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

VOLUME 51 | 2022 | 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

******

Associate Editors: Réjean Belley, Ben Blalock, Tyler Blondi, John Galianos, Don Grubin, Maria Hartwig, Charles Honts, Matt Hicks, Scott Hoffman, Don Krapohl, Thomas Kuczek, Mike Lynch, Ray Nelson, Adam Park, David Raskin, Stuart Senter, Joseph R. Stainback IV and Cholan V.

Subscription information: *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice* is published semi-annually by the American Polygraph Association. Editorial Address is Editor@polygraph.org. Subscription rates for 2022: One year $150.00. Change of address: APA National Office, P.O. Box 8037 Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice* DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

Instructions to Authors

Submission deadline for next publication: July 31, 2022

Scope

The journal *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice* 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.

Jan Widacki's *History of Polygraph Examination*

Warsaw: Polskie Towarzystwo Kryminalistyczne 2021, 178 pages, indexed, illustrations

Donald J. Krapohl

Polygraph professionals may recognize Jan Widacki as the Editor-in-Chief of the journal *European Polygraph*. Those who read more widely may recognize that he is perhaps one of the most well-known and prolific writers on polygraph in all of Europe. A lucky few would be aware of his many accomplishments as an attorney, university professor, and diplomat. While most of his papers were published in his native Polish language, American examiners should recall a 1978 English-language paper by Professor Widacki and Dr. Frank Horvath in which they conducted a groundbreaking study comparing the polygraph with several other forensic method. Now Professor Widacki has produced a textbook in English titled *History of Polygraph Examination*.

In the initial chapter Professor Widacki explores the definition of lying, whether by direct deception, concealment, omission, misdirection, in contrast to mental defect or genuine failures of memory. He also surveys early historical accounts of attempts to identify liars through patterns of behavior and follows up with more recent and current methods that are based on similar premises.

Chapter 2 takes a deep dive into the roots of psychology and physiology that laid the groundwork for the rise of the polygraph. Though the field of polygraph has many contributors, Widacki has provided a direct academic lineage. As all polygraph examiners are aware, John Larson was the first to use multiple channels of physiological recordings and a modern testing approach for lie detection. Larson credited the basis of his work on earlier research by William Marston, who was a student of Hugo Münsterberg, who himself studied under Wilhelm Wundt. Wundt is one of the fathers of experimental psychology who has perhaps one of the most enduring legacies in his field. Widacki has revealed the relatively unrecognized relationship between Wundt and the ultimate development of the polygraph. Chapter 2 also reveals more details on historical figures well known to polygraph examiners, including Mosso, Benussi, Féré as well as the contributions of lesser recognized scientists in eastern Europe and Russia such as Abramowski and Tarkhanov.

In Chapter 3 Widacki examines the array of early methods that were aimed at deception detection. Satisfactory coverage is given to word association tests and several physiological measures including breathing, EEG, blood pressure, electrodermal responses, and muscle tension among others. While polygraph examiners may be familiar with part of this history, Widacki has provided a more informative treatment than many would have seen before.

The first practical use of instrumental lie detection is the focus of Chapter 4. It concentrates almost entirely on the American era from about 1915 to just after World War II. A thorough discussion is offered for the work of Marston, Larson, Keeler, Summers and Reid along with the context in which their contributions took place. The chapter includes the case histories of noteworthy polygraph exams in the early years up to the testing of German prisoners of war near the end of the Second World War, including the test questions.

In Chapter 5 Widacki picks up from the Second World War. In this portion of the text Widacki covers the evolution of modern polygraph techniques. In particular, the contributions and relative merits of the approaches advocated by innovator John Reid and psychophysiologist David Lykken are taken up in detail.

Chapter 6 begins with a brief history of Cleve Backster and his approach to polygraphy. In the following pages Widacki reports on the results of a selection of research studies mostly from the 1960s and 1970s, the profound influence of Cleve Backster on field practices, improvements in polygraph instrumentation, introduction of voice-based systems, and he finishes the chapter with a summary of the movement toward professional standards.

The final chapter of *History of Polygraph Examination* is dedicated to the history of the polygraph in Widacki's home country of Poland, with a smaller section on developments in a handful of other countries. He is rightfully proud of the work of other Polish scientists and writers whose work is often overlooked. Readers are given a glimpse of their contributions, as well as the intersection between Polish and American experts in the 1970s.

In sum, *History of Polygraph Examination* is one of the most complete, scholarly, well organized and thoroughly sourced work on the subject. It is richer in background for the many European contributions to the field than traditional histories coming from the US, which it also covers. The pages are well supplemented with photos and anecdotes that give greater life and depth to the topic. The book was originally released in Polish in 2017. The 2021 English-language version being reviewed here was well translated, with the rare typographic error or unwieldy language (e.g., unfortunately among them, the book's title). It is an easy and interesting read and strongly recommended for the bookshelves of all polygraph examiners, and an essential part of the literature holdings of polygraph schools.

Psychophysiological Detection of Deception in Children and Adolescents

Zuzana Michalicova

Department of Applied Psychophysiology, Institute of Forensic Science of the Police Force, the Presidium of the Police Force of the Ministry of Interior of Slovak Republic

Abstract

The psychophysiological detection of deception in children and adolescents represents an even more controversial area of the already controversial viewpoint on the polygraph in the adult population. Although the development of lying in children itself is substantiated by developmental psychology, the opinions on the child's position in the process of detection of deception and the final reliability of the examination are not clear-cut. The insufficient research in this field is often pointed out, nevertheless. The institutions concerned do not create space for their implementation, and they often incorrectly argue on ethical aspects as an obstacle to conducting research. This paper aims to give a holistic picture of the psychophysiological detection of deception in children and adolescents together with all aspects, which may be involved in the psychophysiological examination, i.e., they affect it in any way possible. We pay attention to the cognitive skills and age specification; we reviewed the theoretical background and current situation. We would like to stimulate interest in any activity related to theory or practice. Regarding the research, we believe it is a matter of time when an increasing number of institutions concerned will consider - in specific cases - the involvement of children and adolescents in the investigation process through the detection of deception.

Keywords: *Psychophysiological Detection of Deception, Children, Lie, Polygraph*

Author Note

We have no conflicts of interests to disclose.

Correspondence concerning this article should be addressed to Department of Applied Psychophysiology, Institute of Forensic Science of the Police Force, Ministry of Interior of Slovak Republic, Sklabinska 1, 812 72 Bratislava. Email: zuzana.michalicova@minv.sk

Psychophysiological Detection of Deception in Children and Adolescents

The history of lying is probably as long as human history itself. It exists as a part of our internal world. We do not know where it starts nor where it ends. However, we know that it helps us to "survive." Lying, hiding of information, falsifying of information, prettifying of information, or however we would like to call it, accompanies us from early childhood. We do not know who taught us to do that, but we soon realized that lying brings us certain benefits.

Although society generally accepts lying, there are some situations in which lying is destructive from the point of view of the legitimate interests of other persons or the whole society. The desire for justice and for rightful retribution motivated people long before Christ to seek ways how they can prove which person is guilty of an offense, and impose a sentence on them. When people started to realize their limits in revealing the lie and they ceased to find their intuition dependable enough, they started to look for more scientific ways to find out whether the person is lying or, on the other hand, whether they can trust the testimony of the accused. While this new approach to detecting lies began to enjoy at least partial popularity, it had to be developed from its unscientific beginnings, through the first scientific experiments, to current targeted scientific research.

Psychophysiological detection of deception, commonly known as "lie detection", is together with other forensic and psychological methods involved in uncovering the circumstances and causes of various forms of undesirable human behavior. The task of psychophysiological detection of deception is simple, thus, to find out whether the person under investigation is telling the truth or not concerning the given verified act. Although it sounds really simple, the approach for the detection of deception is much more complex than the first glance may suggest.

The continuous progress in the field of research of the psychophysiological detection of deception has made the polygraph one of the methods that can verify the truthfulness of the examined person with high reliability and accuracy concerning a specific event. The scientific literature currently reports the mean decision accuracy to be over 90% (Nelson & Handler, 2013). This development has significantly contributed to the fact that detection of deception has become an essential investigative tool in both the criminal and private sectors. Many state and federal institutions around the world have applied it to their investigative practice.

The current state and knowledge in the field of psychophysiological detection of deception are almost entirely limited to adult test population. Originally, methods of recording physiological manifestations were used to look for signals that indicate a possible lie in adults. Only later did the discussion about the possibilities of use with children start. This is mainly related to the mentioned complexity of the examination, which presupposes and requires the presence of a certain level of cognitive capacity and mental maturity. Regarding cognitive maturity, the very question of lying comes to the fore in a psychophysiological examination. The youngest children have no or very little understanding of the nature of the lie itself (Piaget, 1948). To be able, at least on a theoretical level, to talk about the possibility of involving the child in this process, the prerequisites are: the achieved mental processes at such a developmental stage which allow the use of cognitive functions to properly grasp the essence of a lie, or more precisely the truth; the ability to distinguish true information from false information, and, the ability to separate reality from fantasy. Taking this aspect into account, together with other developmental factors, is one of the essential steps in assessing the subject's suitability for this type of examination (APA, 2021).

Development of Lying in Children

My son was about three years old when I first noticed he was lying. Despite being a psychologist, I was overcome by my inner uneasiness, and not very positive scenarios about how his social direction might develop emerged in my mind. I quickly reached for the most available information about lying in childhood and learned that lying in early childhood is a manifestation of a child's intelligence and developing cognitive abilities. I accepted his newly discovered ability as a necessary stage of development. Paul Ekman (1991) in his book *Why Kids Lie* said that although he has been dealing with the question of deception professionally for twenty years, it was not easy to deal with it as a parent. Children can lie, and they can lie much earlier than we adults may think. According to Jean Piaget (1948), a Swiss psychologist who, in his theory of cognitive development, elaborated the aspect of deception in children. Piaget asserted that to the child`s mind, lies represent "a far graver and more pressing problem than do clumsiness or even such exceptional actions as stealing. This is due to the fact that the tendency to tell lies is a natural tendency, so spontaneous and universal that we can take it as an essential part of the child´s egocentric thoughts" (p. 135). A common reason why lying in children is generally considered to be a negative behavior is that it often occurs in connection with other unwanted or maladaptive behavior. Stouthamer-Loeber (1986) and Stouthamer-Loeber and Loeber (1986) in their work specifically addressed the relationship between lying and antisocial behavior and other conduct problems in older children, from the fourth grade onwards (etc. thefts, fights, etc.). According to their findings, the prevalence of children lying did not increase with age. At the same time, their data did not show how lying relates to other problem behavior in terms of time. The question remains whether deception prevents behavior problems or whether it is just a necessary accompaniment to cover up such undesirable behavior. According to comprehensive and complex studies of Hartshorn and May (1928, as cited in Burton, 1963), the consistency in this kind of behavior from one situation to another is caused by the similarities between the situations and not by a consistent personality trait. According to this theory, the emphasis lies on the situational factors of development of the deceit, and the deception itself does not relate to the child's character trait (Ekman, 1991).

However, many studies consistently confirm that the initial emergence and development of childrens' lie-telling is closely associated with their developing cognitive abilities (Evans & Lee, 2013; Talwar, 2018; Talwar & Crossman, 2012; Talwar & Lee, 2008a). The fact of how children understand the meaning of lying and how the concept of deception itself changes with age, represent the significant role in discussing children's lies.

Lying is commonly defined as a statement communicated with the intent to mislead another person into believing something to be true that the lie-teller believes to be false. That is, there is discrepancy between what is said by the lie-teller and what is known to be true (e.g., Ekman, 1991, 2009; Talwar & Lee, 2008a). The aim is to get to the particular result by deceiving another person (Stern, 2018). At the same time, Stern (2018) mentions that an awareness of the falsehood of information, intentional deception, and unequivocal intent are features of a true lie that distinguish it from other false statements. The first two features clearly distinguish a lie from memory errors and the third one distinguishes it from imaginary thoughts. The presence of all three features presupposes a relatively high level of mental development. At the same time, if we want to be a successful liar and effectively inculcate a false statement in the mind of the recipient, we must control our expressive behavior. This behavioral control must be exerted not only during the initial lie statement, but also throughout all related subsequent conversations about the topic (e.g., DePaulo & Jordan, 1982; Ekman, 2009; Talwar, 2018). Ekman (2009) differentiates two main techniques of lying: concealing and falsifying. By concealing the person keeps some information to themselves without saying anything that is not true. By falsifying the person also presents false information as true. Lying is a relatively complex task that requires a sophisticated level of cognitive behavior (Talwar, 2018).

The first scientific studies focused on the development of lying began to appear at the turn of the twentieth century when Developmental Psychology was established. At that time, regarding the prevalence of lying in everyday life, society began to emphasize the promotion of veracity in children's behavior and deception became a point of interest for parents, teachers, and professionals, who worked with children on a clinical or forensic basis. Despite the omnipresence of deception in everyday life, lying was considered a negative and even punishable behavior. In human interactions, it represented a loss of trust and credibility and posed an increased risk for the development of antisocial behavior. Therefore, research focusing on how this behavior arises and develops in children, and the factors that influence its manifestations as a social strategy began (Talwar, 2018).

The German psychologist William Stern (1871-1938) is one of the first leading figures, who in their work focused on the research of lying in children. Based on the studies of his three children, Stern and with his wife documented the nature and diversity of childish lies (Stern & Stern, 1999; Stern, 2018). Jean Piaget (1948) later systematically studied the manner in which children judge and evaluate lying. Following his lead, the little subsequent research that has been done on the development of lying in children has focused primarily on understanding and evaluating the lies (Peterson, Peterson & Seeto, 1983; Piaget, 1948; Siegal & Peterson, 1996; Strichartz & Burton, 1990). Only later did studies begin to focus on the process of lying itself and the cognitive abilities necessary for lie-telling behavior (Evans & Lee, 2013; Talwar, 2018; Talwar & Crossman, 2012; Talwar & Lee, 2008a).

According to Piaget's theory, "the problem of lies is the clash of the egocentric attitude with the moral constraint of the adult" (Piaget, 1948, p.135). As part of his moral theories in the child's development, he analyzed the child´s consciousness of lies, more precisely, the manner in which younger children judge and evaluate lying. Based on his observations and the statements of children up to six years of age, he defined a lie as "making a moral fault by means of language" (Piaget, 1948). Between the ages of six and ten, children see lies simply as something that is not true, making no distinction for unintentional errors of fact. Children between the ages of five and seven, who are more or less aware of the shade of difference between an intentional act and involuntary mistake, tended not to stress this distinction at all and called both types of statements a lie. The awareness of the intentional nature of the act appears with the first "why" questions and this is around the age of three, although not to the same extent as in adults. The identification of error and deception disappears sometime around the age of eight, around the time as disappears the vast majority of the phenomenon of animism and artificialism, and also at the time of losing the other signs of inability to distinguish between the idea of the intentional act and involuntary behavior. In this period, intentional behavior and unintentional error are still not separated at the level of moral reasoning. Only at the age of ten or eleven - "any statement that is intentionally false is a lie" (Piaget, 1948, p. 142). According to Piaget's theory (1948), we would not find a clear form of this definition of deception before the age of 10 or 11.

Children's conceptual and moral understanding of deception increases with coming years, as the child improves the evaluation of the intentional component of the lie. Evidence of direct deception in children comes from observations and experimental studies. Many studies on children's conceptual understanding of truth and deception followed Piaget's work. In general, it has been shown that the recognition between telling the truth and lying appears in the early preschool years. Children under the age of four, and perhaps earlier, can lie (Ekman, 1991; Bussey 1992; Evans & Lee, 2013; Newton, Reddy & Bull, 2000; Strichartz & Burton, 1990). Strichartz and Burton (1990) noted that a four-year-old child developed a set of qualities that define the concept of truth and deception. The child attributes the label "lies" to statements that do not match the facts, without added intention and belief system. Bussey (1992) and Siegal and Peterson (1996) also noted that even a four-year-old child was able to successfully identify deception and evaluate the intention with which the lie was told (to protect or harm another person). According to some studies, some children start to use deception as early as two years of age (Evans & Lee, 2013; Newton, Reddy & Bull, 2000; Wilson, Smith & Ross, 2003). Regarding the question of deception in young children, it is generally believed (Piaget, 1948; Stern, 2018; Strichartz & Burton, 1990) that the first lies in preschool children are erroneous or impulsive statements rather than truly deliberate cases of deception and intentional and conscious betrayal of the truth. It is difficult for a 3-year-old child to understand that giving false information can affect the beliefs of others and subsequently change their behavior (Sodian, 1991). According to Strichartz and Burton (1990), this does not mean that children at this age are unable to distinguish between the belief system and intention, but they do not consider them to be important aspects of their definition of deception when compared to real facts. Only between the ages of 6 and 10, when assessing the veracity of information, do they start to emphasize the belief system, and thus their purely fact-based trust begins to change. According to Stern (as cited in Piaget, 1948), until the ages of 7 to 8, a child has systematic issues in adhering to the truth. Without actual lying due to the deception itself (e.g., without the intention of deceiving someone, and even without realizing what they are doing) they distort the reality according to their desires and fantasies. Children's ability to lie is significantly increased with age, and in middle childhood, children, as skilled liars, can control both their verbal and non-verbal behavior so that they are not easily detected.

Experts and scientists who have dealt with deceptive behavior in children created the so-called experimental paradigm or temptation-resistance paradigm to analyze children's abilities of spontaneous deception (Neville, 2018; Sears, Rau & Alpert, 1965, as cited in Talwar, 2018; Wang & Wang, 2018). In the controlled laboratory conditions, the child is allowed to commit a transgression and was subsequently asked questions about his behavior. A common modified version of this research in preschool children was that during a game where children were tasked with recognizing a toy based on a sound, they were instructed not to look at the toy when the experimenter suddenly had to leave the room. Upon their return, the experimenter asked them if they had looked at the toy. Studies using this paradigm have discovered similar developmental patterns in terms of telling lies to cover their transgression and their subsequent ability to maintain lying in the following statements (Talwar, 2018; Talwar & Lee, 2008a, 2008b). To better understand the age-related differences in the complexity and intricacy of deception in children, experts used this paradigm to closely examine cognitive factors, which were apparently closely related to concealment and deception. Recent empirical studies have found that parts of the complicated cognitive functions, which play a role in improving children's lies, are the theory of mind understanding and executive functioning (Talwar, 2018; Talwar & Lee, 2008a).

Theory of mind understanding is generally defined as the ability to attribute mental states to oneself and others and to recognize that the beliefs, desires, and intentions of others may differ from one`s own. A key concept concerning deception is an understanding of a false belief or an understanding that another person may have a false belief about the true state of events that is different from ours. A prerequisite for telling a successful lie is to have privileged knowledge of the state of the given event (e.g., whether we looked at a toy) to which the recipient of the lie does not have access (Talwar & Lee, 2008a). Research shows that while most of the 4-year-old children understand that others may have false beliefs about the true state of events, many 3-year-old children are unable to understand this and are therefore unable to lie deliberately (Sodian, 1991; Talwar & Lee, 2008a). Talwar (2018) in her research, which focused on the deception process itself, noted that most children at the age of four lie willingly to cover up their transgression. Most of these early lies reflect the ability to represent a false belief that is different from the belief of an actual state or event. At this stage, however, children have difficulties with maintaining a lie. Their subsequent statement, which follows the initial false statement, tends to be inconsistent with the initial lie, and their lie can be therefore easily detected by an adult. The more developed awareness achieved at six to seven years of age has resulted in the successful maintenance of false denials and consistent responses to other questions that followed the initial lie. At this stage, children acquire the ability to maintain the credibility of the lie. The child's ability to maintain consistency between their initial lie and subsequent verbal expression increases with age.

Executive functioning points to a group of psychological processes with higher requirements that serve to monitor and control thoughts and actions. It includes cognitive functions including self-regulation, inhibitory control, planning, attention flexibility, and working memory (Talwar, 2018). These functions begin to appear in late infancy and develop during childhood. They are part of the decisive factors in the whole child's learning process (Raver & Blair, 2016). The inhibitory control and working memory are also directly connected with deceptive behavior in children. The inhibitory control refers to the ability to suppress or stop interfering thought processes and actions, and to adapt behavior to a change in goal or plan (Anderson & Weaver, 2009). The working memory is a system for temporarily holding and processing information in the mind in a controlled manner (Robert, Borella, Fagot, Lecerf & Ribaupierre, 2009). When a child chooses to lie, they must suppress the truth while simultaneously representing and communicating false information that differs from reality to avoid detection. To maintain the lie, they must carefully suppress behavior that may contradict their false statement and at the same time keep in mind the content of their lie (Talwar, 2018; Talwar & Lee, 2008a). The findings of studies on the relationship between executive functioning and deception show that children have difficulties with deception when they lack advanced skills in executive functioning, namely inhibitory control and working memory (Carlson, Moses & Hix, 1998; Talwar & Lee, 2018).

The present empirical conclusions offer us a clear developmental picture of deception in childhood. The ability to tell lies begins to appear in children in the early preschool years and continues to develop into middle childhood. It undergoes a gradual progression, which can be divided into three stages of development of deception (Talwar & Lee, 2008a, 2008b). We talk about a “primary lie” in the period of 2 to 3 years when children can deliberately make false statements for the first time. In these cases, it is not entirely clear whether it is a word play, the fulfillment of hidden desires, or self-protection, but according to available information, children at this age have not created a real concept of lies yet and lies at this stage do not represent conscious and purposeful lies in the true sense of the word (Piaget, 1948; Stern, 2018; Strichartz & Burton, 1990; Talwar & Lee, 2008b). The second stage, a “secondary lie,” points to a significant cognitive shift, which takes place in the third and fourth year of age. By the age of four, most children can tell a lie to conceal their transgression. However, their cognitive abilities are not developed enough to control semantic leaks, their further statements are not consistent with the initial lie, and they are therefore quite easy to detect. The third stage, the “tertiary lie,” emerges around the seventh or eighth year of age. At this stage, children gradually become increasingly sophisticated in controlling semantic leaks. They know how to say a deliberate lie and at the same time make sure that their other statements do not contradict the initial false statement (Talwar & Lee, 2008a, 2008b).

Empirical Basis of Psychophysiological Detection of Deception in Children and Adolescents

The first research focused on the detection of deception in the child population was conducted in a group of juvenile delinquents (Lyon, 1936; Rourke & Kubis, 1948). Based on the available information, it was initially an overall detection of deception in children under eighteen, but without further specification regarding the precise age and developmental level of the child. The aim was to help clarify criminal activities in adolescents. These initial studies were conducted in the form of simple physiological experiments, where only some physiological channels were used alone or in combination to detect lies. The researchers worked with basic physiological levels, the reliability of which was confirmed in studies focused on the detection of deception with an adult sample.

The initial recorded study, which focused on the detection of deception in the under-eighteen age group, for its purpose used continuous relative blood pressure and respiration recording at the same time. In 1935, at the Institute for Juvenile Research in Chicago, Verne Lyon examined suspected deception in 100 cases, which were recommended by the Juvenile Courts of Chicago (Lyon, 1936). The average age of selected delinquents was 15 years. The purpose was to facilitate the court’s decision through an examination of physiological reactions to critical questions. The obtained results suggested that the examined persons may be separated into three groups:

1. Clean physiological records indicating presumptive innocence of the subject (20%)

2. Disturbed physiological records followed by confession (33%)

3. Disturbed physiological records without admissions of guilt (47%).

The first two groups performed slightly better than the third group, with more than 50% of the conclusions supporting the court's decision. The findings confirmed that deception can be detected in children under the age of eighteen.

Another of the first studies, which dealt with some selected aspects of psychophysiological problems in children, was the study performed in the 1940s in New York by Rourke and Kubis (1948) on 80 delinquent boys. According to the records, the aim of their study was to 1) determine the conditions under which the psychogalvanic response (PGR, now known as the electrodermal response or EDR) is a reliable tool for detecting deception, and 2) whether delinquent boys are less subject to detection of lying than their non-delinquent peers. Comparative analysis using one physiological record, two records, and more than two records found that the reliability score was low when the stimulus was presented once, significantly above the average of chance when the stimulus was presented twice and was 97% accurate when the stimulus was presented more than twice. Regarding the second goal, the reliability score was not significantly higher in non-delinquent boys. Therefore, the study shows that when obtaining a sufficient number of records, namely more than two records by an experienced specialist, PGR is an exceptionally reliable indicator of deception. At the same time, the difference in detecting lies in delinquents and non-delinquent peers was not confirmed (Rourke & Kubis, 1948). Therefore, we can assume that the results may be even more accurate if we use all physiological channels of polygraph instrument during the detection of deception.

From June 1964 to December 1970, the Macomb County Sheriff's Department in Michigan used a polygraph within their department and some of the investigated persons were adolescents under the age of eighteen, which underwent the examination due to several reasons. Most of them participated in sexual assaults. The group of adolescents consisted of 29 children from 11 to 17 years. Twenty-six of them were women, most of whom had a confirmatory test, and three men, who appeared to be mostly suspects. Based on the results, the presumed statement was confirmed in 15 of them, 13 of them were untrue and in one person, due to mental disability, an ambiguous result was recorded. All the adult men assessed, who were under suspicion and assessed, showed the same results as the women in the applicant's position, so the examination was supplemented by a confirmatory test. Although exact statistical percentages are not given, the use of the polygraph in children in this department appears to have been successful (as cited in Adang, 1995).

More specific findings regarding the age and reliability of recording of deception came in 1969 with Voronin, Konovalov and Serikov (as cited in Adang, 1995), who tested children aged six up to adults to determine whether a memorized object (numerical card) could be detected. By recording skin resistance as the only physiological channel, the researchers could not correctly identify a single object in 6- to 7-year-old children, and only 12% in 8- to 12-year-old children were correctly identified. In the 14-to 16-year-old group, 53% were detected, and in the 18- to 30-year-old group, the accuracy was up to 87%. However, the final score in the groups with younger children was significantly lower than the score of the older population (Craig, 1997). Based on these results, Voronin with his colleagues noted that in children older than 14, the reliability of the detection of deception increases above the level of coincidence. Under this age limit, the reliability is low due to several reasons.

In 1970, based on a similar principle, Lieblich (1971) gave information detection tasks to 3- to 4-year-old Israeli children. The task was similar to the Guilty Knowledge Test (GKT), which is used in the standard polygraph practice. In this case, skin resistance was also the only physiological criterion recorded during the test. Basically, it was an operational replication of the study performed in an adult sample two years earlier (Lieblich, 1969). Regarding the differences between children and adults from the point of view of development, Lieblich's study aimed to find out whether the structure of recorded physiological reactions in adults also applies to the younger population and whether the relationships between individual physiological functions change with age. Twenty-six children, aged 3 and 4, participated in the Lieblich research. He presented six different stimulus sequences to each child. The stimuli were the names known to the subject, the names less known to the subject, and the unknown names, which the investigator obtained from interviews with the family members of the examined child prior to the testing itself. The order of the presented stimuli in the individual sequences rotated. Lieblich noted that the final values, using similar physiological criteria as in the adult population, were similar to the values obtained based on chance (1:6). The study also shown that the physiological index in children carries a lower amount of information than in adults. However, it was not possible to clearly determine whether this difference results from accidental disruptive influences, lack of attention, or low differentiation in the presented stimuli.

The conclusion of Lieblich's study, that the psychophysiological mechanisms necessary to detect lies, seem to work in children and can be used in the same way as in the adult sample.

One of the first research studies consistent with the standard practice of the polygraph in the investigation process and which included the children's population was performed by Abrams (Abrams, 1975). His research aimed to determine the accuracy of the polygraph in children. There were eight children each selected by their teachers from grades four through eight, inclusive. Their ages were from 9 to 13. However, only 37 children in total participated in the research, while in the 11-year-old group only six children were examined, and in the 12-year-old group, there were only seven children. The division into experimental and control groups was random. During the research, each child received a package of Life Savers sweets with a wild cherry flavor wrapped in a red package. For his research purposes, Abrams used a test known in the polygraph community as the Peak of Tension - Known Solution Test. He performed three tests, each repeated twice, and each consisted of five items. In the first test, the examiner asked whether the child received a pencil, gum, Life Savers, money, or licorice stick. In the second test, the examiner asked that if they received Life Savers, whether it was packed in a green, blue, red, yellow, or white package. In the third test, the examiner asked that if they received Life Savers, whether it was: orange, peppermint, cherry, lemon, or grape flavored. Each test was evaluated by a specialist and another independent, blind evaluator. The overall accuracy of decisions was 77%. The average accuracy for individual categories was 69% for 9-year-old children, for 10-year-old children it was 57%, for 11-year-old children it was 83%, for 12-year-old children it was 96%, and the accuracy for 13-year-old children was 94%.

Based on his study, Abrams concluded that the accuracy of detection of deception in children starts by the 6th grade - or in 11-year-old children with average intelligence. Abrams proposed is necessary to be cautious under the age of 11.

In another study, Abrams and Weinstein (1974) included 16 retardates divided into four different IQ levels (IQ < 50, IQ: 50-64, IQ: 65-69, IQ: 70-75). According to the findings, for those with IQ lower than 69, the accuracy dropped to a point below chance. In the IQ range between 65 and 75 the examiners were able to determine accurately which subjects gave a truthful response, but lying could be detected only at a chance level. An adult with an IQ of 75 would have a mental age of approximately 11 years. Based on these results it might be assumed that the polygraph could not be employed validly with a child at 11 years and less.

In 1996, Bradley et al. (as cited in Craig, 1997) using GKT assessed the second grade and sixth grade children. Similarly, to the previous studies, this research also used skin resistance for the detection of guilt or innocence of the child. The accuracy varied from 82% to 88% and without any influence on the age. However, Bradley noted that gender influenced the correct detection of information when girls were detected more often than boys. His study took into consideration also the developmental differences from the point of view of cognitive and other differences of children from the adults.

Ronald Craig (1997), who works as a professor of psychology at the Edinboro University in Pennsylvania, as part of his dissertation, performed research to determine whether a polygraph could be used with some success for the detection of deception in the adolescent population. It was an experiment that was performed in the form of a mock crime following the current standards of psychophysiological practice. The research involved 84 juveniles, aged 9 to 15 years, with 12 children - 6 boys and 6 girls - in each age category. Each child who took part in the research received $10 in cash and a $3 cinema ticket as compensation. The task of the participants of the experiment was to complete the workbook, specially created for this purpose, before the testing itself. Children were randomly assigned to one of the groups so that the same number of children of each age group and gender were between truthful and deceptive. The task of the children in the guilty group was to tear the last page out of the workbook, hide it in their pocket and deny to the specialist that they tore the page out of the workbook. The innocent group was given a workbook with a pre-torn page. They were told that someone had torn the page in advance and their job was to convince the specialist of their innocence. A $3 ticket to the cinema was to serve as their motivation in case they managed to convince the specialist that they had not ripped the page out of the workbook. For the detection of deception a variation of the Comparison Question Test (CQT) with Directed Lies (DLT) was used. The test format included 11 questions, three of which concerned the torn-out page. Based on the results, it was possible to correctly identify 72.6% of cases as truthful or untruthful. However, better results were recorded in the innocent group, where up to 37 children (88.1%) out of a total of 42 truthful participants were correctly detected. Decision accuracy for the guilty group was significantly lower. Of the 42 children, 24 children (57.1%) were detected.

Current State of Psychophysiological Detection of Deception in Children and Adolescents

Testing of children is not prohibited in the field of detection of deception, but it is partially regulated, and the psychophysiological community worldwide follows the standards of APA (2021), which sets the conditions for testing and the suitability of the examinee for an examination. A uniform and binding minimum age for conducting examinations in the children's population, on which most specialists would agree, has not been established in the field of psychophysiological detection of deception. Based on the data gained from the survey performed by Adang (1994), the lowest age limit accepted for the use of the polygraph among specialists ranged from 6 to 16 years, with a significant difference in the number of tested children, from 4 to 300 tested adolescents by individual specialists. Already at that time, however, the initial emphasis in assessing suitability, even in lower age categories, was laid on the overall maturity, level of development, and intelligence of the child (Adang, 1994). According to data from Craig and Molder's (2003) research, as many as 74.3% of respondents stated that they had tested at least one adolescent under the age of 16, while the minimum recorded age of the person tested was 7 years. According to the respondents, the average minimum age suitable for performing a psychophysiological examination was at the level of 12.84 years, with standard deviation of 1.79. Emerick and Dutton (1993) suggest a specific age criterion of 13 years for the use of the polygraph in a group of "high-risk" sexual offenders. Currently, most programs that apply Post Conviction Sex Offender Testing (PCSOT) to juvenile sex offenders set an age limit for testing of 14 years or older, as well as a minimum functional maturity of 12 years. A child under the age of 14 can be assessed in cases where the multidisciplinary team decides that the results will be clinically useful and at the same time the minimum functional maturity limit is observed (Colorado Sex Offender Management Board, 2020; Oregon Youth Authority, 2016). Until recently, there had been a general consensus based on both research and practical application that mental abilities at the minimum level of an 11- to 12-year-old child are sufficient to perform a valid psychophysiological examination.

However, the updated standards of the APA (2021) makes no mention of age in terms of suitability for polygraph examination. "Possess a basic understanding of right from wrong, and the difference between truth and lies, as demonstrated by an ability to verbalize potential reasons for being honest or dishonest, and the potential rewards or consequences for dishonesty or truthfulness" is stated as one of the minimal requirements of examinee suitability related to developmental issues (APA, 2021, p. 106). Because there is no published research suggesting that any of developmental issues will result in erroneous examination results, APA (2021) including ethical, professional, and empirical practices suggests that the application of normative data and normative interpretation rules to persons whose functional characteristics are outside the normal range should be regarded with caution.

The process of assessing the suitability of the examinee for the examination often involves the individual circumstances of the case and often the moral or ethical stance of the specialist whether or not to perform an examination. In Craig and Molder's (2003) research based on feedback, 13.6% of respondents stated moral developmental limitations as the greatest limitation in testing children and 1.4% stated an ethical or personal prohibition. After all, according to American Association of Police Polygraphers (AAPP, 2001), the specialist is the ultimate authority in deciding the reliability of a given polygraph test in the adolescent population. The ASTM Standard Guide for PDD examinations (ASTM, 2000) similarly states that the specialists themselves should ensure that the examined person is competent to undergo the test to the extent applicable to the law.

Factors Interfering with the Psychophysiological Detection of Deception in Children and Adolescents

Since the origin of psychophysiological detection of deception as a scientifically based method, little scientific research has been conducted on the age group under 18 years (Abrams, 1975, Craig, 1997; Lieblich, 1971; Rourke & Kubis, 1948). There are several reasons for this. One of the crucial factors that interferes with the possibility of using psychophysiological detection of deception in children in the research area is the moral aspect and related ethical matters that question the detection of deception, its validity, and the need for using it in the child population (Chaffin, 2011).

Naturally, each profession is based on ethical principles. The ethical code is one of the first documents of each functional structure and organization. However, special, and increased attention is paid to the ethical side and moral norms when it comes to the interests of the vulnerable and the weak, thus most especially the youngest - children. Chaffin (2011) specifies the vulnerability of this group not only because of their dependence, but also because of their greater susceptibility to abuse, intimidation, pressure, and intrusion into their basic human dignity. Children or adolescents come to the forefront of psychophysiological detection of deception especially when they take part in criminal investigation processes and when they function as a victim, witness, or offender.[1] The right to a fair trial is a fundamental pillar of a democratic society. Children suspected or accused of a crime also have the right to a fair trial and have the same guarantees as any other person who is accused of breaking the law. The guarantees of a fair trial apply from the first interrogation of the child and last through the trial witnesses (European Union Agency for Fundamental Rights, 2015). The task of detection of deception in children as well as in adults is to verify their truthfulness concerning the given investigated criminal offense. In these cases, detection of deception becomes an auxiliary tool in the whole criminal justice process, and therefore it is necessary to respect all the rights of the child in the context of the justice system. This wording primarily refers to the rights of children and adolescents accused, prosecuted, or imprisoned for crimes, as well as children who participate in criminal or non-criminal proceedings as victims and/or witnesses (European Union Agency for Fundamental Rights, 2015). EU law deals with the protection of child victims and witnesses. The Directive on the Rights of Victims 2012/29/EU specifically states that the main focus is on their best interest, which is assessed individually. In addition to this, a child-centered approach must prevail, considering the child´s age, maturity, views, needs, and concerns (European Union

[1] *For the purposes of psychophysiological detection of deception, a child is a human being under the age of 18, as regulated by the Criminal Code of the Slovak Republic, Article 127(1) of Act no. 300/2005 Coll. and at the same time from an international point of view, as set out in Article 1 of the UN Convention on the Rights of the Child (CRC). Pursuant to Article 94(1) of Act no. 300/2005 Coll., an adolescent is considered a person who at the time of the commission of the crime has reached the age of 14 and has not exceeded the age of eighteen.*

Agency for Fundamental Rights, 2015). As the child represents a particularly vulnerable category, the individual criminal and judicial procedures are subject to special laws and are characterized by specific particularities which are conditioned by children's individual characteristics and the state of their mind (Briatkova, 2020). Article 135 of the Criminal Code No. 301/2005 Coll. regulates the position of children or adolescents in legal proceedings and special conditions of interrogation in the territory of the Slovak Republic.

Regarding these facts, the problem referred to by several institutions is the possible effectiveness of the use of the results of psychophysiological detection of deception so that the interests of the child are not harmed (Chaffin, 2011). The United Nations Convention on the Rights of the Child (CRC), Part I, Article 3 (2007) states that "the best interests of the child shall be a primary consideration in all actions concerning children, whether undertaken by public or private social welfare institutions, courts of law, administrative authorities or legislative bodies". From an ethical point of view, the Convention on the Rights of the Child (Part I, Art. 40) also sets out the basic principles for the treatment of children who violate the law.

Another useful aspect to consider when using psychophysiological detection of deception in children and adolescents is the minimum age of criminal liability. However, reports submitted by States which are Parties of the Convention (CRC, 2007) indicate a wide range of minimum ages of criminal responsibility. They range from a very young age of 7 or 8 years to a later age of 14 or 16 years. Some States which are Parties of the Convention, use two minimum ages of criminal responsibility. Children who break the law, who are at or above the lowest minimum age but below the higher minimum age, are considered criminally liable only if they have the required maturity in this regard. An assessment of this maturity should be provided by a forensic clinical psychologist. However, the Convention states that this often happens without the involvement of a psychological expert, and in practice, this may result in considering a low minimum age in cases of serious crimes and the subsequent application of discriminatory practices. In Slovakia, the criminal liability of adolescents is regulated by Article 22 of Act 300/2005 Coll. in the following words: "Who did not reach the age of 14 at the time of committing the crime is not criminally liable." The age limit for sexual offenses is 15 years.

A frequently discussed problem in actual investigative and research practice in children and, at the same time, the most common argument against the use of the polygraph in children are the cognitive abilities of the child. Adequate cognitive abilities of the child are mentioned in most experimental and non-experimental studies in order to deduce the lowest age of the examined person so that sufficient validity of the method is proven (e.g., Adang, 1994; Craig & Molder, 2003). Based on the anonymous survey made by Craig and Molder (2003), in which 101 polygraph examiners participated, insufficient cognitive abilities have become the most common concern as to why psychophysiological examinations may be ineffective and unreliable in the child population.

The basic condition for the detection of deception is for the examinee to know and understand the difference between reality and fantasy, facts and fiction, lies and truth. As cognitive abilities in children develop and deepen, so does the perception, recognition, and use of lies in everyday situations. The process of deception implies the use of complicated cognitive functions (Talwar, 2018). According to the available literature, the level of cognitive maturity required for the evaluation of a lie in the perception of an adult reaches its maturity around the age of ten (Piaget, 1948; Talwar, 2018).

Given all the factors that can influence psychophysiological examinations, it is appropriate to ask the questions of whether, to what extent, at what age, and in what cases is it, therefore, necessary to involve the child in the investigation process of detection of deception.

According to Adang (1995), the most common reasons for a child to undergo a psychophysiological examination is the increasing number of sexual offenses and the associated fact that the embarrassment of reporting these crimes are not covered up as they have been in the past. This argument partially stems from the fact that special attention is paid to sexual offenses abroad and many, not only American states but also European countries, have special post-conviction sexual offenders' treatment programs, not only for adults but also for adolescents (Colorado Sex Offender Management Board, 2020; Emerick & Dutton, 1993; Oregon Youth Authority, 2016). To further highlight the detection of deception in children, Adang (1995) mentions a large number of false allegations made by victims, which include false allegations made by children. The third and most important reason is simply the importance of revealing the truth, regardless of age.

Special Areas of Psychophysiological Detection of Deception in Children and Adolescents

A special category of detection of deception abroad is represented by adolescent sex offenders, where detection of deception is used as part of treatment, prevention, and supervision programs. These programs were adopted for a group of adolescents and derived from programs for adult sexual offenders, known as "Post-Conviction Sexual Offender Testing". The task of PCSOT is to confirm or ascertain the history of sexual assaults, to assist in resolving controversial or unclear aspects of the case, and to monitor an offender´s compliance with probation and treatment contracts (Emerick & Dutton, 1993). Proponents of PCSOT argue that polygraphy assists in overcoming denial, results in better background information about offenders, provides important risk assessment information and helps to monitor adherence to supervision conditions (Grubin, Kamenskov, Dwyer & Stephenson, 2019). The offender undergoes a psychophysiological examination only after a verdict has been pronounced, as part of the therapeutic process. Studies that monitored the behavior of sex offenders showed results in the form of an increased number of confessions and admitted incidents regarding the sexual history among offenders who underwent polygraph (Hidman & Peters, 2001). In England, as part of the PCSOT research, a trial test of sexual offenders took place in 2003-2005, which was assessed as helpful in 90% of cases (Grubin, n.d.). Mandatory testing of high-risk sex offenders released on parole was launched in 2014. Some Police Forces tested sex offenders recorded in a register on a voluntary basis (Grubin, Kamenskov, Dwyer & Stephenson, 2019). However, the use of detection of deception in a group of adolescent sex offenders has been widely criticized, even though in the USA in 1992-2002, according to surveys, up to 70% of PCSOT programs used polygraphs and pointed to it as an effective prevention tool (McGrath, Cumming, Hoke & Bonn-Miller, 2007). It is currently estimated that nearly half of the programs focused on the treatment of adolescent sex offenders across the USA use the polygraph in the therapeutic and treatment process (Jensen, Shafer, Roby & Roby, 2014). Strong opposition to the use of polygraphs in programs for adolescent sexual offenders represents the view of members of the Association for the Treatment of Sexual Abusers (ATSA). Its 2017 standards radically opposed the use of the polygraph in the treatment process (Lee, Lemaster, Hanlin & Johnson, n.d.). Their concerns are, primarily, the lack of research and related questions of validity and reliability that would support the use of the polygraph in adolescents (Chaffin, 2011; Rosky, 2015). Ethical aspects are another factor on which opponents base their critical arguments (Grubin, Kamenskov, Dwyer & Stephenson, 2019). To prevent any manipulation, non-professional performance, or psychological harm to the adolescent, a psychophysiological examination of adolescent sex offenders who are part of the treatment program shall follow the labor standards set by the country and adapted to the adolescent group. The specialist-psychophysiologist is part of a multidisciplinary team and is subject to relevant professional and ethical standards (Colorado Department of Public Safety, 2020; Oregon Youth Authority, 2016). Taking the criticism into account, the existing data conclude that as long as the principles of psychophysiological testing are respected and adhered to, together with a sensitive approach to the offender, the polygraph acts as a supportive tool in treatment and surveillance not only in adult sex offenders (Grubin & Madsen, 2006) but also in adolescent delinquents (Grubin, Kamenskov, Dwyer & Stephenson, 2019; Hidman & Peters, 2001; Jensen, Shafer, Roby & Roby, 2014).

Psychophysiological Examination of Children and Adolescents in the Slovak Republic

In the Slovak Republic, testing children under 18 is not a standard routine. According to the internal records, since the foundation of the Department of Applied Psychophysiology in the Institute of Forensic Science of the Police Force in 1998, which comes under the competence of the Ministry of Interior of the Slovak Republic, there were nine children under 18 years of age, three girls and six boys, tested by polygraph. The minimum recorded age was 14 years. More specifically, three tested adolescents were 14 years old, one was 15 years old, one was 16 years old, and four were 17 years old. The adolescents were witnesses or suspects of a specific crime. Two were the aggrieved persons in the examined cases. These were mainly violent crimes, crimes of a sexual nature, the crime of extortion, and restrictions of personal liberty. There were no obstacles on the part of the specialists that would prevent the psychophysiological examination of minors: cognitive abilities were sufficient to perform the test and appropriate to the age level. Based on the standard working procedure, a clear conclusion of the psychophysiological examination was determined in seven cases (truthful or deceptive) and in two cases the testing was terminated with an inconclusive result. In the investigated cases, 14 tests were performed in total (some adolescents were tested more than once), of which seven tests were concluded with deception indicated, the results of four tests were non-deception indicated, and in three cases the test result was inconclusive. Three of the five adolescents who were deceptive in the post-test part of the examination admitted certain information concerning the investigated act which they had concealed in the pre-test of the examination. One person did not provide any additional information and insisted on their original statement, and one person was not informed about the results due to operational and technical reasons.

Conclusion

Although children, or more precisely adolescents represent a minimal percentage of the total number of examined persons, psychophysiological detection of deception in this area also represents an opportunity to help direct the investigation and, in an ideal state, to contribute to increasing the clearing rate of underage delinquency. However, it is still true that from the point of view of the judicial system in the Slovak Republic, the conclusions of psychophysiological examinations cannot be used as direct evidence in court, but it is also true that they can be considered as forensic-technical means that can help to direct investigations in searching, finding, and taking of evidence, or more precisely they can help to confirm or disprove the facts found in the investigation process (Criminal Code of the Slovak Republic, Article 119(3) of Act no. 301/2005 Coll).

From initial attempts to detect lies using physiological functions, through the promotion of psychophysiological detection of deception to the scientific level, its application in judicial, criminal, police practice, and other social, security, and political areas of life, to the current state, the examination by a "lie detector" has always been a controversial topic and it is still so today. We do not anticipate significant changes in this established trend. Humans project internal fears, suppressed fear, secret desires, personal, and mediated experience into their beliefs. At a more intellectual level, they defend themselves with insufficient reliability, inconclusive data, insufficient research. From the point of view of human nature, we cannot avoid some of them, but we can change some of them with a sober view of the current scientific knowledge and we can contribute to some of them by our own initiative in terms of our scope of authority. Real use in practice is conditioned by scientific research and empirical data. The field of research, especially the one concerning the position of children and adolescents in the psychophysiological examination, is generally underdeveloped in the polygraph community. The fact that research in this area is in general ethically limited contributes to this; in addition, children and adolescents represent a special group that both formally and procedurally require a specific and sensitive approach, preparation, and cooperation with several institutions and experts. In order not to find ourselves in a vicious cycle, it is necessary to look for ways of adhering to professional psychophysiological standards, following our moral beliefs, be concerned at the ethical level, and by ensuring an adequate safe environment for the underage examinee. These considerations can help us contribute to the issue and broaden the boundaries of psychophysiological detection of deception in children.

References

Act 300/2005 Coll. Criminal Code of the Slovak Republic. [Electronic version] [cit. 2021-7-10]. Retrieved from http://www.zakonypreludi.sk/zz/2005-300.

Act 301/2005 Coll. Code of Criminal Procedure of the Slovak Republic. [Electronic version] [cit. 2021-7-10]. Retrieved from http://www.slov-lex/pravne-predpisy/SK/ZZ/2005- /301/.

Abrams, S. (1975). The validity of polygraph technique with children. *Journal of Police Science and Administration,* 3(3), 310-311. © 1975 by Northwestern University School of Law.

Abrams, S., and Weinstein, E. (1974). The validity of the polygraph with retardates. *Journal of Police Science and Administration*, 2(1), 11- 14.

Adang, S.R. (1995). The use of the polygraph with children. *Polygraph*, 24(4), 259-274.

American Association of Police Polygraphists (2001). *AAPP Standards and Principles*. [Electronic version] Retrieved from http://www.americanassociationofpolicepolygraphists.org.

American Polygraph Association. (2021). Model policy for the evaluation of examinee suitability for polygraph testing. [Electronic version] Retrieved from http://www.polygraph.org.

American Society for Testing and Materials. (2000). [Electronic version] Retrieved from http://www.astm.org. doi: 10.1520/E2063-12R17.

Anderson, M.C., and Weaver, C. (2009). Inhibitory control over action and memory. *Encyclopedia of Neurosience.*153-163. doi: 10.1016/B978-008045046-9.00421-6.

Briatkova, M. (2020). Interrogation of a child in criminal proceedings [Výsluch die'a'a v trestnom konaní]. *Projustice.* Scientific Journal for Law and Security Science [Vedecko codborný časopis pre právo a bezpečnostné vedy]. Retrieved from http://www.projustice.sk.

Burton, R. V. (1963). Generality of honesty reconsidered. *Psychological Review*, 70(6), 481-499. doi: 10.1037/h0047594.

Bussey, K. (1992). Lying and truthfulness: Children´s definitions, standards and evaluative reactions. *Child Development*, 63(2), 129-137. Retrieved from http://www.jstor.org. doi: 10.2307/1130907.

Carlson, S.M., Moses, L.J., and Hix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, 69(3), 672-691. doi: 10.1111/j.1467-8624.1998.tb06236.x.

Chaffin, M. (2011). The case of juvenile polygraphy as a clinical ethics dilemma. *Sexual Abuse: A Journal of Research and Treatment*, 23(3), 314–328. Retrieved from http://www.sagepub.com/journalsPermissions.nav. doi: 10.1177/1079063210382046.

Colorado Sex Offender Management Board (2020). Standards and guidelines for the evaluation, assessment, treatment and supervision of juveniles who have committed sexual offenses. [Electronic version] Colorado Department of Public Safety, August 2020. Retrieved from http://www.colorado.gov/dc.

Craig, R.A. (1997). The Use of Physiological Measures to Detect Deception in Juveniles. Draft of the dissertation work. Edinboro University of Pennsylvania, p. 65. Available from the author.

Craig, R.A., and Molder, C. (2003). The use of law enforcement polygraph tests with juveniles. *The Journal of Credibility Assessment and Witness Psychology*, 4(1), 63-74.

DePaulo, B.M., and Jordan, A. (1982). Age changes in deceiving and detecting deceit. *Development of Nonverbal Behavior in Children. Feldman*, R.S. (ed.) New York : Springer Science + Business Media. 151-180. doi: 10.1007/978-1-4757-1761-7_6.

Ekman, P. (1991). *Why kids lie. How parents can encourage truthfulness*. Middlesex, England : A Penguin Books. pp. 212. ISBN: 978-0-14-014322-5.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, marriage*. W.W. Norton & Compnany. p. 416. ISBN: 0393337456.

Emerick, R.L., and Dutton, W.A. (1993). The effect of polygraphy on the self-report of adolescent sex offenders: Implications for risk assessment. *Annals of Sex Research*. 6(2), 83-103. doi: 10.1007/BF00849301.

European Union Agency for Fundamental Rights (2015). *Handbook on European law relating to the rights of the child* [Príručka o európskom práve týkajúcom sa práv die'a'a]. [Electronic version] Belgium, 2015. pp. 256. Retrieved from http://www.echr.coe.int.

Evans, A.D., and Lee, K. (2013). *Emergence of lying in very young children. Developmental Psychology*, 49(10), 1958-1963. ©2013 American Psychological Association. doi: 10.1037/a0031409.

Fu, G., Evans, A.D., Xu, F., and Lee, K. (2012). Young children can tell strategic lies after committing a transgression. *Journal of experimental child psychology*, 113(1), 147-158. doi: 10.1016/j.jecp.2012.04.003.

Grubin, D. [n.d.]. *Polygraph: The use of polygraphy in the assessment and treatment of sex offenders*. pp. 4. Retrieved from https://www.nota.co.uk/wp-content/uploads/2019/01- /NOTA-polygraph-paper-ready-for-publication.pdf.

Grubin, D., Kamenskov, M., Dwyer, G.R., and Stephenson, T. (2019). Post-conviction polygraph testing of sex offender. *International Review of Psychiatry*. Published online 27. March 2019, 37(2), 1-8. doi: 10.1080/09540261.2018.1561428.

Grubin, D., and Madsen, L. (2006). Accuracy and utility of post-conviction polygraph testing of sex offenders. *British Journal of Psychiatry*, 188, 479-483. doi: 10.1192/bjp.bp.105.008953.

Hidman, J., and Peters, J.M. (2001). Polygraph testing leads to better understanding adult and juvenile sex offenders. *Federal Probation*, 65(3), 8-15. Retrieved from http://www.ccoso.org.

Jensen, T., Shafer, K., Roby, C.Y., and Roby, J.L. (2014). Sexual history disclosure polygraph outcomes: Do juveniles and adult sex offender differ? *Journal of Interpersonal Violence*, 30(6), 928-944. Retrieved from http://www.sagepub.com/journalsPermissions.nav. doi: 10.1177/08862605 1 4539766.

Lee, T.T.C., Lemaster, N.M., Hanlin, H.R., and Johnson, M.A. [n.d]. Use of the polygraph with adolescents who have engaged in sexually abusive behavior: A Critical review of the empirical and ethical issues. 1-26. Retrieved from http://www.in-ajsop.org/wp-content/uploads/2018/09/Use_of_the_Polygraph_with_Adolescents_2018.pdf.

Lieblich, I. (1969). Manipulation of contrast between differential GSR responses through the use of ordered tasks of information detection. *Psychophysiology*, 6(1), 70-77. doi: 10.1111/j.1469-8986.1969.tb02885.x.

Lieblich, I. (1971). Manipulation of contrast between differential GSRs in very young children. *Psychophysiology*, 7(3), 436-441.

Lyon, V. (1936). Deception test with juvenile delinquents. *Psychological Bulletin*, 32, 685 – 686. doi: 10.1080/08856559.1936.10533746.

McGrath, R.J., Cumming, G.F., Hoke, S.E, and Bonn-Miller, M.O. (2007). Outcomes in a community sex offender treatment program: A comparison between polygraphed and matched non-polygraphed offenders. *Sex Abuse*. Published online: 4. October 2007, 19(4), 381-393. doi: 10.1007/s11194-007-9058-z.

Nelson, R., and Handler, M. (2013). A brief history of scientific reviews of polygraph accuracy research. [Electronic version] *APA Magazine*, 46(6), 22-28. Retrieved from http://researchgate.net/publication/275956904.

Neville, J. (2018). Identification and child rearing. By Robert R. Sears, Lucy Rau and Richard Alpert. *The British Journal of Psychiatry*, 112(489). pp. 857. doi: 10.1192/bjp.112.489.857.

Newton, P., Reddy, V., and Bull, R. (2000). Children´s everyday deception and performance on false-belief task. *British Journal of Developmental Psychology*, 18(2), 297-317. doi: 10.1348/026151000165706.

Oregon Youth Authority. (2016). Polygraph use with youth who have sexually offended. *Policy Statement*. Part II. [Electronic verison] Last review 2016. pp. 5.

Peterson, C.C., Peterson J.L., and Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development*. 1983, 54(6), 1529-1535. doi: 10.2307/1129816.

Piaget, J. (1948). *The moral judgement of the child* [electronic version]. Glencoe, Illinois: The Free Press. pp. 448. ISBN: 31148004507638.

Raver, C.C., Blair, C. (2016). Neuroscientific insights: Attention, working memory, and inhibitory control. *Future of Children*, 26(2), 95-118. Retrieved from http://www.futureofchildren.org. doi: 101353/foc.2016.0014.

Robert, C., Borella. E., Fagot, D., Lecerf, T., and de Ribaupierre, A. (2009). Working memory and inhibitory control across the life span: Intrusion errors in the Reading Span Test. *Memory and Cognition*, 37 (3), 336-345. doi: 10.3758/MC.37.3.336.

Rosky, J.W. (2015). More polygraph futility: A comment on Jensen, Shafer, Roby and Roby (2015). *Journal of Interpersonal Violence*, 31(10), 1956-1970. doi: 10.1177/0886260515570752.

Rourke, F.L.; and Kubis, J.F. (1948). Studies in the detection of deception: I. Determination of guilt or innocence from psychogalvanic (PGR) records of delinquents and non-delinquents. *American Psychologist*, 3, 255-256.

Siegal, M., and Peterson, C.C. (1996). Breaking the mold: A fresh look at children´s understanding of questions about lies and mistakes. *Developmental Psychology*, 32(2), 322-334. doi: 10.1037/0012-1649.32.2.322.

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Study*, 9(1), 173-188. doi: 10.1111/j.2044-835X.1991.tb00869.x.

Stern, W. (2018). *Psychology of early childhood: Up to the sixth year of age*. [Hardcover ed.] [Electronic version] London: Routledge Taylor and Francis Group. (Volume 16) (hbk). ISBN: 978-1-138-08835-1.

Stern, C., and Stern, W. (1999). Recollection, Testimony, and Lying in Early Childhood. (Trans. James T. Lamiell). American Psychological Association. p. 167. ISBN: 9781557985743.

Stouthamer-Loeber, M. (1986). Lying as a problem behavior in children: A review. *Clinical Psychology Review*, 6, 267-289. doi: 10.1016/0272-7359(86)90002-4.

Stouthamer-Loeber, M., and Loeber, R. (1986). Boys who lie. *Journal of Abnormal Child Psychology*, 14(4), 551-564. doi: 10.1007/BF01260523.

Strichartz, A.F., and Burton, R.V. (1990). Lies and truth: A study of the development of the concept. *Child Development*, 61, 211-220. doi: 10.1111.j.1467-8624.1990.tb02773.x.

Talwar, V. (2018, December). Development of lying and cognitive abilities. *Oxford Handbooks of Lying and Cognitive Abilities*. Oxford University Press. doi: 10.1093/oxfordhb/9780198736578.013.30.

Talwar, V., and Lee, K. (2008a). Social and cognitive correlates of children´s lying behavior. *Child Development*, 79(4), 866-881. doi: 10.1111/j.1467-8624.2008.01164.x.

Talwar, V., and Lee, K. (2008b). Little liars: Origins of verbal deception in children. 2008. In: Itakura, S., Fijita, K. (Eds.) : *Origins of the Social Mind. Evolutionary Psychology and Child Development.* Springer Science + Business Media. 157-178. doi: 10.1007/978-4-431-75179-3 8.

Talwar, V., and Crossman, A.M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review*, 32(4), 337-359. doi: 10.1016/j.dr.2012.06.004.

United Nations Convention on the Rights of Child [Dohovor OSN o právach die'a'a]. (2007). Všeobecný komentár č. 10 (2007) : Práva detí v súdnom konaní. [Electronic version] Výbor pre práva die'a'a, Ženeva, 25.apríl 2007. CRC/C/GC/10 [cit. 2021-7-15]. Retrieved from http://www.justice.gov.sk.

Wang, L., and Wang, Z. (2018). The modified temptation resistance task: A paradigm to elicit children´s strategic lie-telling. *Journal of Visualised Experiments*. Published online © 2018. doi: 10.3791/57189.

Wilson, A.E., Smith, M.D., and Ross, H.D. (2003). The nature and effects of young children´s lies. *Social Development*, 12(1), 21-45. doi: 10.1111/1467-9570.00220.

A Field Assessment of Manually Scoring Electrodermal Data in Self-Centering and Non-Centering Modes

Donald J. Krapohl[1]

Donnie W. Dutton[2]

Abstract

For more than 60 years polygraph manufacturers have offered a useful feature for automatically centering the display of electrodermal data. This feature permits polygraph examiners to attend to question presentation and data management without the distraction of needing to repeatedly recenter a wayward electrodermal tracing. Many, perhaps most, polygraph examiners take advantage of this feature at least occasionally. While it is known the self-centering feature alters the display of the electrodermal data, its effect on manual scores rendered by field examiners has not been previously reported. In this one-year field study we collected 760 cases and tabulated differences in manual scores assigned to the self-centering and non-centering modes of a popular polygraph system to determine the frequency of divergent electrodermal scores as well as the influence of the display modes on polygraph decisions. Differences in electrodermal scores were found at least once in 24.9% of cases between self-centering and non-centering modes, and in the test results of 6.3% of all cases. Among the instances where the choice of EDA mode influenced the final decision, virtually all of them affected whether the results were inconclusive or not-inconclusive, with only one case resulting in an opposite decision (Significant Responses versus No Significant Responses). Implications are discussed.

---

[1]Former APA President and Editor, and currently APA President Elect and employed by the firm Capital Center for Credibility Assessment.

[2]Former APA President and current Director. Mr. Dutton is Vice President of the Capital Center for Credibility Assessment.

The authors are grateful to Kristine Smith and Mark Pszenny for their thoughtful comments to an earlier draft of this paper, and to Terry Ball for helpful historical information.

This article is another in the Best Practices series. The views expressed do not necessarily represent those of the Capital Center for Credibility Assessment or the American Polygraph Association. Questions and comments can be directed to APAkrapohl@gmail.com.

Introduction

A choice of electrodermal filters has been included as options on polygraphs since at least the 1950s for both analog systems and the more recent computerized instruments. The two principal choices, or modes, have been forms of self-centering and non-centering. The self-centering mode, as the name implies, tends to keep the electrodermal tracing near the middle of the display while the other choice allows the tracing to move widely due to less filtering of the signal. The advantage of the self-centering mode is that it relieves the testing examiner of the need to readjust the electrodermal tracing during testing. In addition, certain filters could increase the overall signal-to-noise ratio in the EDA data. An advantage of the non-centering mode is that there may be less distortion of the original signal that comes from the surface of the skin. In terms of how the data from these two filtering approaches are interpreted in manual scoring systems, none of the published scoring systems differentiate between the two modes.

In 2013 an article by the McClatchy News Service asserted that different electrodermal modes can "label truthful people as liars and the guilty as innocent" (Taylor, 2013). The article, which ran in several newspapers across the US, focused on filtering differences in an older model computer polygraph and claimed the differences in filters could influence test results. The potential for filters to alter test results, according to the article, was known to polygraph examiners for more than 10 years.

The suggested impact of electrodermal modes on polygraph results was conjectural as no evidence was offered in the McClatchy piece. US Government examiners who were interviewed for the McClatchy article supported the viewpoint, nonetheless, that filters could affect the final test results. The McClatchy article asserted the polygraph community had not adequately investigated whether electrodermal signal filtering could influence polygraph decisions. Indeed, we were unable to locate any publications suggesting government agencies, police departments or polygraph associations took the possibility seriously.

There is very limited information in the polygraph literature to address the issue. For historical context, Arther (1971, cited in Matte, 1996) reported he had simultaneously recorded electrodermal activity in both self-centering and non-centering mode in his field practice from 1958 to 1965. Arther posited the self-centering mode was better, though no useful quantitative analysis was offered. Conversely, Raskin, Barland and Podlesny (1978) concluded from their field and lab research that the self-centering mode should be eliminated because it can "…greatly distort the relative size and shape of (those) responses." Neither perspective has gained dominance in field practice in the subsequent years.

A more recent article to raise the matter appeared in this journal. Kalafati and Krapohl (2018) asserted that the way in which the self-centering mode filters the data (referred hereafter as the "automatic" mode[3]) can potentially alter the data in ways that can influence polygraph manual scores as compared to the same data in a non-centering mode (identified as the "manual" mode in this article), consistent with the view of Raskin et al. (1978). The Kalafati et al. paper included images of tracings in automatic and manual modes as examples of how this can occur. While it was clear these differences were possible, nothing in the article pointed to how often they might be expected, under what circumstances they tend to occur, or whether there could be a meaningful effect on polygraph results. Therefore, the question of whether the incidence of scoring differences was substantial or rare went unanswered. We could find no standards promulgated by any government agency, professional association, or standards-making body that address how electrodermal data should be displayed for polygraph examiners to score. Therefore, manufacturers can, and do, take

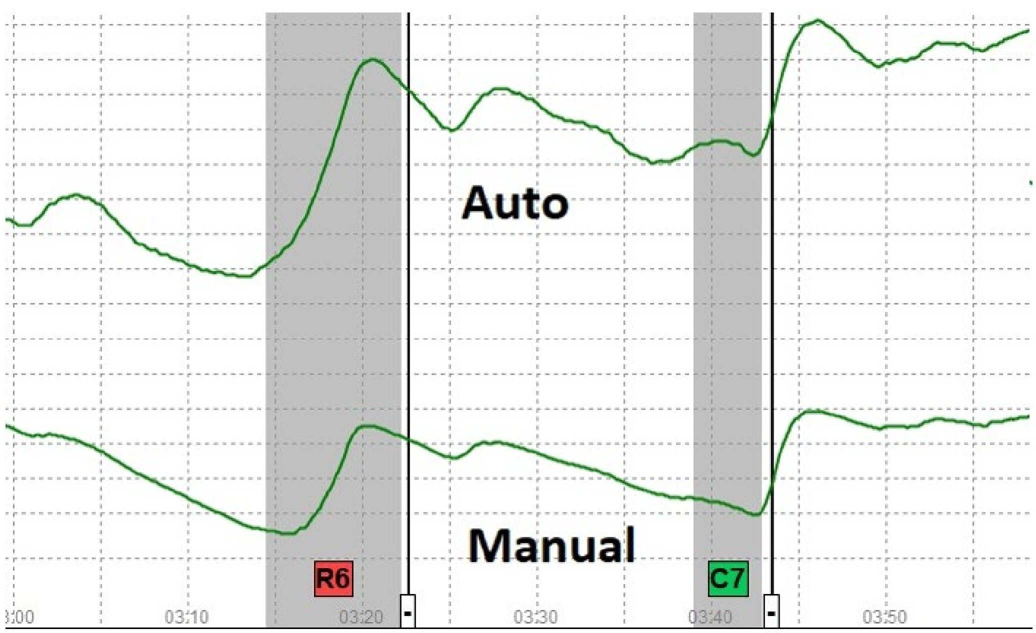
[3]The various polygraph manufacturers may use different terms for the self-centering mode. For example, the Limestone Technologies system refers to this mode as "Auto EDA." In the Stoelting product it is "Autocenter." The Axciton system calls it "Auto", and the Lafayette Instruments system, which was used for this study, calls this mode "Automatic."

quite divergent approaches to their filtering of electrodermal data.

In the same issue of the journal in which Kalafati and Krapohl (2018) published their review of EDA filtering, three of four North American polygraph manufacturers provided responses to address concerns raised in the paper. Among them, Nelson (2018) provided automated analyses of field data to demonstrate the strong statistical relationship between the Lafayette instrument's automatic and manual modes and ground truth. Representatives of a second manufacturer, Stoelting (Cochran & Fuentes, 2018), outlined an explanation of their EDA filter, reporting the two modes cannot produce opposite scores in their polygraph systems.[4] The third manufacturer, Limestone Technologies (2018), more candidly reported that filtering is widely known to make a difference in manual scores in some instances "among various polygraph instrumentation" and recommended examiners assess the data from both modes. None of the instrument manufacturers' responses included an evaluation of the effect of the different filters on manual scoring. Because manual scoring is the dominant method for evaluating polygraph data, this issue became our interest.

A related issue is whether filtering can influence a scorer's interpretation of the onset point of an electrodermal response. Consider Figure 1. The upper tracing shows electrodermal data displayed in the automatic mode and the lower in the manual mode. The manual tracing points to an onset latency of about 2 seconds after question onset, well within the normal response onset range of 1.2 to 8.0 seconds in polygraph testing (Krapohl, Halford, Benson, Mayston & Dutton, 2021). The same data shown in the automatic mode, however, has obscured the response onset point for question R6. In the automatic mode a scorer may regard the response onset as taking place too early and would not use the electrodermal response for scoring.

Figure 1. Display of EDA data in two modes to show how apparent onset latency may be changed according to the mode.



[4]We agree in principle that opposite manual scores for automatic and non-self-centering display modes may be unlikely for this polygraph system. Nonetheless in our limited archive we have one example of where it has occurred.

We were unable to locate any previous research on which to start our investigation. There are scores of polygraph validity studies in print, but none we could find that examined differences in manually scoring for two or more forms of electrodermal data filtering. The knowledge gap came as a surprise given the widely accepted finding the electrodermal channel accounts for about one-half of the diagnostic information available in polygraph data. If, as Kalafati (2018) supposed, manual scores could be influenced by which electrodermal filter the user chose, such information should be considered by field practitioners when they conduct manual scoring. Conversely, if this notion is in error and manual scoring is not affected by filtering modes, the information could be useful in challenges leveled against polygraph decisions in settings such as the courts, offender management programs, hiring systems and others where signal filtering may provide the basis for an objection. It would also directly confront the assertion in the McClatchy News article on EDA modes (Taylor, 2013).

We had at our disposal hundreds of field polygraph examinations conducted on the Lafayette Instruments polygraph system. Though ground truth was not known for any but a small subset, the cases afforded us an opportunity to determine, at least for this polygraph system, whether manual scores for the electrodermal channel were different between the two display modes.

Method

Data Source

Between January 1 and December 31, 2021, a large sex offender management polygraph program submitted 760 cases for independent quality control to the first author. This entire data set was used for the present research. All polygraph examinations had been conducted on Lafayette computerized polygraphs, series LX5000 or LX6, which recorded two channels of breathing, and one each of electrodermal, cardiograph, photoplethysmograph and movement data channels. The data were anonymized to remove personally identifying information. Demographic characteristics of the examinee population could not be captured, though it can be reasonably assumed nearly all examinees were adult males. All examiners were trained at the same education program for government and law enforcement students.

Polygraph Techniques

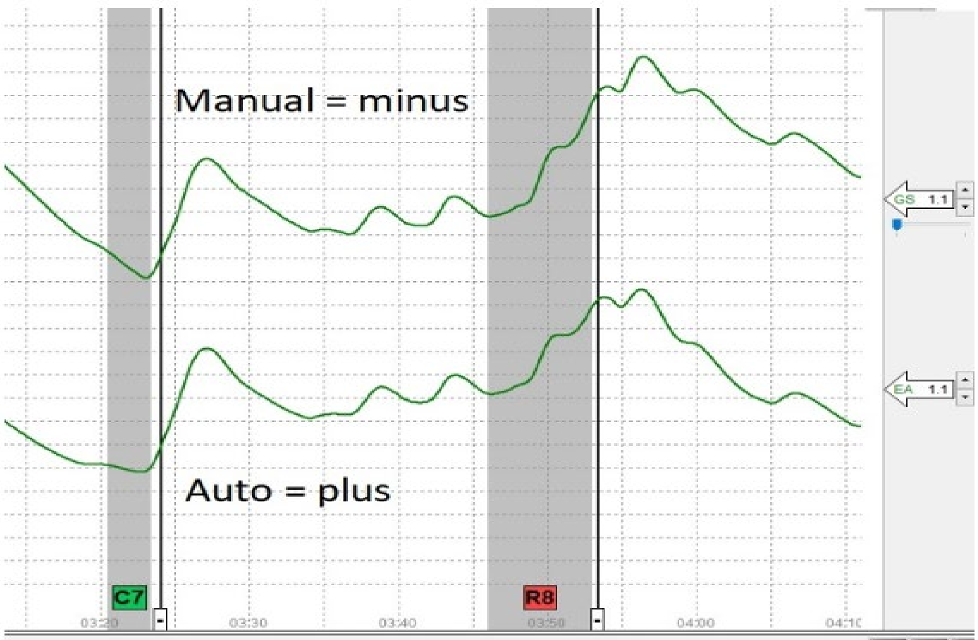
Examiners in this program all use one of three polygraph screening techniques: Air Force Modified General Question Technique (AFMGQT; Krapohl & Shaw, 2015) with either two or three relevant questions covering different test issues; the British One-issue Screening Test (BOST; Krapohl, Grubin, Benson & Morris, 2020) with two relevant questions covering a single issue, and; the Directed Lie Screening Test (DLST; Handler, Nelson & Blalock, 2008) with two relevant questions covering different issues.

Scoring Rules

All the techniques used in this study require the scorer to assess the physiological response of each relevant question against the closest two comparison questions that preceded and followed it. If there was a difference in the response magnitudes between the two comparison questions, scorers used the stronger of the two responses to score against the relevant question. All scorers used the Empirical Scoring System (Blalock, Cushman & Nelson, 2009; Nelson, Krapohl & Handler, 2008) whereby EDA scores could be +2, -2 or 0.

The minimum difference for assigning an electrodermal score in this polygraph program is 10% rather than the more widely employed heuristic of requiring only a visually discernible difference between the response magnitudes. This requirement arose from two considerations. First, “visually discernible” is subjective and can lead to differences in scores based solely on who the scorers are. Krapohl (2020) reported experienced examiners who share the same training can disagree 20% of the time as to whether a response is sufficiently “visually discernible”. A reanalysis of data for a separate study (Krapohl & Cushman, 2006) similarly found pairs of examiners disagreed an average of 18.2% of the time on whether to assign non-zero scores to electrodermal responses. Judgments may be influenced by experience, gain settings, source of training, bias and other factors. The application of an objective criterion - here a minimum difference of 10% - ensured the findings could generalize to the field where the same threshold is implemented. We could conceive of no objective equivalent in the construct "visually discernible." Figure 2 shows an example of differences in response amplitudes that vary with the filtering mode.

Figure 2. Electrodermal response tracings in the manual and automatic modes to show how they may differ in ways that can affect score assignment.



The second consideration that led to the choice of a minimum ratio of difference was that scores assigned to the smallest differences were associated with a lower correlation between manual scores and ground truth (Krapohl, 2020; 2021; but also see Nelson, 2020). The 10% threshold used by this offender management program seeks to incrementally boost decision accuracy as well as to limit the influences of subjective interpretation.

Related to scoring criteria, this offender management program only scores electrodermal responses with onset latencies greater than 1.2 seconds from question onset. This is due to convincing research indicating human physiology is incapable of initiating electrodermal responses (EDRs) in less than 1.2 seconds even under optimal conditions (Sjouwerman & Lonsdorf, 2019).

Consistent with field practices elsewhere, there were exclusionary criteria for scoring electrodermal responses. They included EDRs induced by movements and deep breaths. Anomalous electrodermal data were also excluded, such as non-responsive tracings. Cases in which there were no useful electrodermal data were not considered.

Decision Rules

The polygraph programs that were the source of all cases in this study used the following decision rules. For decisions of Significant Responses (SR) with the AFMGQT or the DLST, a total of -3 or lower for any individual test question was required. For a decision of No Significant Responses (NSR) with these testing techniques, the total of scores for each relevant question must individually be greater than zero. The BOST uses two-stage rules. If the grand total of all scores is +2 or greater the results are NSR. If the grand total is -4 or lower the results are SR. If the results would be Inconclusive, the second stage is implemented. In the second stage, if the total score for any individual question is -6 or lower, the results are SR. All others are Inconclusive. Examiners also resolve Inconclusive results after three charts by running a fourth chart, or if necessary, a fifth chart. No more than five charts are permitted.

Procedure

The electrodermal channel for the 760 cases were scored twice by the quality control reviewers, once in the automatic mode and again in the manual mode. Scores from the other data channels were not altered so that any effect found on polygraph decisions would result solely from changes in EDA scores. When it was not obvious whether an EDR was 10% larger than another, the caliper feature in the Lafayette software was used. The gain setting was sufficient such that the larger of the two compared EDRs was at least 6 chart divisions in amplitude, and up to 12 chart divisions in cases where the amplitudes were especially close in size. The 10% calculation was based on the smaller of the two reactions.

The basic information about each case was recorded in an Excel spreadsheet such as the case number, date of review, testing technique, examiner decision, quality control decision, and whether the case included an EDA score affected by the filtering mode. A separate spreadsheet listed all cases in which the EDA mode affected the score(s), as well as the respective chart number, question number, whether there would be a different result for the two EDA modes, and if so, what those results would be.

The field polygraph reports from the cases were not available to the reviewers. Consequently, there was virtually no information regarding ground truth. For this reason, it would not be possible from this study to determine which EDA mode offered better decision accuracy than the other.

Results

Overall, we found 189 cases (24.9%) out of the 760-case data set in which at least one EDA

Table 1. Percentage of cases by technique for those with differences in EDA scores against the same percentages in the entire data set. Columns may not total to 100% due to rounding.

	Percentage of Cases with Different EDA Scores (n=189)	Percentage of All Cases in Data Set (n = 760)
2-RQ AMGQT	60.3	61.3
3-RQ AFMGQT	24.9	22.2
BOST	14.8	15.7
DLST	0.0	0.0
NA*	0.0	0.3

* Indicates a non-standard method was used and was overturned by the Quality Control review.

score had been different between the automatic and manual modes.

We also wanted to determine whether the prevalence of differences in EDA scores depended on the testing technique. Table 1 lists the percentages of cases by technique. A test of proportions for each technique found no significant differences in percentages between the cases with EDA score differences and the percentage of cases in the entire data set. No significant differences were found between percentages of the entire data set and those in the subset of cases with differences in EDA scores.

Frequency of Differences in EDA Scores

Among the 189 cases containing a difference in EDA score due to the filter, some cases had more than one difference. A total of 41 cases (21.7%) out of 189 had two scores that were different in the case, and 9 (4.8%) had three scoring differences. The remaining 139 cases (73.5%) had a single difference in EDA score per case.

Table 2. Cross tab showing the number of cases by decision for the two different EDA modes in which there was at least one score affected by the filter. n = 189

		Decisions from Manual EDA			
		NSR	INC	SR	Totals
Decisions from Automatic EDA	NSR	66	13	1	80
	INC	15	16	16	47
	SR	0	4	58	62
	Totals	81	33	75	189

Effect of Different EDA Scores on Examination Results

Table 2 is a cross tab for decisions that were based on the EDA in its two filter modes and conventional ESS cutoff scores. Of the 189 cases having at least one difference in EDA score between modes there was no impact on the final results in 140 of the cases (74.1%). This left 49 cases in which changing the EDA mode would also change the test results. For 48 of the 49 cases the change in EDA mode pushed the final decision into or out of the Inconclusive zone. A single case showed that the manual EDA and auto EDA scores produced opposite decisions (SR v NSR). The 49 cases in which decisions were affected represented 6.3% of the entire 760-case data set, and the opposite decision 0.1%.

Table 2 suggests the overall Inconclusive rate was higher for the automatic mode than it was for the manual mode (47 cases versus 33 cases), however this difference was not found statistically significant ($z = 1.77$, ns). Similarly, the apparently higher rate of SR decisions in the manual mode over those of the automatic mode (75 cases versus 62 cases) did not achieve significance (z = 1.39, ns).

Overall Calculation on Frequency of EDA Differences and Examination Results

A first approximation can be calculated for how likely any individual EDA score would be affected by the filtering mode. While we did not track the number of polygraph charts recorded in each case, we do know that a minimum of three presentations of each question is required for all the techniques used by this program. We also know the number of relevant questions for the techniques. If the numbers of charts are approximately equal for those cases producing a difference in EDA score compared to the entire study data set, we can calculate in round numbers that about one EDA score in 20 would be affected by the EDA mode. Likewise, if an EDA score is found to be different between EDA modes, there would be a 1-in-4 chance the test results will be affected by the EDA mode.

Discussion

Possible Factors and Impacts of EDA Modes on Decisions

Though a quarter of all cases had at least one EDA score affected by the filter, the relatively small effect on decisions of 6.3%, and only one opposite result, requires an explanation. There are factors that almost certainly constrained the potential impact of the EDA modes on decisions. One is that polygraph decisions are based on the data from all data channels. While the EDA serves an outsized role in decision-making, data from the breathing, cardiograph and photoplethysmograph also contribute significantly. The additional data could blunt the impact of variability caused by the EDA modes. In addition, polygraph decisions are based on recording from three to five polygraph charts. Consequently, an errant EDA score among all the data collected across the charts can have its influence dampened by the remaining data. Finally, most of the techniques used by this polygraph program have a decision rule such that a decision of SR or Inconclusive to any one question produces an SR or Inconclusive result for the entire test. Therefore, a change of EDA score on a question different from the one that led to an SR or Inconclusive decision will have no consequences on decision-making. These considerations, in combination, could help explain why a quarter of all cases had an EDA score affected by the choice in filter but decisions were only affected in about one case in 16.

Standardization Considerations

Owing to the source of the data, this study could not identify whether either of the EDA modes led to better decision-making. The study has shown there is a non-zero difference in results between the modes in manual scoring. Nevertheless, the prospect that electrodermal modes can "label truthful people as liars and the guilty as innocent" as asserted in the McClatchy News article (Taylor, 2013) appears for the moment to be very limited. In the present study there was only one case among 760 cases in which the electrodermal mode produced opposite decisions, or 0.13%.

The findings in this project do imply possible consequences for the field. For standardization purposes it would appear prudent for the development of professional standards that recognize the potential for disagreements that can arise from the choice of EDA filter. The current policy gap allows two professional polygraph examiners to assign different EDA scores to the same data if displayed in different filtering. This is an obvious problem and is most germane to polygraph cases conducted for evidentiary purposes.

On a more local level, government, law enforcement, and offender management polygraph programs may wish to establish policy regarding which EDA mode will be used by its examiners. Doing so could reduce a 6.3% difference in test results from EDA modes to possibly 0%. For examiners working outside these domains, these examiners would be well served by assessing the current state of the evidence, and consistently using the EDA mode they are prepared to defend if necessary.

Authors' Conclusions

It could be argued that if one EDA filtering approach provided superior decision accuracy over another, that method should be preferred. Taken to its logical conclusion, the best EDA mode for manual scoring should be the one recommended by the manufacturer based on published evidence, ensuring not only better decisions but higher agreement among human scorers because all would be scoring the data displayed in the same way. The variety of EDA filters found within and across different manufacturer products suggests manufacturers are not yet prepared to make such a recommendation (however, see Lykken & Venables, 1971). Our data show the signal is being altered between the automatic and manual modes in ways that affect manual scores, and while we are unable to identify which mode is more correct, the existence of a difference is noteworthy. Because it is accepted that the manual mode imposes less filtering on responses than does the automatic mode, the manual mode arguably captures a more faithful representation of the electrical properties of the skin. Our field experience leads us to conclude the manual EDA mode tends to represent response onset latency better (refer again to Fig. 1). For this reason, the present authors have concluded that the least filtered option is the better choice for manual scoring in all but exceptional cases, e.g. electrodermal data with extreme tonic levels or those with large tonic shifts. This position may be revised once further research evidence is available. In the meantime, professional agreement on which EDA mode to use for manual scoring seems eminently reasonable and likely to reduce inter-scorer differences.

Limitations

The scoring of electrodermal responses in this study required the two compared EDR amplitudes to be objectively different by at least 10%. The present findings may not generalize if scorers use a different, or no specified minimum threshold for differences when assigning scores.

Only Lafayette Instruments polygraphs were used in this study. While the other manufacturers also have more than one EDA mode, the effect of those modes on manual scoring may be different from what we found. We encourage research to explore what those effects might be.

Nothing in this project could determine which EDA mode was better in terms of fostering manual scores that corresponded more closely with ground truth. Ground truth was largely unavailable in this sample. Also, we did not evaluate whether any form of automated analysis would have produced different results. Our findings are restricted to only demonstrating that an examiner's choice of EDA modes in manual scoring can affect polygraph scores and decisions.

The units of measure for this project were polygraph scores and decisions at the case level. Future researchers may wish to control for test technique and number of test charts to further refine the estimated effect of EDA modes on individual scores.

The scoring method used in this study was the Empirical Scoring System. It is unknown whether there would be similar findings with 3- or 7-position scoring, rank order methods, or global analysis. Many questions remain.

References

Arther, R.O (1971). The GSR unit. *Journal of Polygraph Studies,* 5(6), 1 – 4. Cited in Matte, J.A. (1996). *Forensic psychophysiology using the polygraph: Scientific truth verification-lie detection.* Williamsville, NY: J.A.M.

Blalock, B., Cushman, B., and Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph, 38*(4), 281 – 288.

Cochran, M.A., and Fuentes, R.C. (2018). Stoelting Co. response to article, "The difference between the manual and automatic settings for the electrodermal channel and a potential effect on manual scoring." *Polygraph & Forensic Credibility Assessment, 47*(1), 49 – 52.

Handler, M., Nelson, R., and Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph, 37*(2), 100 – 111.

Kalafati, A., and Krapohl, D.J. (2018). The difference between the manual and automatic settings for the electrodermal channel and the potential effect on manual scoring. *Polygraph & Forensic Credibility Assessment, 47*(1), 37 – 44.

Krapohl, D.J. (2020). Electrodermal responses: When is bigger really bigger? *Polygraph & Forensic Credibility Assessment, 49*(2), 104 – 109.

Krapohl, D.J. (2021). Another look at electrodermal response ratio minima. *Polygraph & Forensic Credibility Assessment, 50*(1), 59 – 66.

Krapohl, D.J., and Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph, 35*(1), 55-63.

Krapohl, D.J., Grubin, D., Benson, T., and Morris, B. (2020). Modification of the AFMGQT to accommodate single-issue screening: The British One-issue Screening Test. *Polygraph & Forensic Credibility Assessment, 49*(2), 176 – 183.

Krapohl, D.J., Halford, K.H., Benson, T., Mayston, A., and Dutton, D.W. (2021). Response onset latencies of electrodermal, cardiovascular and vasomotor responses during field polygraph testing. *Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice, 50(*1), 15 – 27.

Krapohl, D.J., and Shaw, P. (2015). *Fundamentals of Polygraph Practice.* Academic Press: San Diego, CA.

Limestone Technologies (2018). Limestone Technologies response to APA EDA publication. *Polygraph & Forensic Credibility Assessment, 47(*1), 48.

Lykken, D.T., and Venables, P.H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology, 8*(5), 656 – 672.

Matte, J.A. (1996). *Forensic psychophysiology using the polygraph: Scientific truth verification-lie detection.* Williamsville, NY: J.A.M.

Nelson, R. (2018). Electrodermal signal processing: A correlation study of auto-centered and manually centered EDA. *Polygraph & Forensic Credibility Assessment, 47*(1), 53 – 65.

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.

Nelson, R. (2020). Bigger is better for automated scoring: Analysis of minimum constraints for RQ/CQ ratios. *Polygraph & Forensic Credibility Assessment, 49*(2), 110 – 120.

Raskin, D.C., Barland, G.H., and Podlesny, J.A. (1978, June). *Validity and Reliability of Detection of Deception*. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S., Department of Justice. Grant Number 75-NI-99-0001.

Sjouwerman, R., and Lonsdorf, T.B. (2019). Latency of skin conductance responses across stimulus modalities. *Psychophysiology, 56*(4), https://doi.org/10.1111/psyp.13307.

Taylor, M. (2013, May 20). *Glitch in Widely Used Polygraph Can Skew Results*. McClatchy Washington Bureau. Accessed December 31, 2021 at https://www.mcclatchydc.com/news/special-reports/article24749257.html

Lafayette Instrument Response to Krapohl & Dutton on EDA Signal Processing Recommendations

Raymond Nelson

As a technology provider, Lafayette Instrument Company believes it reasonable to recommend that customers use the EDA mode that best serves their stated objectives. For many users, that will be the mode that best permits the separation of signal and noise while providing the strongest correlation with the criterion of interest and the highest accuracy rate. No scientific test is expected to be infallible, and tests are often optimized by reducing the error and noise in the data. For this reason, the optimal EDA solution for most users, the one that will be most reliable and most often correct, will be the one that is least often incorrect. In our analyses, a correctly designed auto-centering EDA solution can be the most effective method of optimizing desired effect sizes and criterion coefficients while also achieving the practical goals of accuracy, reliability, and usability of the EDA data. For this reason, Lafayette Instrument Company recommends that most users, especially those primarily interested in phasic vs. tonic EDA data, consider using the auto-centering EDA mode. Lafayette Instrument Company also recognizes that different users may have different goals and objectives, and for this reason will always help users to select the optimal technology and procedures that best serve their individual needs.

Background

Electrodermal activity (EDA) is an umbrella term used to describe complex phenomena associated with the skin, an organ of the integumentary system. Different methods exist to record and monitor EDA. The passive, known as endosomatic, method utilizes the electrical potential (difference in voltage) measured between two contact points, and the active, known as exosomatic, method utilizes an external electrical source applied to the skin and monitored for changes in electrical current or voltage. [Refer to Handler, Nelson, Krapohl & Honts (2010), and Nelson, Handler & Smitley (2017) for more information.] Exosomatic EDA can be measured using either constant-current or constant-voltage circuit designs and can be expressed using either conductance or resistance units.

Polygraph field practitioners have traditionally expressed EDA values in resistance units, whereas research psychophysiologists have expressed a preference for standardized use of conductance units. However, this difference is inconsequential due to the exact mathematical relationship between conductance and resistance. If one is known, the exact value of the other is also known – with no information loss. More importantly, electrical units of either resistance or conductance are not actually synonymous with EDA. Non-living things, such as electronic parts, will also have resistance and conductance. In contrast, EDA is a complex phenomenon subject to human behavior and physiology for which, despite the breadth of our present knowledge, it is likely that much will remain unknown for some time into the future.

Resistance and conductance units are a convenient proxy for EDA. This is similar to the way that EDA is a proxy for autonomic activity, a complex system physiological activity, even though EDA is not synonymous with autonomic activity. In the same way, autonomic activity is a proxy for deception and truth-telling in the comparison question polygraph context. There is a statistically significant and usable correlation between autonomic activity and deception, even though autonomic activity is, of itself, not synonymous with deception and truth-telling. All human physiology is correlated with multiple behavioral and environmental factors. When the strength of these relationships (resistance->EDA-> autonomic activity -> deception or truth-telling) is sufficient, and when we can combine different sources of information in a synergistic model, we can increase the likelihood of correct classifications beyond what can be achieved by chance or random guessing or by unassisted human judgment. As it happens, EDA is among the most important and useful signals during polygraph testing [See Nelson (2019) for a survey of weighting coefficients for polygraph sensors].

EDA has been described as consisting of two types of activity. These include: *electrodermal response* (EDR, known as phasic EDA) and *electrodermal level* (EDL, known as tonic EDA). In the context of a polygraph comparison question, EDRs, or phasic changes in EDA, are responses to test stimuli. EDL is sometimes referred to more descriptively as the EDA *baseline*. Research psychophysiologists who are interested in advancing our theoretical knowledge about EDA will be interested in both EDL and EDR. Some professionals, such as those studying sleep, sedation, or pain, may be primarily interested in tonic EDA. In the context of comparison question test, polygraph professionals are primarily interested in phasic EDA. This is because EDRs have been shown to correlate with the criterion of deception and truth-telling at levels that are both statistically significant and practically useful. In contrast, EDL has not been shown to be significantly correlated or useful in the comparison question polygraph test. Neither manual nor automated polygraph scoring systems are known to make use of tonic EDA. Moreover, excess tonic noise can increase the difficulty and decrease the reliability of correct identification of EDRs.

A useful perspective to understand polygraph test data is that it is *time series* data. That is, data values are recorded repeatedly at a specified and consistent time interval and are strung together in a time series that can be displayed or plotted (printed) to appear as a graphical line when viewed or evaluated visually. The question of interest is this: how do the numerical values change relative to the consistent sampling time interval? Another useful perspective to understand is that all time series data are frequencies. With the exception of sine waves and other pure waveforms, nearly all time series data are a composite of numerous different frequencies – often in a particular spectrum or frequency range of interest.

Observed changes in data values, visualized as changes in a line-plot or graphic waveform, are actually an indication of changes in the frequencies contained in the data. More precisely, *observed changes in time series data values are an expression of changes in the relative strength of the different frequencies contained in the time series waveform*. Mathematical methods, such as Fourier transforms, can be used to transform time series data to the frequency domain. This allows us to deconstruct, analyze, study, and graph those different frequencies and can also be used to reconstruct the time series waveform from the frequency domain information. [Refer to Figure 1 through 4 in Nelson (2018) for a graphic example of time series EDA data transformed to the frequency domain.] Polygraph EDA responses can be thought of as low-frequency information; lower frequency than cardio pulse information, in the range of 60-100 cpm (1Hz to 1.7Hz), and lower than respiration data, in the range of 10-22 cpm (0.17Hz to 0.37Hz).

Phasic EDA frequencies of interest to polygraph testing have been found to be in the range from 0.03Hz, a waveform that takes 33 seconds to complete a full cycle, to 0.5Hz, a waveform that permits us to observe the rapid onset of a phasic change in activity. An EDR with a duration of 5 seconds will have a frequency near 0.2 Hz (1 / 5 = 0.2), while a response with a duration of 12 seconds will have a frequency near 0.08Hz (1/12 = 0.083) and a response of 25 seconds duration will be closer to 0.04Hz. Some of the development and validation experiments that led to the Lafayette EDA signal processing solutions are described in Nelson (2018). Frequencies above 0.5Hz are high-frequency noise that can contribute to unreliability when attempting to accurately locate the onset and peak of EDA response.

Tonic EDA can be thought of as the lowest of recorded frequencies, near 0.01Hz or lower. These frequencies can be observed as waveforms that plot or cycle over a period of approximately 100 seconds or more. When the magnitude of tonic EDA is less than the magnitude of phasic EDA, the tonic EDA will appear visually as a somewhat stable baseline.

An interesting phenomenon observed in polygraph testing is that many individuals are known to have unstable tonic EDA. What this means, in terms of the recorded time series data, is that the magnitude or amplitude of very low-frequency tonic information may be substantially stronger than the magnitude of phasic information. In practical terms, unstable tonic EDA will tend to drift out of range of the visual range of the data plot. This may require user attention and management effort in the form of recentering or relocating the data during testing. Also, unstable tonic EDA may increase problems with unreliability in EDA feature extraction. In more difficult cases, the magnitude of tonic information can begin to eclipse or mask the observable information contained in phasic responses. This is similar to how loud noise from a train or airplane can interfere with our ability to hear a nearby person even though they may be talking directly to us.

Ideally, during polygraph testing, the magnitude of frequencies in the range of phasic EDA will be greater than the magnitude of the tonic EDA frequencies. It is for this reason that engineers, about 70 or 80 years ago, began to include auto-centering EDA solutions in polygraph field instruments. The goal of an auto-centering EDA solution will be to remove excessive tonic information while preserving phasic information of interest. In practical terms, the auto-centering EDA will require less user effort to return to a baseline level.

Auto-centering EDA solutions can be constructed in both hardware and software, and the math is the same in both methods. Hardware filters are constructed of parts such as resistors and capacitors, and in some cases, amplifiers (though polygraph EDA solutions do not typically include amplification designs). Software solutions will be constructed in computer code. Polygraph instruments with a hardware EDA switch are likely to have hardware solutions, while EDA modes that are selected via software settings are more likely to be implemented in computer code. Although the math and design are similar for these two methods, there are practical and economic advantages to a software solution – often referred to as digital signal processing (DSP). Also, different polygraph instruments may include auto-centering EDA solutions with different specifications (corner frequencies and other filter characteristics). Because not all auto-centering EDA designs are the same, documentation and validation of an EDA solution is an important consideration. For this reason, Lafayette Instrument Company has taken the forthright position of publishing the auto-centering EDA design and performance characteristics and believes that the polygraph profession should expect this of all polygraph instruments today.

The prototypical polygraph auto-centering EDA solution is a high-pass filter that reduces or attenuates the amplitude of low-frequency noise below a cutoff or corner frequency, thereby permitting the signal of interest to be more easily extracted from the time series data. Today there is no area of science, technology, or data analysis involving time series data that ignores the importance or value of filters and DSP when seeking to make the most efficient use of available data. Algorithm developers in all areas of science and technology make use of a variety of signal processing options to optimize signal-to-noise ratios and the information they can extract from data. For example, in polygraph data analysis, Harris, Horner & McQuarrie (2000) described the use of a "defuzzification" filter prior to feature extraction with respiration data. [See Nelson (2013) for an example of how this filtering improves the criterion coefficients of respiration data.]

It is important to understand that when dealing with numerical and digital information, there is no such thing as "raw" data. Even an old-fashioned analog polygraph in manual-centering EDA mode will be subject to data filtering, albeit undocumented, through the inertial mass of the moving parts and the friction of the ink pens on the moving paper. All data require some degree of recording and processing to be useful. Filters are a ubiquitous and desirable aspect of all types of technology. For example, virtually any electronic device that deals with time series information or sound will include a 60hz filter to remove inherent noise from the alternating current in our 120V mains.

All data of any type, whether polygraph or any other domain of testing or data analysis, is a combination of *signal* and *noise*. The terms signal and noise are descriptive and metaphorical terms to denote the distinction between data that are useful to a particular goal or objective and data that are not useful and may begin to interfere with a particular goal or objective. Interestingly, what is signal in one context may be noise in another and vice versa. In other words, virtually all information that is not of interest for a particular purpose can be thought of as noise from which the signal of interest must be isolated and extracted. Well-designed signal processing solutions will improve or increase desired signal-to-noise ratios.

The signal of interest in polygraph is phasic activity (higher EDA frequencies) that occurs in response to test stimuli. In this case, noise is any information that is inconvenient, unwanted, uninformative, uninteresting, or interfering with an ability to extract the signal of interest. Tonic activity (low EDA frequencies) is commonly thought of by polygraph examiners as baseline information from which the onset of phasic EDAs can be easily determined. When the magnitude or amplitude of low-frequency tonic EDA is greater than that of the higher frequency phasic EDA of interest, it can become obvious that low-frequency tonic EDA (observed as unstable tonic EDA) may be inconvenient noise.

In the context of polygraph, manually-centered EDA data may be considered the noisiest, most difficult, and least reliable EDA solution. In contrast, a correctly designed auto-centering EDA may offer the best signal-to-noise ratio. An optimal solution will maximize the ability for field examiners to reliably extract useful signal information from inherent noise and will be recognized by an improved correlation with deception, truth, and improved diagnostic coefficients. At this time, there is published evidence that the criterion coefficients of a correctly designed auto-centering EDA solution can equal or exceed that of manually-centered EDA data. [See Nelson (2018) for more information.]

Different signal processing solutions will make different frequencies of interest more or less easily identifiable. Therefore, regardless of which EDA mode is selected for use in field practice, it is within the realm of possibility that different EDA solutions could, under some circumstances, produce different scores. It is also within the realm of possibility that these different scores may, at some times, lead to different test results. This possibility exists because of the simple fact that different is, well, different. It would be neither reasonable nor realistic to expect the outcomes different signal processing solutions to always be the same. If they were always the same, they would not be different. At this point, the matter becomes tautological.

It would be incorrect to try to think of one solution as correct or true and another solution as incorrect, untrue, or even less true. This would be a distortion of the epistemological concept of truth, including what is truth, what kinds of things can be true, and what it means to say that something is true. Different EDA signal processing methods simply allow us to more easily access different aspects of the EDA data. If the goal is to increase the effectiveness and reliability of the polygraph test, a pragmatic solution will be to select the EDA mode that maximizes the criterion coefficients and effect sizes of interest. The criterion of interest to the polygraph test is deception or truth-telling. The effect size of interest may be correct decisions or the constraint of errors to a minimum level.

If one of the practical goals is to avoid the sense of uncertainty that may result from the potential that some polygraph scores may differ for the different EDA signal processing methods (most likely for labile and difficult EDA data), then a simple, though perhaps unnecessary, solution may be to regard as inconclusive any examination for which different results can be achieved via different methods. The problem herein will be that, if all EDA solutions are regarded as equal (which they are not), there will always exist the possibility that some EDA signal processing solution can be devised to provide a different result, thereby nullifying any examination. Another equally unsatisfactory solution will be to arbitrarily, without analysis of criterion coefficients and effect sizes, begin to standardize on a single solution as correct or acceptable or best practice. Adopting solutions without evidence can have the short-term advantage of making people feel more confident while imposing a long-term consequence of reducing their ability to make use of scientific knowledge and evidence when it becomes available. Remembering that polygraph does not physically measure or detect deception and that all scientific test results are probabilities, the question of greatest importance here is this: what solution is correct most often and incorrect least often?

There is little justification for the selection of a signal processing solution that is sub-optimal. An optimal solution will provide an improved signal-to-noise ratio, improved criterion coefficients (i.e., correlation with the criterion and coefficients of determination), and improved outcomes effects (e.g., DEC coefficients, correct decisions, minimized errors, etc.). A correctly designed auto-centering EDA solution, such as that used in both the LXSoftware and LXEdge platforms will be recognizable when these metrics equal or exceed those of the manually-centered and other EDA solutions. A corollary to all of this is that an incorrectly designed auto EDA, one that removes frequencies from the range of interest, may damage the signal of interest to the polygraph test. It is for this reason the polygraph profession should hold some reasonable expectation of transparency and accountability for the design solutions employed in all field testing instruments.

Lafayette Instrument Company has previously published the design and validation of its auto-centering EDA solution, as described in the publication by Nelson (2018). The auto-centering EDA mode in both the LXSoftware and LXEdge platforms was found to equal or exceed the criterion coefficients of the manually-centered EDA data and all previous managed EDA solutions. Description of the design and evolution of the EDA circuit is also available online (Lafayette Instrument Company, 2013). In addition to leading the polygraph profession in terms of EDA signal processing and feature extraction, Lafayette Instrument has also been a leader in terms of making education and information available to polygraph professionals who may have a desire to more fully understand the instrumentation they depend upon. A recent survey of published information could not locate any publications that provided evidence of a more effective solution or any evidence of a similar level of transparency, documentation, and dedication to the scientific validation of polygraph field instrumentation.

All different solutions may have some inherent advantages and disadvantages. Manual EDA provides the greatest access to both tonic and phasic EDA data and may be of interest to those doing basic science research intended to advance our theoretical knowledge about EDA, but may also have the highest level of unwanted noise (lowest signal to noise ratio) and may require the greatest user effort to manage the EDA data for individuals with unstable tonic EDA. The Lafayette detrended EDA mode will provide users with a managed EDA solution that will effectively handle most common manifestations of tonic instability while providing visual information for which phasic responses offer a very high degree of correspondence with manual EDA, but for which tonic information can appear flat and unresponsive for some individuals. The auto-centering EDA mode in the LXSoftware and LXEdge platforms have been shown to maximize the criterion coefficients and effect sizes of interest to polygraph examiners while providing users with a managed EDA solution.

EDA data are closely associated with autonomic activity and a variety of other phenomena in addition to deception and truth-telling. Because correlations are intended to quantify the strength of an imperfect (and therefore probabilistic) relationship, there will always be some margin of error in the EDA data. In simplistic terms, we can expect that an optimal EDA solution is one that maximizes the criterion coefficients and effect sizes of interest. The optimal solution will be the one that produces EDA scores that are most often correct and less often incorrect when compared to other potential solutions.

An evidence-based approach will regard as questionable any standardized solution that is premised on conjecture, persona, ideology, or opinion without evidence in the form of criterion studies and published effect sizes. Standardized solutions have the effect of encouraging people to believe they know the optimal solution to a problem. Although there may be situations when it is acceptable to standardize a solution to simply help mitigate a crisis by reducing chaos, selecting an EDA standard that is either without evidence or inconsistent with published evidence would introduce a potential long-term hazard to the scientific knowledge and integrity of the polygraph profession. Short-term benefits could easily be outweighed by the long-term costs and difficulties associated with the need to adjust or alter field practice standards to align with scientific evidence at a later time.

In practical and applied matters, and especially in the absence of a crisis, it is often preferable to take the time to formulate field practice policies that are based on scientific evidence from the outset. In practical terms, this means that proposed standardized solutions should be based on the study of relationships between different EDA solutions and the criterion of interest, in this case, deception or truth-telling and not merely a comparison of differences between EDA modes. It is worth reminding ourselves of the scientific method of observation, hypothesis formulation, experimentation, analysis, and rejection of hypotheses that are not supported by empirical evidence and the scaffolding of scientific theories that are supported by evidence. In this context, experimentation must include a study of scores and outcomes for different EDA solutions with the criterion of deception and truth. With this in view, it is important to recognize that a comparison of outcomes for different EDA modes without the criterion of deception and truth-telling, although somewhat informative, does not achieve the requirements of a criterion study because it cannot, by design, answer the question as to differences in effect sizes.

Selection of any proposed industry-standard EDA solution, or any recommended EDA solution, should remain an evidence-based scientific process. Such evidence must be in the form of criterion coefficients and effect sizes and cannot be limited to the study of correlation and concordance. Field practice recommendations for a proposed EDA signal processing standard will be most effective if they are correctly informed as to the basis of available evidence with an understanding of the complex issues surrounding the design and selection of a signal processing solution. Field practice standards should also be informed by complete information about different use-case scenarios that can accompany the advantages and disadvantages of different possible solutions.

For more information, please contact polygraph@lafayetteinstrument.com.

References

Handler, M., Nelson, R., Krapohl, D. & Honts, C. (201). An EDA primer for polygraph examiners *Polygraph, 39*(2), 68-108.

Harris, J., Horner, A. & McQuarrie, D. (2000). *An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations*. Johns Hopkins University, Applied Physics Laboratory. SSD-POR-POR-00-7272.

Lafayette Instrument Company (2013, March 1). LX4000 & LX5000 EDA circuit design and evolution. Retrieved Feb 2, 2022, from https://lafayettepolygraph.com/pages/eda-info/

MacLaren, V. V. (2001). A qualitative review of the guilty knowledge test. *Journal of Applied Psychology, 86*(4), 674-683.

Nelson, R. (2018) Electrodermal signal processing: a correlation study of auto-centered EDA and manually-centered EDA with the criterion state of deception and truth-telling. *Polygraph and Forensic Credibility Assessment,* 47(1), 53-65.

Nelson, R. (2019). Literature survey of structural weighting of polygraph signals: Why double the EDA? *Polygraph & Forensic Credibility Assessment, 48(*2), 105-112.

Nelson, R., Handler, M. & Smitley, B. (2017). Practical polygraph: FAQ on electrodermal activity and the electrodermal sensor. *APA Magazine, 50*(3), 60-76.

Final Thoughts on EDA Modes

Donald J. Krapohl

Donnie W. Dutton

We appreciate that the Lafayette Instrument Company, and Mr. Nelson in particular, have championed the issue of transparency in EDA filtering methods. The continuing absence of industry standards for how polygraph data are collected and displayed shows how much reliance polygraph examiners place in the design specifications chosen by the instrument manufacturers. Mr. Nelson's publication of EDA filtering specifications was an important decision, and we recognize it takes some business courage to make that information publicly available. We hope all manufactures will do likewise, and that generally agreed-upon standards will be on the visible horizon. In the meantime, transparency is the best first step.

We offer no disagreement with Mr. Nelson's response to our paper, save one. In his response he cited an earlier paper (Nelson, 2018) in which various EDA filters were compared against one another and offered that the current Lafayette auto-centering filter "…was found to equal or exceed the criterion coefficients of the manually-centered EDA data and all previous managed EDA solutions." This statement is technically correct but may mislead those who do not read the Nelson (2018) paper and look specifically at the difference in correlation coefficients between the automatic and manual modes. They were different by only 0.01. A test of significance for the two coefficients with the modest sample size would reveal that any apparent difference was illusory and not meaningful.

Relatedly, the feature extraction in Nelson (2018) was automated and used logged R/C ratios of EDR amplitudes. The relationship between this approach and conventional manual scoring has not been established. As such, the study did not perfectly model what human scorers do, lending its findings more to algorithmic than human approaches to EDA analysis. Taken all together, there is insufficient reason to conclude that auto-centering the EDA signal improves human scorer performance over the use of manual mode.

In fairness, our paper did no better at resolving which filtering approach performed best because we only examined the effect of the filters on polygraph decisions, not on accuracy. We did argue for our preference for the manual mode except under special circumstances due in part to a demonstrated possibility the automatic mode can influence the appearance of EDR onset latency. We concede our opinion is shaped by our field experiences and that the incidence of onset latency distortion has not been systematically investigated. We and others have seen EDR onset latency distortion in the field due to the self-centering filter. We have yet to count how often. So, in sum, the basis for our preference for manual EDA as well as Mr. Nelson's argument for self-centering EDA both lack sufficient empirical support.

Sadly, this leaves us where we began. There is no evidence that either self-centering or non-centering EDA modes will lead to higher decision accuracy in human scoring. We only know they can lead to different decisions. Perhaps one day Mr. Nelson will be shown to be correct about the superiority of auto-centering, or we regarding the importance of manual filtering. Perhaps we are both correct under different conditions. What we have shown with our data, for which no one disagrees, is that the reliability of polygraph decisions among examiners may be affected by their choice of EDA settings. That fact, we would argue, is cause for concern.

Reference

Nelson, R. (2018) Electrodermal signal processing: a correlation study of auto-centered EDA and manually-centered EDA with the criterion state of deception and truth-telling. *Polygraph and Forensic Credibility Assessment, 47*(1), 53-65.

Multinomial Reference Distributions for Seven-Position Likert Scores of Comparison Question Polygraph Tests

Raymond Nelson

Abstract

Multinomial reference tables were calculated for seven position Likert scores of comparison question polygraph tests. Although Likert score are somewhat inherently subjective, with no theoretical justification for the boundaries between scale items, calculation of the multinomial reference model was possible under the null hypothesis to the analytic theory of the polygraph. Reference tables of this type can be used to calculate posterior probabilities associated with observed test scores and are useful as a likelihood function in Bayesian analysis. They can also be used to calculate numerical cutscores to achieve desired posterior effect sizes. Although multinomial tables have been previously available for three-position scoring methods, including the U.S. Federal three-position scoring method and the Empirical Scoring System, this project advances the polygraph profession by providing a multinomial reference model with application to the tradition seven position scoring scale.

Introduction

The seven-position scoring method (Bell, Raskin, Honts & Kircher, 1999; Department of Defense, 2006; National Center for Credibility Assessment, 2017) is a semi-objective procedure for manual scoring of comparison question test (CQT) data. The traditional Likert scale for comparison question polygraph scores is referred to by polygraph professionals as the *seven-position scale* and includes the values [-3 -2 -1 0 +1 +2 +3]. Scoring procedures involve the assignment of seven-position scores to analysis spots using the analytic theory of the polygraph test. An analysis spot consists of a relevant and comparison question. The seven-position scale is an example of a *Likert scale* (Likert, 1932), and seven-position scores can be thought of as *Likert scores*. Likert scales, and Likert scores are a common solution for numerically coding data in psychology and social sciences because they are an effective way to transform subjective information to numerical data. Numerical data is useful because it serves to document an analytic process. Documentation is an important aspect of reproducible analyses (Peng, 2011), which is fundamental to scientific research and scientific testing.

A prototypical Likert scale will solicit information about the degree to which a person agrees with a stated idea, providing a range of options including non-agreement (often coded numerically as 0) and often including options as to the degree of agreement with an idea. The degree of agreement is often coded using integers (i.e., 1, 2, 3) that signify the degree or level of agreement. In practice, Likert scales exist in many forms, with some scales including the use of negative integer scores that signify the degree of agreement in an inverse or opposite direction. For example, a person may be asked to respond to a statement such as "I like Mondays," with options ranging from (-2, -1, 0, 1, 2) indicating strong disagreement, mild disagreement, neutral attitudes toward Mondays, mild agreement, and strong agreement. The number of response items in a Likert scale is determined by the information requirements of each use-case, and may commonly include 3, 5 or 7 options, but can be adapted for any number of items as indicated by the use-case or context. Likert type scales have been used when manually scoring comparison question polygraph tests data since the early 1960s (Kubis, 1962, Backster, 1963) and continues to this day (Bell et al., 1999; Handler and Nelson 2008; National Center for Credibility Assessment, 2017).

Likert scores are assigned by the examiner using the analytic theory of the polygraph test that greater changes in physiological activity are loaded at different types of test stimuli (i.e., relevant and comparison questions) as a function of deception or truth-telling in response to relevant target stimuli. [See Nelson, (2016) for a discussion.] Evaluation of the loading of scores is easily accomplished by summation. Truthful persons have been shown to have generally greater changes in physiological responses, and corresponding numerical scores, loaded at comparison stimuli. In contrast, greater changes in physiological activity of deceptive persons have been shown to be loaded at relevant target stimuli. Because all test data is subject to variability, and because a goal of testing is to separate random or unexplained variation from systematic or explained variation, testing procedures include the use of multiple repetitions of an array consisting of multiple relevant and comparison questions, in addition to other procedural questions.

Some polygraph scoring methods make use of a three-position Likert scale, including the Federal three-position method (Department of Defense, 2006) and the Empirical Scoring System (ESS; Nelson et.al, 2011). Whereas traditional Likert scores are a solution for the quantification of information that is inherently subjective, three-position polygraph scores can be argued as objective when feature extraction is based on objective methods. This nuance is largely inconsequential for field practitioners, though potentially important when considering that objective feature extraction will remove analytic subjectivity as a source of unexplained or uncontrolled variation. Remaining sources of uncontrolled variation will be limited to the test administration and to the individual examinee. Multinomial reference distributions are available for the three-position and ESS-M scoring methods (Nelson, 2017; 2018a). However, no multinomial reference models have been previously available for the seven-position scoring method. Reference distributions are useful in both practical and scientific ways because they can be used as a likelihood function for Bayesian analysis and quantification of the practical likelihoods of deception or truth-telling. This project involved the calculation, using Monte Carlo methods, of multinomial reference tables for seven-position Likert scores of comparison question polygraph test data.

Method

Multinomial reference distributions were calculated for the seven-position Likert scores of comparison questions polygraph tests consisting of a question sequence that includes two to four relevant questions along with comparison questions and other procedural questions. A distribution of this type, like all distributions, is a list of all possible values under the design, along with the statistical likelihoods associated with each score, under the theory of the test.

Under the analytic theory of the polygraph test, different scores within the Likert scales occur with loaded, unequal frequency depending on whether an examinee has been deceptive or truthful. As it often happens in many scientific endeavors, it is more realistic and useful to calculate the distribution under a null hypothesis that directly for a theory, because the null hypothesis can often be easily characterized as random. In contrast, calculating a distribution under an alternative hypothesis is often difficult. Under the null hypothesis to the analytic theory physiological activity, and resulting numerical scores, are not loaded at the different types of test stimuli in any systematic way. In the polygraph testing context, the null hypothesis holds that changes in physiology that contribute to each of the different possible scores will occur with equal frequency, resulting in aggregated scores that do not differ significantly from zero.

Field polygraph tests consists of up to five iterations of the sequence of test questions, with a floor requirement of three iterations. Polygraph data analysis is multivariate in that different sources of data are integrated to achieve an analytic result. Data sources include respiration data, electrodermal (EDA) data, cardiovascular activity. Vasomotor activity can also be included but is not presently a required standard practice. Tables were calculated separately for examinations with two, three and four relevant questions. [See Nelson & Krapohl (2017) for an example of a manual score sheet for comparison question polygraphs.] A separate table was calculated for subtotal scores of up to five iterations of individual RQs, consistent with field practices for multiple issue screening polygraphs in which each relevant question may be evaluated with an assumption of independent criterion variance.

Seven-position Likert scores of comparison question polygraph tests are multinomial because each score is taken from multiple possible values. Each score – for each sensor and each iteration of each relevant question – can take one of seven possible values. When the number of possible scores is limited to two items the distribution is referred to as binomial, a special case of the multinomial. A distribution is referred to as *multinomial* when each score is taken from more than two possible values. Table 1 shows the number of scores for each of the different testing conditions.

The maximum possible score will be [number of scores (n) * max score (+3)]. Similarly, the minimum possible score will be [number of scores * the min score (-3)]. The range of possible total scores is [2 * max + 1] because the score zero (0) exists in between the + and – scores. Table 2 shows the range of possible scores for each of the testing configurations. In practice, polygraph scores are often loa-

Table 1. Number of scores for different testing conditions

	4 RQs	3 RQs	2RQs	RQ subtotals
With vasomotor	80	60	40	20
Without vasomotor	60	45	30	15

ded at statistically significant levels after three iterations of the question sequence. For this reason, field polygraph tests often consist of a minimum of three iterations of the test question sequence. When fewer than five iterations are used, it is simply as if the optional fourth and fifth iteration have resulted in no non-zero scores. In other words, the fourth and fifth charts contribute no information to the test result when they are not completed.

Table 2. Range of possible scores for different testing conditions

	4 RQs	3 RQs	2RQs	RQ subtotals
With vasomotor	481 (-240 to +240)	361 (-180 to +180)	241 (-120 to +120)	121 (-60 to +60)
Without vasomotor	361 (-180 to +180)	271 (-135 to + 135)	181 (-90 to +90)	91 (-45 to +45)

For each test configuration, there is only one way to achieve the maximum or minimum possible score – all score must be either the max (+3) or min (-3) score withing the Likert scale. While there is only one way to achieve the maximum or minimum possible total score, there are numerous ways to achieve a total score near zero. For each test configuration there is a mathematically finite combination of sensor scores, and a finite number of ways to achieve each possible total score within the range of possible total scores. One way to study the distribution of possible scores is to manually enumerate each possible combination of scores.

When the number of combinations is large it will be much more efficient to calculate the number of ways to achieve each possible score using combinatoric formulae. For comparison question polygraph tests consisting of five iterations of four relevant questions with four recording sensors the number of possible combinations of seven-position scores will be 4.053622e+67[1]. Unfortunately, when the number of possible combinations is very large these calculations can become unwieldy even for powerful computing platforms. In these situations, Monte Carlo methods – statistical methods developed to calculate large and difficult multinomial problems during the Manhattan project – can be a more efficient way of calculating a distribution of values. [Refer to Eckhardt (1987), Metropolis, (1987), and Metropolis & Ulam (1949) for more information on Monte Carlo methods.] Table 3 shows the number of possible combinations for different comparison question test configurations with seven position Likert scores.

Table 3. Number of possible combinations of 7 position scores.

	4 RQs	3 RQs	2RQs	RQ subtotals
With vasomotor	4.053622e+67	35.080219e+50	6.366806e+33	7.979227e+16
Without vasomotor	35.080219e+50	1.070069e+38	2.253934e+25	4.747562e+12

In field practice the entire range of seven-position scores is rarely, if ever, used for some of the sensors. By convention, respiration scores are commonly constrained to a range of three possible scores [-1, 0, +1]. The entire range of seven position values scores may be used for EDA scores. Cardio data are commonly less productive than electrodermal and more productive than respiration data. Cardio data often produce more numerical information than

[1]For readers who don't work with scientific notation for large numbers, this number can be thought of intuitively as the number 4053622 followed by an additional 61 zeros. As a point of trivia, a number with 60 zeros is named *novemedecillion*.

respiration scores yet less than EDA data, and rarely receive the maximum or minimum possible value from the seven-position Likert scale. Vasomotor scores are commonly assigned a numerical score using only the middle three values of the seven position scale.

A common characteristic of multivariate analysis – those that combine different sources of information into a statistical model or statistical classifier – is that different sources of data do not contribute equally to an optimal model. [Refer to Nelson (2019) for a survey of structural weighting coefficients for comparison question polygraph test scores.] For this reason, multinomial distributions for this project were calculated using a weighting scheme that was devised by constraining the range of the Likert scale for each of the recording sensors. Table 4 shows the constrained Likert scales for each of the polygraph recording sensors.

Table 4. Number of possible combinations of seven-position scores.

	Respiration scores	EDA scores	Cardio scores	Vasomotor scores
Likert scale range	[-1 0 +1]	[-3 -2 -1 0 +1 +2 +3]	[-2 -1 0 +1 +2]	[-1 0 +1]

These scales produced the following weighting coefficients for the sensor data: respiration = .14, EDA = .43, cardio = .28, and vasomotor = .14. When vasomotor data are omitted, the weighting coefficients for Table 4 are as follows: respiration = .17, EDA = .5, cardio = .33. Evaluation of the weighting coefficients reported in Nelson (2019) produced the following weighting coefficients when vasomotor data was not included: respiration data = .22, EDA = .49, and cardio data = .28. Using only those studies that include the vasomotor sensor, aggregated weighting coefficients were the following: respiration = .21, EDA = .41, cardio = .18, and vasomotor = .22.

Evaluation of the summed numerical scores from Kircher and Raskin (1988) produced the following weighting coefficients: respiration = .08, EDA = .59, cardio = .16, vasomotor = .16. A more recent study by Honts and Reavy (2015) included seven position manual scores that resulted in the following weighting coefficients: respiration = .13, EDA = .39, cardio = .26, vasomotor = .22. With consideration for the observed variation in reported weighting coefficients in previous studies, the EDA data is found to consistently contribute more information to effective classification than the other sensors. The Likert scales shown in Table 4 are thought to be a reasonable and convenient integer-based approximation to an optimal structural model.

Constraining the multinomial scales for each of the sensor has the effect of reducing the number of possible multinomial combinations. However, the resulting combinatoric distribution is still too large to compute an exact multinomial calculation, and still are more suited for Monte Carlo methods.

Design of the Monte Carlo model

The Monte Carlo distributions were calculated by sampling from the Likert scale scores, for each sensor, under the null-hypothesis to the analytic theory. Scientific experiments are often designed to test a null hypothesis. An alternative hypothesis can be accepted when the probability of obtaining the observed data under the null hypothesis is sufficiently low. In this method, a hypothesis is supported by the data only when the null hypothesis is rejected. A large amount of scientific knowledge has been acquired by testing of hypotheses in this manner. When applied to the polygraph context, this method asks the following question of the test: *are the test loaded, at a statistically significant level, for deception or truth-telling?* Expressed under the null hypothesis the question is asked this way: *are the aggregated test scores similar to zero?* When a test does not work – when the data are not loaded for deception or truth-telling – numerical scores do not differ significantly from zero and a test result is inconclusive.

Monte Carlo distributions, under the null hypothesis, are constructed by sampling, with replacement, from the Likert scale values. When sampling under the null hypothesis each of the values within the Likert scales is selected with uniform or equal probability weighting. Also, whereas under the theory of the test scores are systematically loaded and are therefore covariant, under that null hypothesis each score is unrelated to other scores and therefore independent. The number of sample scores for each case is determined by the number of relevant questions and the number of recording sensors, as shown in Table 1. Three million sample cases were included in the Monte Carlo space.

After sampling the sensor scores, sensor scores for each case in the Monte Carlo space were then aggregated to total scores. Results of event-specific (single-issue) diagnostic exams can be classified as deceptive or truthful using both grand total and subtotal scores, while results of multiple issue screening polygraphs are commonly classified using the subtotal scores for the individual relevant questions. [See to Nelson (2018b) for a description of the procedural rules used to classify polygraph test results.]

Total scores in the Monte Carlo space were then counted for their frequency of occurrence, and these sums were divided by the number of cases in the Monte Carlo space. The resulting probability values indicate the frequency of occurrence of each possible numerical total within the multinomial combinatoric distribution of scores if it were possible to calculate the combinatoric formula. As stated earlier, Monte Carlo methods are an effective way of calculating models for which the combinations of interactions is very complex, or the numbers are very large, to a degree that they remain unwieldy even for today's computing platforms.

Multinomial distributions were calculated for tests with four recording sensors, including a single score for abdominal and thoracic respiration, electrodermal activity, cardiovascular activity, and the optional vasomotor activity. When the optional vasomotor sensor is not used it is simply as if there are no non-zero vasomotor scores. In other words, the optional vasomotor sensor contributes nothing to the test scores when it is not used. Multinomial tables, shown in Appendices A-H, were calculated for single issue and multiple issue polygraph examinations consisting of up to

Table 5. Segment of a multinomial table for single issue polygraphs with three RQs.

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-8	0.029	0.238	0.224	3.5	2.0
-7	0.031	0.269	0.253	2.9	1.8
-6	0.032	0.301	0.285	2.5	1.5
-5	0.034	0.334	0.318	2.1	1.3
-4	0.035	0.370	0.352	1.8	1.1
-3	0.036	0.406	0.388	1.6	1.0
-2	0.037	0.443	0.425	1.4	0.9
-1	0.038	0.481	0.462	1.2	0.7
0	0.038	0.519	0.500	1.0	0.6
1	0.038	0.557	0.462	1.2	0.7
2	0.037	0.594	0.425	1.4	0.9
3	0.036	0.630	0.388	1.6	1.0
4	0.035	0.666	0.352	1.8	1.1
5	0.034	0.699	0.318	2.1	1.3
6	0.032	0.731	0.285	2.5	1.5
7	0.031	0.762	0.253	2.9	1.8
8	0.029	0.790	0.224	3.5	2.0

five iterations of two, three, or four relevant questions, with and without the optional vasomotor sensor. All computation was done in the R Language for Statistical Computing (R Core Team, 2021). Appendix I includes the code used to calculate the multinomial tables.

Table in the appendices cover the entire spectrum of validated comparison question test formats described by American Polygraph Association (2011). Table 5 shows a portion of a multinomial distribution for a single-issue polygraph test with three relevant questions.

One of the important practical uses for probability tables is as a likelihood function for Bayesian analysis. [Refer to Nelson (2018c) for more information on Bayesian analysis and its application to polygraph testing.] Columns of greatest practical interest are shaded lightly in yellow in Table 5. Each column is calculated from the column immediately to the left. The rightmost column, titled *oddsLL05*, indicates the lower limit of the credible interval for the posterior odds of deception or truth-telling with one-tailed alpha set at .05. These values were calculated using the Clopper-Pearson method. [Refer to Nelson (2018d) for information on the Clopper-Pearson interval.] Each value in the rightmost column is associated with a total numerical score in the leftmost *score* column and tells us that when accounting for expected variability in test data, repetition of an examination will produce a similar classification with 95% likelihood. The column labeled *odds* shows the posterior conditional odds of truth or deception under an equal prior.

Working from the left side of the table, the leftmost column shows the range of total scores (though truncated in this example). The column labeled pmf shows an estimate of the probability mass function for each possible total score. The pmf is the expected proportion of all possible test results, under the null hypothesis, that are expected to be equal to the scores in the leftmost score column. The column labeled *cdf* shows the cumulative distribution function and is calculated as a running sum of each successive row in the pmf column. The pmf can be thought of as the probability, under the null hypothesis, of observing a test score that is equal to or lower than each possible scores in the score column. The *cdfContCor* column is a continuity correction to the cdf. Using a *continuity correction* in this way assures that the actual probability value, for deception or truth-telling, always exceeds the calculated statistical value. The *odds* column is calculated from the *cdfContCor* column, and provides a more practical and intuitive discussion, compared to decimal probabilities, of the probability of deception or truth-telling.

Using the Multinomial Reference Model

Multinomial reference tables can be used to calculate the posterior odds associated with each possible total score. This is done by locating an observed test score in the left column, labeled *score*, and then finding the value in the corresponding row using the column labeled *odds*. For example, using the information in Table 5, a total numerical score of eight (8) can located in the score column, and corresponds to a posterior odds of 3.5 to 1 for truth-telling, assuming an equal prior, as shown in the *odds* column.

A second use for the multinomial reference tables shown in the appendices is to calculate, prior to any testing, the numerical cutscores that are required to achieve desired posterior effect sizes. For example, a cutscore can be selected so that test results equal or exceed a desired posterior odds. Table 6 shows the multinomial cutscores that correspond to a posterior odd of 2 to 1 or greater. using the *two-stage-rule* (TSR: Senter, 2003; Senter & Dollins, 2003; Krapohl, 2005; Krapohl & Cushman, 2006) for single issue exams, and the subtotal-score-rule (SSR: Department of Defense, 2006, National Center for Credibility Assessment, 2017; Senter Waller & Krapohl, 2008) for multiple is-sue exams. Table 7 shows the cutscores using seven-position multinomial tables calculated without the vasomotor sensor.

Another way to calculate cutscores – and potentially reduce inconclusive results while maintaining desired classification effect sizes – is to work with the expected variation in test sample data. This requires calculation of an expected interval of variability, typically using a=.05, that give the lower limits and upper limits of the space in which we expect an unknown value to exist. Table 8 shows the multinomial cutscores using alpha=.05 for the lower

Table 6. Seven-position multinomial cutscores to achieve posterior odds >=2 to 1 with vasomotor data.

	4 RQs	3 RQs	2RQs
Single issue exams: TSR*	+6 / -6 (-10)	-+5 / -5 (-8)	+4 / -4 (-5)
Multiple issue exams; SSR	-3 / (+1)	-3 / (+1)	-3 / (+2)

(values in parens indicate correction for statistical multiplicity when using the subtotal scores)

Table 7. Seven-position multinomial cutscores to achieve posterior odds >2 to 1 without vasomotor data.

	4 RQs	3 RQs	2RQs
Single issue exams: TSR	+5 / -5 (-9)	+5 / -5 (-7)	+4 / -4 (-5)
Multiple issue exams: SSR	-3 (+1)	-3 / (+1)	-3 / (+2)

(values in parens indicate correction for statistical multiplicity when using the subtotal scores)

limit of the Clopper-Pearson interval for the posterior odds of deception and truth-telling. Scores that equal or exceed these cutscores can be said to be indicative or deception or truth-telling at a 1-alpha * 100% = 95% level of confidence (referred to as a credible interval in Bayesian analysis) that other data, collected and analyzed in the same way, would lead to a similar classification. This can be thought of as the 95% likelihood that other data would be similarly indicative of deception or truth-telling if acquired and analyzed the same way. Table 9 shows the cutscores for the lower limit of the Clopper Pearson interval without the vasomotor sensor data.

Table 8. Seven-position multinomial cutscores to assure the 95% CI exceeds the equal prior with vasomotor data.

	4 RQs	3 RQs	2RQs
Single issue exams: TSR*	+3 / -3 (-11)	-+3 / -3 (-9)	+3 / -3 (-7)
Multiple issue exams; SSR	-4 / (+1)	-4 / (+2)	-4 / (+2)

(values in parens indicate correction for statistical multiplicity when using the subtotal scores)

Table 9. Seven-position multinomial cutscores to assure the 95% CI exceeds the equal prior without vasomotor data.

	4 RQs	3 RQs	2RQs
Single issue exams: TSR	+3 / -3 (-13)	+4 / -4 (-10)	+4 / -4 (-7)
Multiple issue exams: SSR	-4 / (+1)	-4 / (+2)	-4 / (+2)

(values in parens indicate correction for statistical multiplicity when using the subtotal scores)

Numerical scores that do not meet these cutscores do not provide a 95% level of confidence that they will be replicated by similar testing and analysis procedures. In other words, this method of cutscore selection answers the question: *if not certain, what cutscores provide us with a 95% level of confidence that other test data will be similarly indicative of deception or truth-telling?* Or: *how sure can one be that the observed test result is indicative of deception or truth-telling?*

Discussion

Multinomial distributions are an expression of the analytic theory of the polygraph test, and were previously published for three-position scoring methods, including the Federal three-position scores (Nelson, 2018a) and for the empirical scoring system (ESS, Nelson, 2017). Empirical reference distributions have also been described (Nelson & Handler, 2015). However, no previous publications were found that described a theoretical (mathematical and statistical) justification for seven position scores or suggested cutscores.

Theoretical distributions are important part of any scientific activity because they are calculated using only facts and information that are subject to mathematical and logical proof. When our knowledge about a theoretical distribution is strong and sufficient, we can use that knowledge to make statistical inferences about an observed individual case. For example, a six-sided die is multinomial in that when cast it will display one of six possible values. For an unbiased die all of the possible scores will occur with frequency when data are collected for numerous trials. Of course, some variation will always be observed However, if we observe that some numbers occur with a frequency that exceeds normally expected variation, it supports a conclusion about that particular die.

Throughout the 20th and 21st centuries, there has been a consistent trend in social sciences to conform, to the extent possible, to the procedures and expectations of other scientific disciplines, including the formation of testable hypothesis, rejection of ideas that cannot be replicated or which cannot be supported with reproducible analysis, and the mathematical quantification of probabilities associated with conclusions made under inherent uncertainty (Baker, 2016; Meehl, 1954; Peng, 2015; Popper, 1959; 1972). These expectations extend to both scientific research and also to scientific testing of individuals. Examples of this are abundant and include the conception of the Likert scale itself, for which its intended purpose is to permit the numerical statistical analysis of subjective information. Other examples include the scientific concepts of a null hypothesis and alternative hypothesis, and the requirement that scientific ideas are testable and falsifiable.

A limitation of this study is that it does not include the study of outcome effect sizes. This is somewhat mitigated by the fact that numerous studies have already demonstrated useful effect sizes for CQT procedures and is further mitigated by the similarity of multinomial seven-position cutscores with traditional seven-position cutscore that were developed largely through heuristic experience. This project is limited to the application of multinomial theory to the seven position Likert scale under the null hypothesis. Multinomial probabilities are an expression of the likelihood of obtaining a score equal to or more extreme than an observed test score if the test data are not loaded systematically for deception or truth-telling. Posterior odds shown in the appendices can be thought of as the likelihood of deception or truth-telling after updating an equal prior with the multinomial likelihood value for an observed test score.

A more important limitation surrounds the seven position scoring method in general. Although the multinomial distribution under the null hypothesis may be objective, there is no theoretical justification (subject to mathematical and logical proof) as to the boundaries between the different seven position scale scores. Attempts to find empirical or observational basis for scoring boundaries does not satisfy the scientific requirement for theoretical justification. Also, attempts to apply mathematical ratio constraints or rules to the seven position scale values are without theoretical justification and are therefore arbitrary. Whereas the seven position Likert scale is inherently limited due either subjective or arbitrary boundaries between scale items, three-position scoring methods, based on the rank ordinal comparison of relevant and comparison response pairs, are objectively multinomial and are supported by an inherently stronger theoretical and scientific foundation.

Although multinomial distributions were previously introduced to polygraph data analysis for three-position scoring methods, this project advances the scientific foundation of the comparison question test through the calculation of multinomial references distributions for the traditional seven-position scoring method for CQT data. No previous effort was found to attempt to calculate a theoretical distribution for this method. Future research may want to investigate the degree to which the use of multinomial cutscores and multinomial may be useful to reducing the occurrence of inconclusive results or to improve effect sizes of interest to field polygraph examiners, polygraph programs and referring professionals.

References

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph, 2011, 40*(4), 196-305. [Electronic version] Retrieved February 12, 2022, from http://www .polygraph.org.

Backster, C. (1963). *Standardized Polygraph Notepack and Technique Guide: Backster Zone Comparison Technique*. Cleve Backster: New York.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604), 452-454.

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

Department of Defense (2006). *Federal psychophysiological detection of deception examiner handbook*. Reprinted in *Polygraph*, 40 (1), 2-66.

Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method. *Los Alamos Science (15)*, 131–137.

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

Honts, C. R. & Reavy, R. (2015). The comparison question polygraph test: A contrast of methods and scoring. *Physiology and Behavior, 143*, 15-26.

Kircher, J. C. & Raskin, D.C. (1988). Human vs. computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology,* 72(2), 291-302.

Kubis, J. F. (1962). *Studies in Lie Detection: Computer Feasibility Considerations*. RADC-TR 62-205, Contract AF 30(602)-2270. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-55.

Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science (1987 Special Issue dedicated to Stanislaw Ulam),* 125–130.

Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. https://doi.org/10.1037/11281-000

Metropolis, N., Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association,* 44(247), 335–341.

National Center for Credibility Assessment (2017). *PDD 503-Analysis II. Test Data Analysis: Numerical Evaluation Scoring System Pamphlet.* Available from the author.

Nelson, R. (2015). Scientific basis for polygraph testing. *Polygraph,* 41(1), 21-61.

Nelson, R. (2016). Scientific (analytic) theory of polygraph testing. *APA Magazine,* 49(5), 69-82.

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

Nelson, R. (2018a). Multinomial reference distributions for three-position scores of comparison question polygraph examinations. *Polygraph & Forensic Credibility Assessment,* 47 (2), 158-175.

Nelson, R. (2018b). Practical polygraph: A survey and description of decision rules. *APA Magazine 2018,* 51(2), 127-133.

Nelson, R. (2018c). Practical polygraph – a tutorial (with graphics) on posterior results and credible intervals using the ESS-M Bayesian classifier. *APA Magazine,* 51(4), 66-87.

Nelson, R. (2018d). Five-minute science lesson: Clopper-Pearson credibility intervals for Bayesian analysis of multinomial polygraph scores. *APA Magazine,* 51(3), 61-70.

Nelson, R. (2019). Literature survey of structural weighting of polygraph signals: Why double the EDA? *Polygraph & Forensic Credibility Assessment,* 48(2) 105-112.

Nelson, R. & Handler, M. (2015). Statistical reference distributions for comparison question polygraphs. *Polygraph,* 44(1), 91-114.

Nelson, R. & Krapohl, D. (2017). Practical polygraph: A recommendation for combining the upper and lower respiration data for a single respiration score. *APA Magazine,* 50 (6), 31-41.

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

Peng. R. (2011). Reproducible research in computational science. *Science, 334(6060)*, 1226-1227.

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance,* 12(3), 30-32.

Popper, Karl (1959). *The Logic of Scientific Discovery*. New York, NY: Basic Books

Popper, K. R. (1972). *Conjectures and refutations: The growth of scientific knowledge (4th ed.).* London: Routledge Kegan Paul.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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.

Senter, S., Waller, J. & Krapohl, D. (2008). Air Force Modified General Question Test Validation Study. *Polygraph,* 37(3), 174-184.

Appendix A: Multinomial Reference Table – 4 RQs without PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-27	0.002	0.011	0.010	104	14
-26	0.003	0.013	0.012	82	13
-25	0.003	0.017	0.015	65	13
-24	0.004	0.021	0.019	52	11
-23	0.005	0.025	0.023	42	11
-22	0.006	0.031	0.028	34	9.4
-21	0.007	0.038	0.034	28	8.7
-20	0.008	0.046	0.042	23	7.9
-19	0.009	0.055	0.050	19	7.0
-18	0.010	0.065	0.060	16	6.3
-17	0.012	0.076	0.071	13	5.6
-16	0.013	0.090	0.083	11	5.0
-15	0.015	0.105	0.097	9.3	4.4
-14	0.017	0.122	0.113	7.8	3.9
-13	0.018	0.140	0.131	6.7	3.5
-12	0.020	0.160	0.150	5.7	3.0
-11	0.022	0.182	0.171	4.8	2.7
-10	0.024	0.206	0.194	4.2	2.4
-9	0.026	0.231	0.219	3.6	2.1
-8	0.027	0.259	0.245	3.1	1.8
-7	0.029	0.287	0.273	2.7	1.6
-6	0.030	0.317	0.302	2.3	1.4
-5	0.031	0.349	0.333	2.0	1.2
-4	0.032	0.381	0.365	1.7	1.1
-3	0.033	0.414	0.398	1.5	1.0
-2	0.034	0.449	0.432	1.3	0.8
-1	0.034	0.483	0.466	1.1	0.7
0	0.034	0.517	0.500	1.0	0.6
1	0.034	0.551	0.466	1.1	0.7
2	0.034	0.586	0.432	1.3	0.8
3	0.033	0.619	0.398	1.5	1.0
4	0.032	0.651	0.365	1.7	1.1
5	0.031	0.683	0.333	2.0	1.2
6	0.030	0.713	0.302	2.3	1.4
7	0.029	0.741	0.273	2.7	1.6
8	0.027	0.769	0.245	3.1	1.8
9	0.026	0.794	0.219	3.6	2.1
10	0.024	0.818	0.194	4.2	2.4
11	0.022	0.840	0.171	4.8	2.7
12	0.020	0.860	0.150	5.7	3.0
13	0.018	0.878	0.131	6.7	3.5
14	0.017	0.895	0.113	7.8	3.9
15	0.015	0.910	0.097	9.3	4.4
16	0.013	0.924	0.083	11	5.0
17	0.012	0.935	0.071	13	5.6
18	0.010	0.945	0.060	16	6.3
19	0.009	0.954	0.050	19	7.0
20	0.008	0.962	0.042	23	7.9
21	0.007	0.969	0.034	28	8.7
22	0.006	0.975	0.028	34	9.4
23	0.005	0.979	0.023	42	11
24	0.004	0.983	0.019	53	11
25	0.003	0.987	0.015	65	13
26	0.003	0.989	0.012	82	13
27	0.002	0.992	0.010	104	14

Appendix B: Multinomial Reference Table – 3 RQs without PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-24	0.002	0.009	0.008	126.2	12.2
-23	0.003	0.012	0.010	95.4	11.2
-22	0.004	0.015	0.014	72.7	10.2
-21	0.004	0.020	0.018	55.8	10.0
-20	0.005	0.025	0.023	43.3	8.8
-19	0.007	0.032	0.029	34.0	8.2
-18	0.008	0.040	0.036	26.9	7.2
-17	0.009	0.049	0.045	21.5	6.5
-16	0.011	0.060	0.055	17.2	5.8
-15	0.013	0.073	0.067	14.0	5.1
-14	0.015	0.088	0.081	11.4	4.5
-13	0.017	0.106	0.097	9.3	4.0
-12	0.020	0.125	0.116	7.7	3.4
-11	0.022	0.147	0.136	6.3	3.0
-10	0.024	0.172	0.159	5.3	2.6
-9	0.027	0.198	0.185	4.4	2.3
-8	0.029	0.227	0.213	3.7	2.0
-7	0.031	0.258	0.243	3.1	1.7
-6	0.033	0.292	0.275	2.6	1.5
-5	0.035	0.327	0.309	2.2	1.3
-4	0.037	0.364	0.345	1.9	1.1
-3	0.038	0.402	0.383	1.6	0.9
-2	0.039	0.441	0.421	1.4	0.8
-1	0.039	0.480	0.461	1.2	0.7
0	0.040	0.520	0.500	1.0	0.6
1	0.039	0.559	0.461	1.2	0.7
2	0.039	0.598	0.421	1.4	0.8
3	0.038	0.636	0.383	1.6	0.9
4	0.037	0.673	0.345	1.9	1.1
5	0.035	0.708	0.309	2.2	1.3
6	0.033	0.742	0.275	2.6	1.5
7	0.031	0.773	0.243	3.1	1.7
8	0.029	0.802	0.213	3.7	2.0
9	0.027	0.828	0.185	4.4	2.3
10	0.024	0.853	0.159	5.3	2.6
11	0.022	0.875	0.136	6.3	3.0
12	0.020	0.894	0.116	7.7	3.4
13	0.017	0.912	0.097	9.3	4.0
14	0.015	0.927	0.081	11.4	4.5
15	0.013	0.940	0.067	14.0	5.1
16	0.011	0.951	0.055	17.2	5.8
17	0.009	0.960	0.045	21.5	6.5
18	0.008	0.968	0.036	26.9	7.2
19	0.007	0.975	0.029	34.0	8.2
20	0.005	0.980	0.023	43.3	8.8
21	0.004	0.985	0.018	55.8	10.0
22	0.004	0.988	0.014	72.7	10.2
23	0.003	0.991	0.010	95.4	11.2
24	0.002	0.993	0.008	126.2	12.2

Appendix C: Multinomial Reference Table – 2 RQs without PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-19	0.003	0.011	0.010	101.3	8.3
-18	0.004	0.016	0.014	72.8	7.6
-17	0.006	0.021	0.019	53.0	6.9
-16	0.007	0.029	0.025	39.0	6.1
-15	0.009	0.038	0.033	29.1	5.7
-14	0.011	0.049	0.043	22.0	5.2
-13	0.014	0.063	0.056	16.8	4.4
-12	0.017	0.080	0.072	13.0	3.9
-11	0.020	0.100	0.090	10.1	3.4
-10	0.023	0.123	0.112	8.0	3.0
-9	0.027	0.150	0.137	6.3	2.5
-8	0.030	0.181	0.165	5.0	2.2
-7	0.034	0.215	0.198	4.1	1.8
-6	0.037	0.252	0.233	3.3	1.5
-5	0.041	0.292	0.272	2.7	1.3
-4	0.043	0.335	0.314	2.2	1.1
-3	0.045	0.381	0.358	1.8	0.9
-2	0.047	0.428	0.404	1.5	0.8
-1	0.048	0.476	0.452	1.2	0.6
0	0.048	0.524	0.500	1.0	0.5
1	0.048	0.572	0.452	1.2	0.6
2	0.047	0.619	0.404	1.5	0.8
3	0.045	0.665	0.358	1.8	0.9
4	0.043	0.708	0.314	2.2	1.1
5	0.041	0.748	0.272	2.7	1.3
6	0.037	0.785	0.233	3.3	1.5
7	0.034	0.819	0.198	4.1	1.8
8	0.030	0.850	0.165	5.0	2.2
9	0.027	0.877	0.137	6.3	2.5
10	0.023	0.900	0.112	8.0	3.0
11	0.020	0.920	0.090	10.1	3.4
12	0.017	0.937	0.072	13.0	3.9
13	0.014	0.951	0.056	16.8	4.4
14	0.011	0.962	0.043	22.0	5.2
15	0.009	0.971	0.033	29.1	5.7
16	0.007	0.979	0.025	39.0	6.1
17	0.006	0.984	0.019	53.0	6.9
18	0.004	0.989	0.014	72.8	7.6
19	0.003	0.992	0.010	101.3	8.3

Appendix D: Multinomial Reference Table – Question Subtotals without PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05	odds2RQ	odds3RQ	odds4RQ	odds2LL05	odds3LL05	odds4LL05
-14	0.004	0.009	0.007	141	4.1	11.9	5.2	3.4	2.4	1.5	1.1
-13	0.006	0.015	0.012	85	3.8	9.2	4.4	3.0	2.1	1.4	1.0
-12	0.008	0.023	0.019	53	3.9	7.3	3.8	2.7	1.9	1.2	0.9
-11	0.011	0.034	0.028	34	3.5	5.9	3.2	2.4	1.6	1.1	0.9
-10	0.016	0.050	0.042	23	3.0	4.8	2.8	2.2	1.4	1.0	0.8
-9	0.021	0.071	0.060	16	2.7	3.9	2.5	2.0	1.3	0.9	0.7
-8	0.027	0.098	0.085	11	2.4	3.3	2.2	1.8	1.1	0.8	0.7
-7	0.034	0.132	0.115	7.7	1.9	2.8	2.0	1.7	1.0	0.7	0.6
-6	0.041	0.173	0.152	5.6	1.6	2.4	1.8	1.5	0.8	0.7	0.6
-5	0.048	0.220	0.197	4.1	1.3	2.0	1.6	1.4	0.7	0.6	0.5
-4	0.054	0.275	0.248	3.0	1.0	1.7	1.4	1.3	0.6	0.5	0.5
-3	0.060	0.335	0.305	2.3	0.8	1.5	1.3	1.2	0.6	0.5	0.5
-2	0.064	0.399	0.367	1.7	0.6	1.3	1.2	1.1	0.5	0.4	0.4
-1	0.067	0.466	0.432	1.3	0.5	1.1	1.1	1.1	0.4	0.4	0.4
0	0.068	0.534	0.500	1.0	0.4	1.0	1.0	1.0	0.4	0.4	0.4
1	0.067	0.601	0.432	1.3	0.5	1.7	2.3	3.0	0.6	0.8	1.0
2	0.064	0.665	0.367	1.7	0.6	3.0	5.2	8.9	1.0	1.5	2.1
3	0.060	0.725	0.305	2.3	0.8	5.2	12	27	1.5	2.4	3.2
4	0.054	0.780	0.248	3.0	1.0	9.2	28	85	2.1	3.3	3.8
5	0.048	0.827	0.197	4.1	1.3	16.7	68	280	2.8	3.7	4.3
6	0.041	0.868	0.152	5.6	1.6	31.0	173	>500	3.4	4.1	4.4
7	0.034	0.902	0.115	7.7	1.9	59.2	456	>500	3.6	4.4	4.5
8	0.027	0.929	0.085	11	2.4	117.2	>500	>500	4.0	4.5	4.5
9	0.021	0.950	0.060	16	2.7	241.5	>500	>500	4.2	4.5	4.5
10	0.016	0.966	0.042	23	3.0	>500	>500	>500	4.4	4.5	4.5
11	0.011	0.977	0.028	34	3.5	>500	>500	>500	4.5	4.5	4.5
12	0.008	0.985	0.019	53	3.9	>500	>500	>500	4.5	4.5	4.5
13	0.006	0.991	0.012	85	3.8	>500	>500	>500	4.5	4.5	4.5
14	0.004	0.995	0.007	141	4.1	>500	>500	>500	4.5	4.5	4.5

Appendix E: Multinomial Reference Table – 4 RQs with PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-29	0.002	0.009	0.008	122	18
-28	0.002	0.011	0.010	97	18
-27	0.003	0.014	0.013	78	16
-26	0.003	0.017	0.016	62	15
-25	0.004	0.021	0.019	51	14
-24	0.004	0.026	0.024	41	13
-23	0.006	0.031	0.029	34	11
-22	0.006	0.038	0.034	28	10
-21	0.007	0.045	0.041	23	9.2
-20	0.008	0.053	0.049	19	8.1
-19	0.010	0.063	0.058	16	7.2
-18	0.011	0.074	0.069	14	6.4
-17	0.012	0.086	0.080	12	5.7
-16	0.014	0.100	0.093	9.7	5.1
-15	0.015	0.115	0.108	8.3	4.5
-14	0.017	0.133	0.124	7.1	4.0
-13	0.018	0.151	0.142	6.1	3.5
-12	0.020	0.171	0.161	5.2	3.1
-11	0.022	0.193	0.182	4.5	2.7
-10	0.024	0.217	0.205	3.9	2.4
-9	0.025	0.242	0.229	3.4	2.1
-8	0.027	0.268	0.255	2.9	1.9
-7	0.028	0.296	0.282	2.5	1.7
-6	0.029	0.325	0.311	2.2	1.5
-5	0.030	0.355	0.340	1.9	1.3
-4	0.031	0.386	0.371	1.7	1.1
-3	0.032	0.418	0.402	1.5	1.0
-2	0.033	0.451	0.435	1.3	0.9
-1	0.033	0.484	0.467	1.1	0.8
0	0.033	0.516	0.500	1.0	0.7
1	0.033	0.549	0.467	1.1	0.8
2	0.033	0.582	0.435	1.3	0.9
3	0.032	0.614	0.402	1.5	1.0
4	0.031	0.645	0.371	1.7	1.1
5	0.030	0.675	0.340	1.9	1.3
6	0.029	0.704	0.311	2.2	1.5
7	0.028	0.732	0.282	2.5	1.7
8	0.027	0.758	0.255	2.9	1.9
9	0.025	0.783	0.229	3.4	2.1
10	0.024	0.807	0.205	3.9	2.4
11	0.022	0.829	0.182	4.5	2.7
12	0.020	0.849	0.161	5.2	3.1
13	0.018	0.867	0.142	6.1	3.5
14	0.017	0.885	0.124	7.1	4.0
15	0.015	0.900	0.108	8.3	4.5
16	0.014	0.914	0.093	9.7	5.1
17	0.012	0.926	0.080	12	5.7
18	0.011	0.937	0.069	14	6.4
19	0.010	0.947	0.058	16	7.2
20	0.008	0.955	0.049	19	8.1
21	0.007	0.962	0.041	23	9.2
22	0.006	0.969	0.034	28	10
23	0.006	0.974	0.029	34	11
24	0.004	0.979	0.024	41	13
25	0.004	0.983	0.019	51	14
26	0.003	0.986	0.016	62	15
27	0.003	0.989	0.013	78	16
28	0.002	0.991	0.010	97	18
29	0.002	0.993	0.008	122	18

Appendix F: Multinomial Reference Table – 3 RQs with PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-25	0.002	0.009	0.008	118	15
-24	0.003	0.012	0.011	91	13
-23	0.003	0.016	0.014	71	13
-22	0.004	0.020	0.018	55	12
-21	0.005	0.025	0.023	44	11
-20	0.006	0.031	0.028	34	9.4
-19	0.007	0.039	0.035	28	8.6
-18	0.009	0.047	0.043	22	7.7
-17	0.010	0.058	0.053	18	6.8
-16	0.012	0.070	0.064	15	6.1
-15	0.014	0.083	0.077	12	5.3
-14	0.016	0.099	0.091	10	4.6
-13	0.018	0.117	0.108	8.3	4.1
-12	0.020	0.137	0.127	6.9	3.5
-11	0.022	0.159	0.148	5.8	3.1
-10	0.024	0.183	0.171	4.8	2.7
-9	0.026	0.210	0.196	4.1	2.3
-8	0.029	0.238	0.224	3.5	2.0
-7	0.031	0.269	0.253	2.9	1.8
-6	0.032	0.301	0.285	2.5	1.5
-5	0.034	0.334	0.318	2.1	1.3
-4	0.035	0.370	0.352	1.8	1.1
-3	0.036	0.406	0.388	1.6	1.0
-2	0.037	0.443	0.425	1.4	0.9
-1	0.038	0.481	0.462	1.2	0.7
0	0.038	0.519	0.500	1.0	0.6
1	0.038	0.557	0.462	1.2	0.7
2	0.037	0.594	0.425	1.4	0.9
3	0.036	0.630	0.388	1.6	1.0
4	0.035	0.666	0.352	1.8	1.1
5	0.034	0.699	0.318	2.1	1.3
6	0.032	0.731	0.285	2.5	1.5
7	0.031	0.762	0.253	2.9	1.8
8	0.029	0.790	0.224	3.5	2.0
9	0.026	0.817	0.196	4.1	2.3
10	0.024	0.841	0.171	4.8	2.7
11	0.022	0.863	0.148	5.8	3.1
12	0.020	0.883	0.127	6.9	3.5
13	0.018	0.901	0.108	8.3	4.1
14	0.016	0.917	0.091	10	4.6
15	0.014	0.930	0.077	12	5.3
16	0.012	0.942	0.064	15	6.1
17	0.010	0.953	0.053	18	6.8
18	0.009	0.961	0.043	22	7.7
19	0.007	0.969	0.035	28	8.6
20	0.006	0.975	0.028	35	9.4
21	0.005	0.980	0.023	44	11
22	0.004	0.984	0.018	55	12
23	0.003	0.988	0.014	71	13
24	0.003	0.991	0.011	91	13
25	0.002	0.993	0.008	118	15

Appendix G: Multinomial Reference Table – 2 RQs with PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05
-20	0.003	0.011	0.010	104	11
-19	0.004	0.015	0.013	76	9.5
-18	0.005	0.020	0.018	56	8.5
-17	0.007	0.027	0.023	42	8.1
-16	0.008	0.035	0.031	31	7.0
-15	0.010	0.045	0.040	24	6.3
-14	0.012	0.057	0.051	19	5.6
-13	0.015	0.072	0.065	14	4.9
-12	0.018	0.090	0.081	11	4.2
-11	0.021	0.111	0.100	9.0	3.7
-10	0.024	0.134	0.123	7.2	3.1
-9	0.027	0.161	0.148	5.8	2.7
-8	0.030	0.191	0.176	4.7	2.3
-7	0.034	0.225	0.208	3.8	1.9
-6	0.037	0.261	0.243	3.1	1.6
-5	0.039	0.301	0.281	2.6	1.4
-4	0.041	0.342	0.321	2.1	1.2
-3	0.044	0.386	0.364	1.7	1.0
-2	0.045	0.431	0.408	1.4	0.8
-1	0.046	0.477	0.454	1.2	0.7
0	0.046	0.523	0.500	1.0	0.6
1	0.046	0.569	0.454	1.2	0.7
2	0.045	0.614	0.408	1.4	0.8
3	0.044	0.658	0.364	1.7	1.0
4	0.041	0.699	0.321	2.1	1.2
5	0.039	0.739	0.281	2.6	1.4
6	0.037	0.775	0.243	3.1	1.6
7	0.034	0.809	0.208	3.8	1.9
8	0.030	0.839	0.176	4.7	2.3
9	0.027	0.866	0.148	5.8	2.7
10	0.024	0.889	0.123	7.2	3.1
11	0.021	0.910	0.100	9.0	3.7
12	0.018	0.928	0.081	11	4.2
13	0.015	0.943	0.065	14	4.9
14	0.012	0.955	0.051	19	5.6
15	0.010	0.965	0.040	24	6.3
16	0.008	0.973	0.031	32	7.0
17	0.007	0.980	0.023	42	8.1
18	0.005	0.985	0.018	56	8.5
19	0.004	0.989	0.013	76	9.5
20	0.003	0.992	0.010	104	11

Appendix H: Multinomial Reference Table – Question Subtotals with PLE

Score	pmf	cdf	cdfContCor	odds	oddsLL05	odds2RQ	odds3RQ	odds4RQ	odds2LL05	odds3LL05	odds4LL05
-14	0.005	0.012	0.010	101	5.1	10	4.7	3.2	2.7	1.7	1.3
-13	0.007	0.019	0.015	64	5.3	8.0	4.0	2.8	2.3	1.5	1.1
-12	0.010	0.028	0.024	42	4.7	6.4	3.5	2.5	2.1	1.3	1.0
-11	0.013	0.041	0.035	28	4.0	5.3	3.0	2.3	1.8	1.2	1.0
-10	0.017	0.059	0.050	19	3.6	4.4	2.7	2.1	1.6	1.1	0.9
-9	0.022	0.081	0.070	13	3.1	3.6	2.4	1.9	1.4	1.0	0.8
-8	0.028	0.109	0.095	9.5	2.6	3.1	2.1	1.8	1.2	0.9	0.8
-7	0.034	0.143	0.126	6.9	2.2	2.6	1.9	1.6	1.1	0.8	0.7
-6	0.041	0.184	0.164	5.1	1.8	2.3	1.7	1.5	0.9	0.7	0.7
-5	0.047	0.231	0.207	3.8	1.4	2.0	1.6	1.4	0.8	0.7	0.6
-4	0.053	0.283	0.257	2.9	1.2	1.7	1.4	1.3	0.7	0.6	0.6
-3	0.058	0.341	0.312	2.2	0.9	1.5	1.3	1.2	0.6	0.6	0.5
-2	0.062	0.403	0.372	1.7	0.7	1.3	1.2	1.1	0.6	0.5	0.5
-1	0.064	0.467	0.435	1.3	0.6	1.1	1.1	1.1	0.5	0.5	0.5
0	0.065	0.533	0.500	1.0	0.4	1.0	1.0	1.0	0.4	0.4	0.4
1	0.064	0.597	0.435	1.3	0.6	1.7	2.2	2.8	0.7	0.9	1.1
2	0.062	0.659	0.372	1.7	0.7	2.8	4.8	8.1	1.1	1.7	2.4
3	0.058	0.717	0.312	2.2	0.9	4.8	11	24	1.7	2.7	4.0
4	0.053	0.769	0.257	2.9	1.2	8.3	24	70	2.4	4.1	5.4
5	0.047	0.816	0.207	3.8	1.4	15	55	213	3.3	5.1	5.6
6	0.041	0.857	0.164	5.1	1.8	26	134	>500	3.9	5.3	6.0
7	0.034	0.891	0.126	6.9	2.2	48	333	>500	4.9	5.8	6.1
8	0.028	0.919	0.095	9.5	2.6	91	>500	>500	5.0	6.0	6.2
9	0.022	0.941	0.070	13	3.1	177	>500	>500	5.5	6.1	6.2
10	0.017	0.959	0.050	19	3.6	362	>500	>500	5.8	6.2	6.2
11	0.013	0.972	0.035	28	4.0	>500	>500	>500	6.0	6.2	6.2
12	0.010	0.981	0.024	42	4.7	>500	>500	>500	6.1	6.2	6.2
13	0.007	0.988	0.015	64	5.3	>500	>500	>500	6.2	6.2	6.2
14	0.005	0.992	0.010	101	5.1	>500	>500	>500	6.2	6.2	6.2

Appendix I – R Code to Calculate the Seven Position Multinomial

```
sevenPositionMultinomialFn <- function(nSim=(3*10^6),
                                       numberSensors=4,
                                       numberChart=5,
                                       numberRQs=3,
                                       weighted=TRUE,
                                       multi=FALSE ) {
  # R function to simulate the multinomial distribution of 7 position
scores
  # under the theory of the polygraph test
  # Jan 11, 2022
  # Raymond Nelson
  ####
  # nSim input is the number of simulations to run
  # numberSensors is 4 with PLE and 3 without PLE
  # numberChart is 5, but can be 3 or 4
  # numberRQs is 3, but can be 2 or 4 or "rand"
  # weighted=TRUE will alter the scales for the sensors
  # pneumo -1, 0 +1
  # cardio -2, -2, 0, +1, +2
  # eda -3, -2, -1, 0, +1, +2, +3
  # vasomotor -1, 0, +1
  # multi=TRUE will calculate the multiplicity corrected odds and lower
limits
  #
  # output is a data frame
  ####

  # nScores <- 80 # 4 sensors * 4 RQs * 5 charts
  # nScores <- 60 # 4 sensors * 3 RQs * 5 charts
  # nScores <- 40 # 4 sensors * 2 RQs * 5 charts
  # nScores <- 20 # 4 sensors * 1 RQ * 5 charts
  # nScores <- 27 # 3 sensors * 3 RQs * 3 charts
  # nScores <- 60 # 3 sensors * 4 RQs * 5 charts
  # nScores <- 45 # 3 sensors * 3 RQs * 5 charts
  # nScores <- 30 # 3 sensors * 2 RQs * 5 charts
  # nScores <- 15 # 3 sensors * 1 RQ * 5 charts

  #### initialize a matrix by sampling from the 7 position scale ####
  if(!isTRUE(weighted)) {
    # for unweighted 7 position scores
    simDAT7 <- matrix(sample(seq(-3, 3, by=1), size=nSim*nScores,
replace=TRUE),
                      ncol=nScores)
    simDAT7 <- data.frame(simDAT7)
    names(simDAT7) <- paste0("x", 1:nScores)
  }

  #### weighted 7 position scores ####

  if(isTRUE(weighted)) {
```

```
    respirationScale <- c(-1:1)
    EDAScale <- c(-3:3)
    cardioScale <- c(-2:2)
    vmScale <- c(-1:1)

    nSensorScores <- numberCharts * numberRQs

    # respiration scores
    simDATResp <- matrix(sample(respirationScale, size=nSim*nSensorScores,
replace=TRUE),
                         ncol=nSensorScores)
    simDATResp <- data.frame(simDATResp)
    names(simDATResp) <- paste0("p", 1:nSensorScores)

    # EDA scores
    simDATEda <- matrix(sample(EDAScale, size=nSim*nSensorScores,
replace=TRUE),
                        ncol=nSensorScores)
    simDATEda <- data.frame(simDATEda)
    names(simDATEda) <- paste0("e", 1:nSensorScores)

    # cardio scores
    simDATCardio <- matrix(sample(cardioScale, size=nSim*nSensorScores,
replace=TRUE),
                           ncol=nSensorScores)
    simDATCardio <- data.frame(simDATCardio)
    names(simDATCardio) <- paste0("c", 1:nSensorScores)

    # vasomotor scores
    if(numberSensors == 4) {
      simDATVm <- matrix(sample(vmScale, size=nSim*nSensorScores,
replace=TRUE),
                         ncol=nSensorScores)
      simDATVm <- data.frame(simDATVm)
      names(simDATVm) <- paste0("v", 1:nSensorScores)
    } else {
      simDATVm <- NA
    }

    # combine the sensor scores
    simDAT7 <- cbind.data.frame(simDATResp, simDATEda, simDATCardio)
    if(!is.na("simDATVm")) {
      simDAT7 <- cbind(simDAT7, simDATVm)
    }

    # calculate the total score
    simDAT7$permutationScore <- apply(simDAT7[,1:nScores], 1, sum)

  }
```

```
#### count the number of occurrences for the range of scores####

{
  # library("dplyr")
  simDAT7Dist <- count(simDAT7, permutationScore)

  # check the max and min values
  maxVal <- max(simDAT7Dist$permutationScore)
  minVal <- min(simDAT7Dist$permutationScore)
  diffVal <- maxVal - abs(minVal)

  # set the max and min values as equal
  while(diffVal != 0) {
    thisMax <- which.max(c(maxVal, abs(minVal)))
    for(i in 1:length(diffVal)) {
      if(thisMax == 1) {
        simDAT7Dist <- rbind(c((minVal - 1), 0), simDAT7Dist)
      } else {
        simDAT7Dist <- rbind(simDAT7Dist, c((maxVal + 1), 0))
      }
    }
    maxVal <- max(simDAT7Dist$permutationScore)
    minVal <- min(simDAT7Dist$permutationScore)
    diffVal <- maxVal - abs(minVal)
  }
}

#### fix any missing rows ####

{
  simDAT7DistX <- simDAT7Dist[1,]
  # simDAT7DistX <- simDAT7DistX[1:102,]
  i=2
  for(i in 2:nrow(simDAT7Dist)) {
    if( simDAT7Dist[i,1] > (simDAT7DistX[nrow(simDAT7DistX),1]+1) ) {
      while( simDAT7DistX[nrow(simDAT7DistX),1] < (simDAT7Dist[i,1]-1)
) {
        simDAT7DistX <- rbind( simDAT7DistX,
c(simDAT7DistX[nrow(simDAT7DistX),1]+1, 0) )
      }
      simDAT7DistX <- rbind( simDAT7DistX, simDAT7Dist[i,] )
    } else {
      simDAT7DistX <- rbind( simDAT7DistX, simDAT7Dist[i,] )
    }
  }
  simDAT7Dist <- simDAT7DistX
}

#### initialize an output data frame ####

{
  maxVal <- max(simDAT7Dist$permutationScore)
  simDAT7Dist2 <- data.frame(permutationScore=seq(-maxVal, maxVal,
```

```
by=1))
    # private function
    findCount <- function(x, y=simDAT7Dist) {
      # a function to get count for each score in the simDAT7Dist2range
      # using the simDAT7Dist data frame
      y$n[y$permutationScore == x]
    }
    # call the function to populate the data frame n (count) for each
score
    simDAT7Dist2$n <- as.numeric(apply(simDAT7Dist2[1], 1, findCount))
  }

  #### calculate the probability mass function ####

  {
    simDAT7Dist2$pmf <-
      round( simDAT7Dist2$n / nSim, 9)
    # smooth it
    simDAT7Dist2$pmf <- rowMeans(cbind(simDAT7Dist2$pmf,
rev(simDAT7Dist2$pmf)))
  }

  #### calculate the cumulative distribution function and continuity ####

  {
    simDAT7Dist2$cdf <- round( cumsum(simDAT7Dist2$pmf) , 9)
    # calculate the continuity correction
    simDAT7Dist2$cdfContCor <- 0
    for(i in 2:nrow(simDAT7Dist2)) {
      if(simDAT7Dist2$permutationScore[i] <= 0) {
        simDAT7Dist2$cdfContCor[i] <- round(mean(c(simDAT7Dist2$cdf[(i-
1):i])), 9)
      } else {
        # use the lower tail for scores > 0
        simDAT7Dist2$cdfContCor[i] <- 1 -
round(mean(c(simDAT7Dist2$cdf[(i-1):i])), 9)
      }
    }
  }

  #### calculate the odds from the continuity corrected cdf column ####

  simDAT7Dist2$odds <- (1-simDAT7Dist2$cdfContCor) /
simDAT7Dist2$cdfContCor


  #### calculate the lower limit of the Clopper Pearson CI ####

  simDAT7Dist2$oddsLL05 <-
    apply(simDAT7Dist2['cdfContCor'], 1, clopperPearsonFn, n=nScores,
a2=.05, odds=TRUE, output="lower")

  #### calculate the multiplicity corrected odds and lower limits ####
```

```
  if(isTRUE(multi) && numberRQs==1) {
    # multiplicity for deceptive subtotals of single issue exams
    # 2 RQs
    simDAT7Dist2$odds2RQ <- apply(simDAT7Dist2['odds'], 1,
mutiplicityOddsFn, n=2, inv=FALSE)
    # 3 RQs
    simDAT7Dist2$odds3RQ <- apply(simDAT7Dist2['odds'], 1,
mutiplicityOddsFn, n=3, inv=FALSE)
    # 4 RQs
    simDAT7Dist2$odds4RQ <- apply(simDAT7Dist2['odds'], 1,
mutiplicityOddsFn, n=4, inv=FALSE)
    # multiplicity for truthful subtotals of multiple issue exams
    theseRows <- which(simDAT7Dist2$permutationScore > 0)
    # 2 RQs
    simDAT7Dist2$odds2RQ[theseRows] <-
      apply(simDAT7Dist2['odds'], 1, mutiplicityOddsFn, n=2, inv=TRUE)
[theseRows]
    # 3 RQs
    simDAT7Dist2$odds3RQ[theseRows] <-
      apply(simDAT7Dist2['odds'], 1, mutiplicityOddsFn, n=3, inv=TRUE)
[theseRows]
    # 4 RQs
    simDAT7Dist2$odds4RQ[theseRows] <-
      apply(simDAT7Dist2['odds'], 1, mutiplicityOddsFn, n=4, inv=TRUE)
[theseRows]
    # clopper pearson lower limit
    simDAT7Dist2$p2RQ <- simDAT7Dist2$odds2RQ / (1+simDAT7Dist2$odds2RQ)
    simDAT7Dist2$p3RQ <- simDAT7Dist2$odds3RQ / (1+simDAT7Dist2$odds3RQ)
    simDAT7Dist2$p4RQ <- simDAT7Dist2$odds4RQ / (1+simDAT7Dist2$odds4RQ)
    # 2 RQs
    simDAT7Dist2$odds2LL05 <-
      apply(simDAT7Dist2['p2RQ'], 1, clopperPearsonFn, n=nScores, a2=.05,
odds=TRUE, output="lower")
    # 3 RQs
    simDAT7Dist2$odds3LL05 <-
      apply(simDAT7Dist2['p3RQ'], 1, clopperPearsonFn, n=nScores, a2=.05,
odds=TRUE, output="lower")
    # 4 RQs
    simDAT7Dist2$odds4LL05 <-
      apply(simDAT7Dist2['p4RQ'], 1, clopperPearsonFn, n=nScores, a2=.05,
odds=TRUE, output="lower")
  }

  return(simDAT7Dist2)

}


mutiplicityOddsFn <- function(odds=1.38, n=3, inv=FALSE) {
  # R function to calculate a multiplicity correction for an odds ratio
  # Raymond Nelson
  # Jan 11, 2022
  # odds can be a whole number or decimal
  # n is the number of simultaneous odds or decisions
```

```
  # inv will invert the calculation
  ###
  # p <- odds / (1 + odds)
  # mp <- 1- ( 1 + exp( log(p /(1-p)) / n ) )^-1
  if(isTRUE(inv)) n <- 1/n
  mp <- 1 - ( 1 + exp( log(odds) / n ) )^-1
  oddsM <- mp / (1-mp)
  return(oddsM)
}


clopperPearsonFn <- function(p=,
                             n=,
                             a2=.05,
                             odds=FALSE,
                             output="both") {
  # Clopper-Pearson exact binomial confidence interval
  # for posterior odds of seven position scores
  # Jan 11, 2022
  # Raymond Nelson
  ####
  # p input is a decimal probablity value
  # n input is the sample size (number of scores)
  # a2 is the one-tailed alpha level for the lower limit post odds
  # odds is a boolean value to obtain the output in the form of odds
  # p input should always be in decimal probability form when odds=TRUE
  # output="both" will give lower and upper limits
  # output="lower" will return the lower limit
  # output="upper" will return the upper limit
  #
  # ouput is a named vector with the lower limit (LL) and upper limit
(UL)
  # of the Clopper-Pearson interval
  #
  # Clopper-Pearson interval is an exact interval based on the binomial
  # actual coverage always exceeds the nominal level
  ####
  # n = number of scores
  # n = 15 for subtotal scores with PLE sensor
  # n = 20 for subtotal scores without PLE sensor
  # n = 30 for 2 RQs without PLE
  # n = 40 for 2 RQs with PLE
  # n = 45 for 3 RQs without PLE
  # n = 60 for 3 RQs with PLE
  # n = 60 for 4 RQs without PLE
  # n = 80 for 4 RQs with PLE
  ####
    # set p on the upper tail
    if(p < .5) p = 1-p
    # p compliment
    q <- 1 - p
    # degrees of freedom
    v1=trunc(round(2*(n*q+1),0))
    v2=trunc(round(2*n*p,0))
```

```
    v3=trunc(round(2*(n*p+1),0))
    v4=trunc(round(2*n*q,0))
    # truncated degrees of freedom this way
    # give the same result as Excel
    if(v1 == 0) v1 = 1
    if(v2 == 0) v2 = 1
    if(v3 == 0) v3 = 1
    if(v4 == 0) v4 = 1
    # calculate the f statistics for the upper and lower limit
    FLow <- qf((1-a2),v1,v2)
    FUp <- qf(a2,v3,v4)
    # calculate the limits
    LL <- (1+FLow*(q+1/n)/p)^-1
    UL <- (1 + q / (1 / n + p) * FUp)^-1
    # transform the result to odds form
    if(isTRUE(odds)) {
      LL <- LL / (1 - LL)
      UL <- UL / (1 - UL)
    }
    # output either both tails or the lower
    ifelse(output=="both",
           outVector <- c(lowerLimit=LL, upperLimit=UL),
           ifelse(output=="lower",
                  outVector <- LL,
                  outVector <- UL ) )
    return(outVector)
}
```

AMERICAN POLYGRAPH ASSOCIATION

DEDICATED TO TRUTH