated since all of the studies were experimental and conducted in laboratory settings.

Use in Japan

The CIT has been the dominant polygraph technique since the 1950's and is currently used almost exclusively by Japanese law enforcement for criminal investigations. Approximately 5,000 are conducted on an annual basis. Polygraph examiners in Japan traditionally visit crime scenes, and are required to be experts in crime scene analysis. The Japanese consider the process of selecting key items for a CIT is an investigation within itself and requires a careful analysis of the crime scene (National Research Institute of Police Science, 2009). The Japanese investigators, who are aware of the requirements of the CIT, protect much more "hold back" information than their US counterparts.

However, field tests conducted in the US have been problematic Ben-Shakhar (2016). They resulted in lower accuracy rates than seen in the laboratory research. In the past this has been blamed on the use of countermeasures Ben-Shakhar (2016). The lack of available field studies makes it difficult to estimate the validity of the CIT in actual criminal investigations conducted in the United States. Regardless of this finding, Elaad and Ben-Shakhar (2003), opined that the CIT may be the most valid application of psychological principals derived from the field of behavioral science.

Advantages

The CIT is an especially useful tool when a suspect has failed a standard CQT polygraph with no admissions and a crime scene is subsequently discovered. The CIT will allow investigators to re-evaluate the suspect's possible involvement in the crime from a different and more specific perspective.

Another advantage to the CIT involves the emotional state of a suspect who is innocent, but is not hopeful he can convince his accusers of this. Based on polygraph theory, the negative emotional state of an examinee is less likely to affect the accuracy of the CIT than a CQT. While the CQT is not recommended for use immediately after a suspect has been interrogated, there is no reason why an interrogation should impact accuracy of the CIT unless crime scene information was divulged (Krapohl, McLoughan, and Senter, 2009). Likewise, individuals who have been accused by the media, advocates of the victim, family members, or others, can also become sensitized to the relevant questions on a CQT, making the CIT a more viable alternative.

Researchers have also concluded that the CIT may be more appropriate for testing psychopathic offenders who demonstrate a lack of emotion associated with the defensive type response underlying CQT theory. There may be an increase in orienting response, upon which the CIT is based, associated with psychopathy, thus there is reason to believe that psychopathic offenders may be more suitable for CIT testing than others (Verschuere, Crombez, DeClercq and Koster, 2004).

Research highlighting the CIT's strong theoretical foundation has caused some to postulate that results of a properly constructed and administered CIT will meet the four major criteria for the admissibility of scientific evidence in U.S. federal courts as stated in *Daubert* (Ben-Shakhar, Bar-Hillel, and Kremnitzer, 2002). While this view is subject to further debate, it is not unreasonable to believe that CIT results could be used to assist in establishing probable cause of an examinee's involvement in a crime in support of a search warrant.

Method

Equipment

Since a prototype ERS instrument had yet to be fabricated prior to commencing this study, all but the last of the study's examinations were conducted with a standard Axciton eight channel field polygraph instrument currently in use with numerous federal and local law enforcement agencies. This instrument was utilized as the software and hardware components are interchangeable with the ERS. In order to be consistent with ERS instrumentation, only one pneumograph tube, the electrodermal finger plates and the seat sensor cushion were utilized. This eliminated the possibility that the analyst could be influenced by tracings that would not be collected by an ERS instrument. The last test in the study was conducted with the ERS prototype.

Participants

The test subjects were all graduate students in the George Washington University Forensic Psychology program. The 4 male and 24 female subjects were reflective of the male to female ratio of the student population in the forensic science discipline at the university. All of the test subjects were Caucasian. Two of the male subjects were serving police officers who had previously taken polygraph examinations as part of their hiring process. Each subject volunteered to participate in the study without expecting any compensation, though at the conclusion of the examination each was given a ten dollar gift card to a local coffee shop that is popular with the students.

The two individuals who conducted and evaluated the examinations were highly experienced former federal law enforcement and/or counterintelligence polygraph examiners who are currently conducting polygraph examinations for a defense contractor.

Design

Subjects were scheduled by and reported to a George Washington University administrative assistant and were then assigned a role of innocent bystander or criminal in a mock crime scenario similar to the standard paradigm used in polygraph research studies. They filled out standard consent forms and were verbally briefed on their role in the experiment by one of the authors. Those subjects who were designated to be “criminals” were given a vehicle's Virginia tag number and location in writing and told to take the elevator to the first underground parking level and steal the contents of a briefcase located inside that vehicle. The briefcase, containing a watch, was covered by a coat. They then placed the watch in a drawer in the professor's lounge. No post crime instructions were given or emphasis placed on the key items. Subjects assigned to the role of innocent bystander were instructed to walk through the parking level before the crime occurred, thus not observing its commission. The subjects were then introduced to the ERS analyst as suspects in the larceny that occurred in the parking garage. A pretest consisting of gathering background information, as well as an explanation of the instrument, was conducted and testing began.

Data collection

The examination consisted of the four keys, or evidence connecting items, the model of vehicle that was broken into, the door used to enter the vehicle, the item covering the briefcase and the item stolen. These were a Lexus SUV, the rear driver's door, a coat and a watch. The following questions were asked:

1. Regarding the vehicle that was broken into, was it a Ford Mustang, Chevy Impala, Toyota Rav4, Lexus SUV, Honda Accord?
2. Regarding the door used to enter the vehicle was it the driver's door, passenger's door, rear hatch, rear driver's door, rear passenger's door?
3. Regarding the item covering the briefcase, was it a shirt, dress, blanket, coat, sweater?
4. Regarding the item that was stolen was it a wallet, laptop, IPad, watch, pen?

It has been suggested that in this case, the Lexus SUV is different than the foils in that they were specific vehicle models and the SUV is a vehicle type. This criticism has some merit. All of the vehicles should have been listed as the brand name's SUV; such as Cadillac SUV, Honda SUV, or Toyota SUV. In this research however, in only four to the first charts collected did the electrodermal response have the largest amplitude, thus mitigating this concern.

This study implemented the ICAS method, a modified version of Lykken's (1998) system. While adopting the same numerical scoring rules as those proposed by Lykken, the scores for each chart were not added across the board. In the ICAS method, three charts were used for each key question. This procedure differed from the one chart for each key question employed by Lykken. This was done due to the findings of Meijer, Verschuere and Ben-Shakhar (2011) who noted that more keys resulted in greater accuracy, if less than five key questions could be formulated, there should be a repetition of a minimum of three salient keys. Though Elaad and Ben Shakhar (2002) did find that even one key could render significant results. Since American law enforcement releases a great deal of information regarding criminal acts, the accuracy of one to four keys charts would be tested as this would serve as a realistic number of keys that could be held back from the public.

Under the Lykken scoring system, a score of "2" is assigned if the largest amplitude electrodermal response occurs in response to the key. The second greatest response receives a "1." The scores from each chart are added together and if the sum exceeds half of the total possible score, the test is determined recognition indicated (RI). Any score below one half of the possible total was determined no recognition indicated (NRI). In the ICAS system the sum of the evaluation of three charts was taken for each key. If the sum equaled or exceeded three, that key was considered RI. If half or more of the keys are determined to be RI, the entire test is considered RI. These cutoff scores were anecdotally derived. The difference in the numbered versions of ICAS scoring system reflect the number of key questions; thus ICAS-4 consists of four key questions while the ICAS-3 contains only three.

At no time was the key question placed as the first question on the chart. On the second chart the question order was reversed. Question spacing was reduced to 12 to 15 seconds as this was all that was required in most cases for the subject's tracings to return to a stable base line. Breaks between tests lasted for only half a minute as there was only a need to allow the subjects the opportunity to move about. To include the pretest, the entire examination lasted approximately 45 minutes.

The First Chart Method (FCM) evaluated only the first of the three charts conducted for each key. The purpose was to discover if conducting one chart per key question would suffice to achieve accurate results. This was consistent with Lykken (1998) and the Gaines' "232" method (Gaines, 2013). Numerical scoring for the FCM required that the key had to have the most or second most significant reaction of all of the questions on the chart to be considered RI. If half or more of the keys are determined to be RI, the entire test is considered RI.

The Most Salient Key Method (MSK) was used to test the theory that the key thought to be the most significant to the guilty subject might be able to alone determine recognition. The use of a single key is consistent with Elaad and Ben-Shakhar's (2002) study that found the results of one key to be to significant in the detection of concealed information.

In this case the item stolen was considered to be the most salient to the subjects as stealing it was the object of the scenario. The item would also be presumed to be the most salient to an actual offender. Scoring for the three charts testing the single key was conducted as described for the ICAS system. The "2 cut off" method resulted from a postulation that by adjusting the threshold for determining recognition a greater accuracy for RI subjects could be achieved. Thus the threshold for evaluating the sum of three charts for each key was reduced from three to two. Per the ICAS scoring methodology, if half or more of the keys are determined to be RI, the entire test is considered RI.

Results

During the first two tests the item that was actually stolen, the watch, was inadvertently left off of the question list for that key and thus was not asked. This error was corrected in the remaining tests. Since the Lykken and ICAS-4 would have identified the correct result based upon the responses to the first three keys this was not deemed problematic, however it did impact the MSK method by reducing the N for this methodology by two.

The Lykken scoring method did not do as well in this study as he found in 1959 when he achieved 89% accuracy for the guilty, 100% for the innocent for a total of 93% accuracy with an N of 50. Here, Lykken's modified methodology could only boast 57% for the guilty, 100% for the innocent and a total of 71% accuracy with an N of 28. Since Lykken used a single chart per key question; only the first chart collected for each key was evaluated. Thus, due to this modification, it could be argued that fatigue may have played a factor in the results. However, they were similar to that seen with Gaines' (2014) research, where Lykken scoring received 65% for the guilty, 95% for the innocent and a total of 80% accuracy with an N of 40. It should be noted that Lykken employed painful electroshocks to his subjects after telling them that they were reacting to the key question. Subjects were also told that they would receive additional shocks if they reacted again. Lykken failed to admit that this may have motivated those programed as guilty into stronger physiological reactions (Lykken, 1959).

In the "2 cut off" method, adjustment to the traditional three as the cut off score expectedly resulted in a high degree of accuracy for the guilty at 93% but was disastrous for the innocent at 50%. This rendered any chance of recommending the "2 cut off" method's use in the field to be nil. The FCM performed little better at 79% for guilty, 79% for innocent and 79% total. The MSK system at 75% for the guilty, 86% innocent and 81% total was not noteworthy.

Of the ICAS methodologies the ICAS-4 was the most accurate with 86% for the guilty, 93% for the innocent and 90% total. The inclusion of the ICAS-2 and ICAS-3 in this study resulted in the realization that with the ICAS method an even number of keys is most desirable. As seen with the ICAS-3, a majority of the keys, rather than only half must be judged to be RI in order for the test to be determined RI. This may have impacted the research results for this methodology.

	ICAS-4	ICAS-3	ICAS-2	Lykken	FCM	MSK	2 cut off
Guilty	86%	65%	86%	57%	50%	79%	50%
Innocent	93%	93%	79%	100%	86%	86%	93%
Total	90%	79%	83%	71%	68%	83%	72%

Raw Data

	Car	Door	Cover	Item	Total	Program	Lykken	ICAS-4	ICAS-3	ICAS-2	FCM	MSK	2 cut off
1	2,1,2	2,2,0	2,0,0	N/A	11	Guilty	RI	RI	RI	RI	RI	NA	RI
2	1,1,0	2,2,2	1,2,2	N/A	13	Guilty	RI	RI	RI	RI	NR*	NA	RI
3	0,0,0	1,1,0	1,2,2	1,2,0	10	Guilty	NR*	RI	NR*	NR*	NR*	RI	RI
4	1,0,2	0,1,2	2,1,1	1,2,2	15	Guilty	NR*	RI	RI	RI	NR*	RI	RI
5	1,2,1	0,0,0	0,2,2	2,2,2	14	Guilty	NR*	RI	RI	RI	NR*	RI	RI
6	0,2,1	0,2,0	0,0,0	0,0,0	5	Guilty	NR*	NR*	NR*	RI	RI	NR*	RI
7	1,2,2	0,1,0	2,0,2	2,0,0	12	Guilty	RI	RI	RI	RI	RI	NR*	RI
8	0,1,1	2,0,2	2,0,0	1,1,2	12	Guilty	RI	RI	RI*	RI	RI	RI	RI
9	0,1,0	0,1,2	2,2,2	2,2,1	17	Guilty	NR*	RI	RI	RI	RI	RI	RI
10	2,0,0	2,0,1	0,0,0	0,1,2	8	Guilty	NR*	RI	NR*	RI	RI	RI	RI
11	0,0,0	1,0,0	0,0,1	0,0,2	4	Guilty	NR*	NR*	NR*	NR*	NR*	NR*	NR*
12	2,2,2	0,2,2	1,1,2	2,1,0	17	Guilty	RI	RI	RI	RI	RI	NR	RI
13	2,1,1	2,2,1	2,2,2	2,1,2	20	Guilty	RI	RI	RI	RI	RI	NR	Ri
14	0,0,1	2,2,0	0,1,2	0,1,2	11	Guilty	NR*	RI	RI	RI	NR*	RI*	RI
15	1,0,0	1,1,0	0,0,0	1,2,0	6	Innocent	NR	NR	NR	NR	NR	RI*	RI*
16	0,1,1	0,0,2	0,1,0	0,0,1	6	Innocent	NR	NR	NR	NR	NR	NR	RI*
17	0,0,0	0,1,2	0,2,0	0,0,1	6	Innocent	NR	NR	NR	RI*	NR	NR	RI*
18	0,1,2	2,0,0	0,0,0	1,0,1	7	Innocent	NR	NR	NR	RI*	RI*	NR	RI*
19	0,1,1,	0,0,0	0,0,1	0,0,2	5	Innocent	NR	NR	NR	NR	NR	NR	RI*
20	0,0,1	0,0,2	0,0,0	0,0,0	3	Innocent	NR	NR	NR	NR	NR	NR	NR
21	0,0,0	1,0,0	1,1,0	1,0,0	4	Innocent	NR	NR	NR	NR	NR	NR	RI
22	0,2,1	2,1,0	0,2,0	0,0,0	8	Innocent	NR	RI*	RI*	RI*	NR	NR	RI*
23	1,0,0	0,0,0	0,0,0	0,0,0	1	Innocent	NR	NR	NR	NR	NR	NR	NR
24	1,0,0	0,0,0	0,0,0	0,0,0	1	Innocent	NR	RI*	NR	NR	NR	NR	NR
25	0,1,0	0,1,0	0,1,0	0,0,0	3	Innocent	NR	NR	NR	NR	NR	NR	NR
26	1,1,0	0,1,0	0,0,2	0,1,0	6	Innocent	NR	NR	NR	NR	NR	NR	RI*
27	0,0,0	1,0,0	0,0,1	0,0,0	2	Innocent	NR	NR	NR	NR	NR	NR	RI
28	0,0,0	0,0,0	0,0,0	2,0,2	4	Innocent	NR	NR	NR	RI*	NR	RI*	NR

Guilty = Recognition Indicated (RI).

Innocent = No Recognition (NR). Results printed in italics with an asterisk identify an incorrect decision.

Binomial Test

One Tail	Guilty	Innocent	
ICAS-4	P = .0065	P = .0009	P < .05
LYKKEN	P = .0898	P = .0001	

The Binomial test at P < .05 indicates that the accuracy of the ICAS-4 is better than chance for both innocent and guilty subjects.

Discussion

In evaluating accuracy of the CIT with the polygraph, Gaines (2013) found that his “232 system” resulted in an accuracy rate of 82% for guilty and 87% for the innocent with an overall accuracy rate of 84% with an N of 89. The “232” in the system’s name came from the weight assigned to each of the components in conducting the test data analysis. Since the ERS only utilizes the EDA sensor for evaluating data, the Gaines “232” evaluation system could not be evaluated as part of this study.

Study Limitations

Due to the gender and educational dynamics of the study population, which consisted of predominately Caucasian female graduate students, the ability to generalize the findings to a more diverse population is therefore at this point unknown. This concern may be alleviated through pending field trials conducted by a local police department.

The study design was such that the subjects were tested immediately after they committed the mock crime. This may have had some impact on memory of the items selected as question keys.

Both of the ERS analysts who conducted the examination for this study were highly experienced polygraph examiners who have conducted thousands of polygraph examinations to include CIT tests. Though the evaluation of EDA responses is quite simple, the ability of novice ERS analysts to achieve similar rates of success may not be as great.

Conclusion

At a 90% accuracy rate, the research results indicate that the ERS, when utilized with the CIT test and the ICAS-4 scoring methodology is a viable means of detecting recognition. What was surprising was that the accuracy of the ERS was demonstrated to be superior to the “232” scoring system, where Gaines (2013) analyzed all of the components of the polygraph in his study of the CIT. It is postulated that the greater number of charts conducted in the ERS ICAS-4 protocol resulted in the increased accuracy rate. This theory is consistent with the findings of Meijer, Verschuere, and Ben-Shakhar (2011).

Additional research using the ERS ICAS-4 scoring protocol in conjunction with a polygraph instrument is recommended as both the ERS and Gaines 232 test data analysis systems could then be evaluated side by side. If possible, any future ERS laboratory research should also include a more gender and racially diverse study population. Plans are currently underway to have a local police department conduct a field study of the ERS. This research would seek to determine the accuracy and utility of the ERS in support of actual criminal investigations.

The ERS is of interest to the polygraph profession as while not able to replace the polygraph, it supplements it in criminal investigations. Polygraph examiners assigned to local police departments may also be utilized to conduct quality control of CIT examinations conducted by ERS analysts and supervise their operations.

The authors would like to express their appreciation to Richard Cooter J.D. Psy.D. Director of the Forensic Psychology Program at George Washington University, for providing administrative support, office space and allowing his graduate students to participate in this study.

References

Ben-Shakhar, G. (2016). *Psychophysiological Detection of Deception:A review of the recent research and potential forensic applications.* The Psychology of Crime Policing and the Courts edited by Andreas Kapardis, David P. Farrington Routledge Press

Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the guilty knowledge technique in court. *Law and Human Behavior*, 26 No. 5

Ben-Shakhar, G. & Freudy, J. (1990). Theories and Applications in the Detection of Deception. New York Springer-Verlag New York Inc.

Bureau of Labor. (2016). *Occupational Employment and Wages, May 2016 33-3021 Detectives and Criminal Investigators.* Retrieved 6 June 2017 from https://www.bls.gov/oes/current/oes333021.htm

Elaad, E. & Ben-Shakhar, G. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination*. Journal of Applied Psychology,* 88

Elaad, E. & Ben-Shakhar, G. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A Meta-Analytic Review. *Journal of Applied Psychology,* 88 No.1

Gaines, K. (2013). Utility and Numerical Evaluation of the Guilty Knowledge Test. *Polygraph,* Vol 43 No.3

General Services Administration. (2017). *FY 2017 Per Diem Rates for ZIP 20176.* Retrieved June 6, 2017 from https://www.gsa.gov/portal/category/100120

Handler. M., Nelson, R. Krapohl, D. & Honts. C., (2010). An EDA Primer for Polygraph Examiners. *Polygraph,* 39 No. 2

Hira, S. & Furumitsu, I. (2009). Tonic Arousal During Field Polygraph Tests in Guilty vs. Innocent Suspects in Japan. *Applied Psychophysiology and Biofeedback,* 34 No. 3: 173-6

Kaplan, R., Van Damme, I., & Levine, L. (2012). Motivation matters: differing effects of pre-goal and post-goal emotions on attention and memory. *Frontiers in Psychology, 3*(404), 1-9.

Kraypohl, D., McCloughan, J. & Senter, S. (2009). How to use the Concealed Information Test *Polygraph,* 38 No.1

Levine, L. & Edelstein, R. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion, 23*(5), 833-875

Lykken, D. (1959). The GSR in the Detection of Guilt. *Journal of Applied Psychology,* Vol. 43, No. 6

Lykken, D. (1998). *Tremor in the Blood.* Perseus Publishing Reading, MA

MacLaren, V. (2001). A Quantitative Review of the Guilty Knowledge Test. *Journal of Applied Psychology,* Aug 86 No.4 647-83

Meijer, E., Verschuere, B., & Ben-Shakhar, G. (2011). *Practical Guidelines for Developing a CIT*. Retrieved January 20, 2018 from https://www.researchgate.net

National Research Institute of Police Science. (2009). *The Polygraph Examination in Japan.* Unpublished slide presentation

Verschuere, B., Crombez, G., DeClercq & Koster, E. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology*, 41

Credibility Assessment Using Bayesian Credible Intervals: A Replication Study of Criterion Accuracy Using the ESS-M and Event-Specific Polygraphs with Four Relevant Questions

Raymond Nelson[1]

Criterion accuracy was evaluated for event-specific polygraph exams with four relevant questions. The sample included n=15 innocent cases and n=15 guilty cases, selected from an archive of confirmed field cases that was compiled by the Department of Defense in 2002. All cases employed relevant questions that described both direct and indirect involvement in the crime under investigation. Physiological responses were extracted from recorded computer software designed to execute the scoring procedures described in the published literature. Numerical scores were assigned using the Empirical Scoring System. A multinomial likelihood function was used to calculate a statistical value for the numerical scores. The cases were classified as either deceptive, truthful or inconclusive using two-stage decision rules and a naïve-Bayes classifier for which the 95th percentile limit of the credible interval for the posterior odds of deception or truth-telling was calculated using the Clopper-Pearson method. They were classified as deceptive when the 95th percentile limit of the credible interval for the posterior odds of deception exceeded the prior odds of one to one. Similarly, the samples were classified as truthful when the 95th percentile limit of the posterior odds of truth-telling exceeded the prior odds. Results for two of the sample cases (7%) were inconclusive because the 95th percentile limit of the posterior odds did not exceed the prior odds. Correct classifications were made for 93% of the 28 cases where the posterior odds were statistically significant (where the 95th percentile limit of the credible interval exceeded the prior odds). Test sensitivity to deception was observed at .87, and test specificity to truth-telling was also observed at .87. These results are consistent with previously published descriptions of event-specific examinations with four relevant questions.

Previous publications have described the structure of event-specific examinations with four relevant questions, known to field examiners as variants of the MGQT format, including the AFMGQT version 1 and version 2 (Department of Defense, 2006a, 2006b), and the Utah four-question format (Handler & Nelson, 2008; 2009), known to some field examiners as the “Raskin technique” due to his role in the development of this approach. These examinations can consist of both primary relevant questions that describe an examinee’s direct involvement in the issue under investigation, along with secondary relevant questions that attempt to describe an examinee’s indirect involvement or level of involvement. These examinations are traditionally interpreted with an assumption of independent criterion variance. However, previous studies have not supported the validity of the independence hypothesis for these examination for-

[1] Raymond Nelson is one of the developers of the Empirical Scoring System (ESS) and Objective Scoring System, version 3 computer scoring algorithm, and has published numerous studies on all aspects of the polygraph test. Mr. Nelson is a polygraph field examiner and psychotherapist with expertise in sexual offending, victimization, trauma and development in addition to other experience in testing, data analytics and statistics. Mr. Nelson serves as an expert witness in legal matters involving both polygraph and psychology/psychotherapy. Mr. Nelson is a past president, and currently elected member of the APA Board of Directors, and has helped with policy development at the state, local and national level. Mr. Nelson is a research specialist with Lafayette Instrument Company, which develops and markets polygraph technologies. The views an opinion expressed herein are those of the author and not the APA or LIC.

mats (Barland, Honts & Barger, 1989; Podlesny & Truslow, 1993).

In response to inquiry from field practitioners, Nelson, Handler, Oelrich and Cushman (2014) described the rationale for a more generalized usage of polygraph test formats for event-specific diagnostic exams. Nelson and Handler (2017) described a general procedural rationale for the selection of a test formats for screening and diagnostic exams, indicating that while statistical multiplicity may lead to reduced precision for screening with more relevant questions, precision of diagnostic exams may be increased through the use of more relevant questions and the acquisition of a greater volume of test data.

Raskin, Honts, Nelson & Handler (2015) reported the results of a Monte Carlo analysis of these examinations, and suggested that criterion accuracy with four relevant questions can equal or exceed that or other examination formats when these exams are evaluated with an assumption of non-independent criterion variance. The present study was designed as a replication, in an attempt to increase the available published information about four relevant question event-specific polygraphs, including: test sensitivity, specificity, false-negative and false-positive errors, and inconclusive results using confirmed field cases.

Data

Examination data for this study were obtained from an archive of confirmed field polygraph exams that was compiled at the Department of Defense in 2002. Cases consisted N=30 confirmed field polygraph examinations. There were n=15 confirmed innocent examinations of this type in the confirmed case archive, along with a random selection of n=15 matching confirmed deceptive cases. Investigation target issues included: theft/larceny, murder, sexual assault, aggravated assault, false statements/false swearing, arson, robbery, child abuse, fraud, and illegal drugs. Archival data indicate that all sample cases were confirmed by information other than examinee confession. All examinations consisted of four relevant questions presented in a sequence with other questions designed to elicit responses that can be compared with responses to the relevant test stimuli, in addition to other procedural questions. Relevant questions included a combination of questions about direct involvement and indirect involvement in the issue under investigation. An important characteristic of contemporary field polygraph test formats is that all relevant questions subject to numerical evaluation are presented subsequent to the presentation of at least one comparison stimuli. Another important characteristic is that responses to each of the relevant stimuli are evaluated using the comparison stimulus immediately preceding and immediately subsequent to the relevant stimuli depending on which comparison stimuli has elicited the greater change in physiological activity.

Analysis

Data for each case was exported to a structured ASCII text format, including time-series data for all recording sensors, along with event markers indicating the onset, end and verbal answer for all test stimuli, along with other annotations. Data were imported to the R statistical computing environment (R Core Team, 2017) for signal processing and feature extraction. Response features were those described in previous publications (Krapohl & McManus, 1999; Nelson, Krapohl & Handler, 2008). Those features include: amplitude of increase in EDA, amplitude of increase in blood pressure, and suppression or reduction of respiration activity. Numerical scores were assigned to each of the sensors for each stimulus presentation using the Empirical Scoring System (Nelson et.al., 2011).

Posterior odds of deception or truth-telling were calculated for each case using a multinomial likelihood function for ESS scores (ESS-M) and a naïve-Bayes classifier (Nelson, 2017). The 95^{th} percentile one-tailed limit of the Bayesian credible interval was calculated using the Clopper-Pearson method. Classifications of deception or truth-telling were made using two-stage rules (TSR; Senter, 2003; Senter & Dollins, 2003). The TSR requires that cases would be classified as deceptive when the 95th percentile limit of the credible interval for the posterior odds of deception exceeded the prior odds of one to one using the grand total score. Similarly, cases would be classified as truthful when the 95th percentile limit of the posterior odds of truth-telling exceeds the prior odds using the grand total score. When results are inconclusive using the grand total score, the TSR would permit a deceptive classification if the 95th percentile limit of the multiplicity-corrected posterior odds of deception for the lowest subtotal score has exceeded the prior odds. Cases would be unclassified, and therefore inconclusive, when 95th percentile limits of the grand total and lowest subtotal score have not exceeded the prior odds.

Results

The mean score for innocent cases was 13.3 (sd=10.0), and the mean score for guilty cases was 14.1 (sd=12.7). Results with the naive-Bayes classifier and ESS-M scores are shown in Table 1. Two cases were inconclusive, including one guilty and one innocent case. In addition one of the innocent cases was incorrectly classified as deceptive, and one of the guilty cases was incorrectly classified as truthful. Twenty-six of the cases were classified correctly. A detection efficiency coefficient (Kircher, Horowitz & Raskin, 1988) was calculated in order to provide a single statistical metric to encompass correct, incorrect, and inconclusive results with both guilty and innocent cases. The detection efficiency coefficient was .83.

Excluding inconclusive results, 93% of the decisions from the naive-Bayes ESS-M classifier were correct. Several metrics of classification accuracy were calculated, including test sensitivity to deception, specificity to truth-telling, false-negative and false-positive errors, inconclusive results, and unweighted criterion accuracy. Confidence intervals, shown in Table 1, were calculated for all metrics using a parametric bootstrap.

Table 1. Criterion accuracy of ESS-M scores of event-specific exams with four relevant questions.

Unweighted accuracy	.93 {.87 to .98}
Unweighted inconclusive	.07 {.02 to .12}
Sensitivity	.87. {77. to .95}
Specificity	.87. {77. to .95}
False negative	.07 {<.01 to .14}
False positives	.07 {<.01 to .14}
Guilty inconclusive	.07 {<.01 to .14}
Innocent inconclusive	.07 {<.01 to .14}

Results for two of the sample cases (7%) were inconclusive because the 95th percentile limit of the posterior odds did not exceed the prior odds. Correct classifications made for 93% of the 28 cases where the posterior odds were statistically significant (where the 95th percentile limit of the credible interval exceeded the prior odds). Test sensitivity to deception was observed at .87, and test specificity to truth-telling was also observed at .87. These results are consistent with previously published descriptions of event-specific examinations with four relevant questions. Incorrect classifications were made for two of the sample cases, including one innocent case (7%) and one guilty case (7%).

Discussion

This project was an attempt to replicate previous work on event-specific diagnostic polygraphs with four relevant questions. This project also replicates previous work involving the use of a multinomial referenced distribution and naive-Bayes classifier for ESS-M scores. Results from this study are consistent with other reported results involving event-specific polygraphs with four relevant questions evaluated with an assumption of non-independent criterion variance.

To further investigate the differences between the traditional approach to these examinations and results using an evidence-based statistical classifier, the detection efficiency coefficient and results were re-calculated using the subtotal-score rules (SSR) and traditional numerical cutscores. Traditional numerical cutscores for these examinations are +3 for truthful classification of the subtotal scores, and -3 for deceptive classification of the subtotal scores. The SSR requires that all subtotal scores exceed the traditional numerical cutscore in order to classify a case as truthful, while any deceptive subtotal score would result in a classification of the case as deceptive. The SSR does not permit both truthful and deceptive decision within a single exam. The detection efficiency coefficient using the SSR and traditional cutscores was .75. Of the 30 cases, 10 (33%) were inconclusive using the SSR and traditional cutscores, including 1 guilty case and 9 innocent cases. Correct classifications of deception and truth-telling were made for 14 of the guilty cases (93%) and 4 of the innocent cases (27%). The unweighted accuracy, excluding inconclusive results, was .83 using the TSR and traditional cutscores.

Results from this project suggest that decision accuracy could benefit substantially from a change from the traditional decision rules to others for which published evidence has found better performance.

Traditional approaches using the SSR for the interpretation of polygraph tests that use a combination of four primary and secondary relevant questions are known to produce accuracy rates that underperform compared to other well-known testing approaches. Interpretation of these exams using the evidence-based TSR and cutscores that are informed by sound statistical theory can produce classification accuracy rates that may equal or exceed that of other highly-regarded polygraph formats for event-specific diagnostic exams.

Like all projects, this project is not without limitations. Among the obvious limitations herein, is the small sample size. Additionally, incomplete information was available regarding the examinee demographics, and no information is available concerning how the sample cases came to be included in the confirmed case archive. These limitations notwithstanding, the present results support a recommendation for continued interest in the four relevant question event-specific format for field practitioners and researchers within the polygraph profession. In addition, these results support continued interest in the TSR and the ESS-M naive-Bayes model for statistical quantification and classification of polygraph test results.

References

Barland, G. H., Honts, C. R. & Barger, S.D. (1989). Studies of the accuracy of security screening polygraph examinations. DTIC AD Number A304654. Department of Defense Polygraph Institute.

Clopper, C. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 26, 404–413.

Department of Defense (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retrieved from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02- 10-2006.pdf on 3-31-2007. Reprinted in Polygraph, 40(1), 2-66.

Department of Defense (2006). Psychophysiological Detection of Deception Analysis II -- Course #503. Test data analysis: DoDPI numerical evaluation scoring system. Available from the author. (Retrieved from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf on 3-31-2007).

Handler, M. & Nelson, R. (2008). Utah approach to comparison question polygraph testing. European Polygraph, 2, 83-110.

Handler, M. & Nelson, R. (2009). Utah approach to comparison question polygraph testing. Polygraph, 38(1), 15-33.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

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

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

Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. Polygraph & Forensic Credibility Assessment, 46 (2), 81-115.

Nelson, R. & Handler, M. (2017). Practical polygraph: how to select a polygraph test format. APA Magazine 2017, 50 (2), 72-81.

Nelson, R., Handler, M., Oelrich, M. & Cushman, B. (2014). APA Research Committee Report:proposed Usage for an Event-specific AFMGQT Test Format. Polygraph, 2014, 43(4), 155-167.

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

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

Podlesny, J. A. & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. Journal of Applied Psychology, 78, 788-797.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raskin, D. C., Honts, C., Nelson, R. & Handler, M. (2015). Monte Carlo estimates of the validity of four relevant question polygraph examinations. Polygraph, 44(1), 1-27.

Senter, S. (2003). Modified general question test decision rule exploration. Polygraph, 32, 251- 263.

Senter, S. M. & Dollins, A. B. (2003). New Decision Rule Development: Exploration of a two-stage approach. Report number DoDPI00-R-0001. Department of Defense Polygraph Institute Research Division, Fort Jackson, SC. Reprinted in Polygraph, 37(2), 149-164.

AMERICAN POLYGRAPH ASSOCIATION

DEDICATED
TO
TRUTH