

Bases Científicas de la Examinación Poligráfica

Raymond Nelson

Abstract

Se revisa la literatura científica para la evaluación poligráfica con preguntas de comparación y su aplicación en contextos diagnósticos y exploratorios. La revisión resume la literatura de todos los aspectos del procedimiento de la prueba, incluyendo la entrevista de pretest, la recolección de datos de prueba, el análisis de datos prueba y una propuesta de las bases fisiológicas y psicológicas de la prueba poligráfica. Se resume información de la precisión del polígrafo para exámenes de diagnóstico y exploratorios. Se revisa la evidencia de las amenazas en contra de la precisión del polígrafo y de la contribución de los resultados poligráficos para la validez incremental o la precisión de la decisión incremental por parte de los consumidores profesionales de los resultados de la prueba poligráfica. Se describe al polígrafo como una prueba probabilística y no-determinista que involucra tanto el registro de fisiología como métodos estadísticos. En general, se necesitan pruebas probabilísticas, modelos estadísticos y pruebas científicas cuando no es posible una observación determinista ni una medición física. Los polígrafos diagnósticos de evento específico han mostrado que proporcionan una precisión media de .89 con un rango de confianza del 95% desde .83 y hasta .95. Los polígrafos exploratorios de asunto-múltiple han mostrado que proporcionan tasas de precisión, con una media de .85 y un rango de confianza de 95% de .77 a .93.

Palabras clave: *Polígrafo, detección de mentiras, detección de señales, análisis de datos de prueba, bases científicas.*

Las exámenes poligráficos, al igual que otras pruebas científicas, puede tomar la forma tanto de pruebas de diagnóstico como de pruebas exploratorias. La diferencia entre los exámenes diagnósticos y exploratorios es que las exámenes diagnósticos involucran la existencia de un problema conocido, en la forma de síntomas, evidencia, alegatos o incidencias circunstanciales que sugieren que un individuo podría tener algún involucramiento, para el que los resultados de la examinación intenta dar soporte a una conclusión diagnóstica positiva o negativa. Las pruebas exploratorias incluyen todas las pruebas que se conducen en la ausencia de un incidente conocido, alegato conocido o problema conocido.

Nota del autor: Raymond Nelson es un examinador poligráfico y psicoterapeuta quien es autor de numerosos artículos y estudios sobre muchos aspectos de la prueba poligráfica. Mr. Nelson es un investigador asociado con Lafayette Instrument Company, es el director curricular de un programa de entrenamiento poligráfico acreditado y es miembro electo de Mesa Directiva de la American Polygraph Association – actualmente sirve en el rol de Presidente. Mr. Nelson ha enseñado en varios programas de entrenamiento poligráfico, es un ponente y presentador activo en conferencias internacionales, y ha sido testigo experto en diversas cortes judiciales tanto en asuntos poligráficos como terapéuticos. Las perspectivas y opiniones expresadas en esta publicación son del autor y no necesariamente las de Lafayette Instrument Company o de la American Polygraph Association. Consultas y correspondencia puede enviarse a raymond.nelson@gmail.com

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

El propósito de los exámenes diagnósticos es la de lograr una conclusión que podría servir como base de una acción. Esta acción a menudo afectará el futuro de un individuo en términos de sus derechos, libertades o salud. Por esta razón, es difícil imaginar una justificación ética para la selección de una técnica de examinación que proporcione algo menor al nivel factible más alto de precisión diagnóstica. Los exámenes diagnósticos alcanzan altos niveles de precisión en la decisión, en parte, mediante la restricción de la prueba a un solo tema de preocupación.

Por el contrario, las pruebas exploratorias, tienen el propósito de añadir validez incremental a las decisiones en el manejo de riesgos que se realizan en ausencia de un problema conocido. Esto se logra tanto por la recopilación de información como por la investigación de la posible participación de un individuo en uno o varios temas de interés. Las pruebas exploratorias no deben de ser utilizadas de forma aislada como la base de una acción que podría afectar los derechos, libertades o salud de un individuo. La ausencia de un problema conocido es la característica que define una prueba exploratoria (Wilson, & Jungner, 1968; Raffle, & Muir Gray, 2007). Las pruebas poligráficas exploratorias tienen el objetivo de agregarse incrementalmente al manejo de riesgos mediante una combinación de metas pequeñas que podrían incluir tanto la habilidad de discriminación del resultado de prueba, como de la capacidad del proceso de prueba para desarrollar información. Los programas poligráficos exploratorios pueden también involucrar una tercera meta en la forma de impedimento incremental de problemas (American Polygraph Association, 2009a; 2009b). Los objetivos disuasivos podrían lograrse al impedir a las personas de alto-riesgo el acceso a ambientes de alto-riesgo (por ejemplo la exploración de aplicantes a policía o la exploración de seguridad gubernamental/operacional), o mediante la disuasión de (o decremento) del no-cumplimiento con políticas, reglas y normas (por ejemplo, políticas de seguridad operacional).

Discusión

De acuerdo con la Asociación Americana de Poligrafía (2011), las exámenes poligráficos constan de tres fases: 1) entrevista de pre-test, 2) una fase in-test de recolección de datos de prueba, y 3) análisis de datos de prueba. Cada una de estas fases tiene un efecto importante tanto en la precisión de la prueba como en la utilidad del resultado de la prueba poligráfica. Por esta razón, todos los supuestos y procedimientos que son considerados fundamentales para el resultado poligráfico, idealmente deberán basarse en un conocimiento con aceptación generalizada o en evidencia y constructos teóricos para los que existan soporte empírico publicado y replicado.

Entrevista de pre-test poligráfica

En su forma más básica, una entrevista no es más que una conversación con un propósito (Hodgson, 1987), y, como lo indican Kahn y Cannel (1957), el éxito de muchos esfuerzos profesionales depende en parte de la capacidad de obtener información de otros. La entrevista de pre-test del polígrafo, está destinada a orientar al examinando a los procedimientos de prueba, al propósito de la prueba y a las preguntas objetivo de investigación. La premisa básica de la entrevista sugiere que las personas reportan información más útil e interesante cuando se lo pide alguien que escucha con interés y construye rapport mediante el uso de la conversación y de preguntas de entrevista. Las entrevistas de pretest poligráficas intentan permitir a los examinados veraces ajustarse - o habituarse - al impacto cognitivo y emocional de escuchar y responder a las preguntas estímulo de la prueba que

describen su posible involucramiento en conductas problemáticas, al tiempo que también sensibilizan o incrementan la conciencia y el potencial de respuesta en los examinados no veraces a las preguntas del examen que describen sus conductas pasadas.

La entrevista de pretest poligráfica es un proceso que consta de varias etapas (American Polygraph Association, 2009a, 2009b; Department of Defense Polygraph Institute, 2002), que incluye: una entrevista de narrativa-libre (Powell & Snow, 2007), una entrevista semi-estructurada (Lindlof & Taylor, 2002), o una entrevista estructurada (Drever, 1995), una revisión a conciencia de las preguntas estímulo de prueba, y la práctica de una prueba de orientación. El primer objetivo de la entrevista pretest consiste en establecer una identificación positiva e introducción, y clarificar los roles del examinador y del examinado. El examinador también presentará al examinado la sala de evaluación, incluyendo el uso de dispositivos de grabación de audio y vídeo, y de todos los sensores poligráficos que posteriormente serán ajustados al examinado.

La siguiente etapa del proceso consiste en realizar una determinación inicial de la idoneidad del examinado y obtener el consentimiento informado para la prueba. Esto se hace después de revisar los derechos de la persona examinada durante la prueba, incluyendo el derecho de suspender el examen en cualquier momento. Idealmente, el consentimiento informado incluirá además información acerca de quien recibirá la información y los resultados de la prueba, y de donde obtener mayor información acerca de las fortalezas y debilidades del procedimiento poligráfico. El examinador entonces involucrará al examinado en una discusión respecto a los antecedentes del caso y de los antecedentes personales del examinado, con el fin de continuar con el establecimiento de un rapport de prueba adecuado e idóneo. El examinador proporcionará también más información acerca de las bases psicológicas y fisiológicas de la prueba del polígrafo y proporcionará respuestas a las preguntas que el examinado podría tener acerca de los procedimientos de prueba.

Deberá llevarse a cabo una prueba de práctica o prueba de familiarización como parte de una práctica de campo estandarizado (American Polygraph Association, 2009a, 2009b; Department of Defense, 2006a). El propósito de esta prueba es orientar al examinado ante el procedimiento de prueba antes de comenzar el examen real. La investigación realizada por Kircher, Packard, Bell & Bernhardt (2001) ha demostrado que esto puede contribuir en un incremento en la precisión de la prueba. Un punto de vista científico, apoyado por estudios recientes con examinados no-ingenuos que eran plenamente conscientes de los detalles de los procedimientos de prueba (Honts y Reavy, 2009; Honts y Alloway, 2007; Nelson, Handler, Blalock y Hernández, 2012; Rovner, 1986), sostiene que la efectividad de las pruebas científicas basadas en evidencia no depende del sistema de creencias del examinado. El propósito de la prueba de familiarización no es demostrar o convencer al examinado de creer que la prueba del polígrafo es infalible, pero sí de orientar al examinado hacia los procedimientos de prueba. Independientemente de las actitudes o creencias del examinador o del examinado con respecto a la prueba de familiarización, los estudios científicos (Bradley y Janisse, 1981; Horneman y O'Gorman, 1985; Horowitz, Kircher y Raskin, 1986; Kirby, 1981; Widup, R, Jr y Barland, 1994) han demostrado que el uso de una prueba de familiarización no daña y podría en ocasiones incrementar la precisión del resultado del examen poligráfico. La razón real de este efecto podría tener más que ver con la certeza de que el instrumento y los sensores están ajustados y funcionando adecuadamente y de que el examinado ha tenido la oportunidad de practicar el cumplimiento de las instrucciones de comportamiento.

La siguiente etapa de la entrevista de pre-test será una entrevista de narrativa libre, una entrevista estructurada o una entrevista semi-estructurada. Las entrevistas de narrativa-libre se caracterizan por el uso de un lenguaje sencillo y común, con ausencia de técnicas coercitivas, como una oportunidad para que el entrevistado comunique detalles al nivel de su propia elección, y propiciando la elaboración. Las entrevistas de narrativa libre realizadas durante las pruebas de polígrafo pueden incluir preguntas directas o de exploración con respecto al incidente presunto o conocido, antes de proceder a la construcción de las preguntas de la prueba poligráfica. Las estrategias de entrevista de narrativa-libre son muy útiles durante las investigaciones diagnósticas, pero no son idóneas para su uso en pruebas de exploración, que se llevan a cabo en ausencia de un incidente presunto o conocido. Las entrevistas de Pretest para exámenes exploratorios que se realizan durante los exámenes poligráficos exploratorios, así pertenezcan a seguridad operacional, de pre-empleo en fuerzas públicas, o de supervisión posterior a una condena, tomarán la forma de una entrevista estructurada o de una entrevista semi-estructurada.

Las entrevistas estructuradas se diferencian de las entrevistas semi-estructuradas en que se llevan a cabo al pie de la letra, sin desviación del protocolo de la entrevista (General Accounting Office, 1991; Campion, Campion, & Hudson, 1994; Kvale, 1996). En contraste, las entrevistas semi-estructuradas se realizan con un contenido y esquema de preguntas en el cual se permite que el entrevistador pueda presentar las preguntas de entrevista de manera tal que sean individualizadas con base en la personalidad, nivel de educación y rapport entre el entrevistador y el entrevistado. Aunque las entrevistas estructuradas son las preferidas por los investigadores y administradores de programas por su consistencia, las entrevistas estructuradas hacen poco uso de la habilidad y experiencia del entrevistador.

Las entrevistas semiestructuradas intentan hacer un uso más eficiente de la habilidad y experiencia del entrevistador para un acceso más rico de información con respecto al contenido de la entrevista. Tanto las entrevistas estructuradas como semi-estructuradas deberán anclarse en un programa o protocolo de entrevista definido, con definiciones operacionales claramente formuladas que describan los asuntos comportamentales de preocupación. En comparación con los métodos de entrevistas estructuradas, las estrategias de entrevista semi-estructurada fomentan y dependen de mayores habilidades de entrevista. Se sabe que las habilidades de entrevista semi-estructurada son más eficaces para obtener información rica y detallada en comparación con los protocolos de entrevistas estructuradas. Al igual que los métodos de entrevista estructurada, los protocolos de la entrevista semi-estructurados requieren que todos los temas y las preguntas de entrevista sean abordados en algún momento durante la entrevista.

En la última etapa de la entrevista de pre-test - después de la entrevista de narrativa libre o de la entrevista semi-estructurada - el examinador desarrollará y revisará las preguntas de prueba con el examinado (American Polygraph Association, 2009a, 2009b; Department of Defense, 2002). El lenguaje de las preguntas de prueba se ajusta para asegurar una correcta comprensión y dar cuenta de la información o admisiones que el examinado ha proporcionado durante la entrevista o durante el desarrollo de las preguntas del examen. Las preguntas relevantes describirán el posible involucramiento conductual del examinado en el o los asuntos de preocupación. Estas preguntas

evitarán en general asuntos relacionados con memoria, intentos o motivación. De cualquier manera, algunos protocolos de prueba de investigación permitirán preguntas de prueba relacionadas con la memoria, si es que el examinado admite el acto conductual alegado y el asunto de la memoria o motivación es el objetivo de la investigación (American Polygraph Association, 2009a).

Cuando un examen de polígrafo consta de series múltiples de preguntas de prueba, el examinador revisará cada serie de preguntas por separado, y luego llevará a cabo la fase in-test de recolección de datos de prueba para cada serie de preguntas antes de la revisión y recolección de datos para cada serie de preguntas subsecuente. Cuando un polígrafo consta de series múltiples de preguntas de prueba, no hay evaluación o discusión de los resultados de cualquier serie individual de preguntas hasta que todas las series de preguntas de prueba hayan sido completamente registradas y analizadas. Si previamente no se ha conducido una prueba de familiarización, podría conducirse después de revisar las preguntas de prueba y antes de proceder con la fase in-test de la prueba. Anteriormente, algunos formatos de prueba poligráfica empleaban un procedimiento análogo a una prueba de familiarización, aunque era después de la primera presentación de los estímulos de prueba durante la fase in-test del examen.

Recolección de datos de prueba In-Test

La segunda fase del examen poligráfico es la recolección de datos en el in-test. Este se puede llevar a cabo utilizando cualquiera de una variedad de formatos de prueba validados de diagnóstico o exploratorios (American Polygraph Association, 2011b; Department of Defense, 2002). Todas las técnicas poligráficas de diagnóstico y exploratorias incluyen preguntas relevantes (RQ) que describen la posible participación del examinado en las conductas que son asuntos bajo investigación. Las preguntas relevantes efectivas deberán de ser simples, directas y deberán evitar terminología legal o clínica y palabras para las que el significado correcto pudiera ser ambiguo, confuso o no reconocible para personas no familiarizadas con vocabulario legal o profesional. Cada pregunta relevante deberá abordar un asunto comportamental único.

Preguntas relevantes del caso específico

Los polígrafos diagnósticos se construyen bajo el supuesto de criterio de varianza no-independiente. El significado científico y probabilístico de esto es que las RQs tienen una fuente común o compartida de varianza de respuesta porque los estados de criterios externos de las diferentes RQs pueden (y lo hacen) afectarse mutuamente. El significado práctico de esto es que todas las RQs deben abordar conductas dentro de un incidente único de preocupación.

Los polígrafos exploratorios Multi-tema, conducidos en ausencia de un alegato o incidente conocido, puede construirse con preguntas relevantes que describen distintos comportamientos para los que se asume que los estados de criterio externo varían de forma independiente (es decir, se asume que los estados de criterio externo serán exclusivos o no interactuarán o se afectarán entre sí). Hay evidencia de que la varianza de respuesta para estas preguntas realmente no es independiente (Barland, Honts y Barger, 1989; Podlesny y Truslow, 1993; Raskin, Honts y Kircher, 2014), y por esta razón las prácticas de campo no permiten resultados de prueba tanto positivos como negativos dentro

de un solo examen. Independientemente de si se conducen con fines de diagnóstico o exploratorios, todos los exámenes de polígrafo se interpretan en última instancia, a nivel de la prueba como un todo, aunque las puntuaciones subtotales para las RQs individuales pueden ser evaluadas de acuerdo a los procedimientos estandarizados.

En la actualidad, la mayoría de los exámenes de polígrafo en los Estados Unidos se conducen con una variante de la técnica de preguntas de comparación (CQT). La CQT fue inicialmente descrita en la publicación de Summers (1939), cuando él era jefe del Departamento de Psicología en la Fordham University Graduate School en Nueva York. El CQT fue popularizado dentro de la profesión poligráfica por Reid (1947) y Backster (1963). Es la más comúnmente utilizada y exhaustivamente investigada de las familias de técnicas poligráficas en uso hoy en día. Además de las RQs, estas técnicas poligráficas incluyen también preguntas de comparación (CQs; referidas previamente en la literatura poligráfica como preguntas de control). Cuando se califica una prueba, los examinadores evaluarán numéricamente y estadísticamente las diferencias en las respuestas a las RQs y las CQs.

La forma tradicional de preguntas de comparación, son las preguntas comparativas de mentira probable (PLC), aunque algunos CQTs contemporáneos basados en evidencia hacen uso de la comparación de mentira-dirigida (DLC). Los examinadores que usan PLC manipulan al examinado para que nieguen un asunto comportamental común que no es el objetivo bajo investigación. Las preguntas de comparación de mentira-probable, han sido la base para algunas críticas de la técnica poligráfica debido a su naturaleza manipulativa, y también por la incertidumbre en torno a la veracidad de los examinados con respecto a estas preguntas. (Office of Technology Assessment, 1983; Furedy, 1989 ; Lykken, 1981; Saxe, 1991). Algunas de estas críticas descansan sobre un supuesto inexacto de que el polígrafo mide mentiras reales *per se*. El polígrafo, al igual que muchas pruebas científicas, mide respuestas a estímulos. El instrumento poligráfico en realidad no mide *mentiras*, pero sí discrimina el engaño y la veracidad mediante el uso de modelos de probabilidad y datos de referencia estadísticos que describen las diferencias entre los patrones de reacción de las personas veraces y de quienes engañan cuando responden a las RQs y CQs.

Aunque no es un *control* en un sentido estricto, las CQs cumplen con una función similar a un control en el sentido de que permiten que un examinador analice y compare eficientemente la varianza diagnóstica y otras fuentes de varianza. La varianza de respuesta de las CQs no es completamente independiente del asunto objeto de investigación de la misma manera que lo son los controles científicos - porque las respuestas en las CQs y las RQs vienen del mismo examinado. Este modelo de prueba puede ser pensado como análogo a la forma en que se adquieren datos de los sujetos en un diseño ANOVA de mediciones repetidas de dos-vías - en la que cada sujeto sirve como su propio set de control. De esta manera, cada examen poligráfico sirve como una forma de experimento científico en un solo sujeto.

Las preguntas de Comparación de Mentira Dirigida (DLC) se han sugerido como una alternativa al uso de las PLCs (Barland, 1981; Research Division Staff, 1995a; 1995b). Las DLCs se utilizan en técnicas poligráficas desarrolladas por el gobierno de los Estados Unidos (U.S.) para su uso en programas poligráficos exploratorios, y en técnicas poligráficas diagnósticas desarrolladas por

investigadores de la Universidad de Utah (Honts & Raskin, 1988; Kircher, Honts & Raskin, 1997) y en el U.S. Department of Defense (Honts & Reavy, 2009). La mayor diferencia entre las técnicas PLC y las DLC es que las técnicas DLC son transparentes y pueden ser utilizadas sin la necesidad de manipular o manejar al examinado para que niegue un asunto comportamental común.

Las DLCs han demostrado en numerosos estudios, resumidos por Blalock, Nelson, Handler y Shaw (2011; 2012), que realizan tareas de clasificación con eficiencia similar y distribuciones estadísticas similares de calificaciones numéricas (American Polygraph Association, 2011) en comparación con los exámenes PLC. Algunos investigadores han sugerido que las DLC son éticamente menos complicadas que las PLC, ya que no requieren que el examinador manipule psicológicamente al examinado (Honts y Raskin, 1988; Honts y Reavy, 2009; Horowitz, Kircher, Honts y Raskin, 1997; Kircher, Packard, Bell, y Bernhardt, 2001; Raskin & Kircher, 1990). Los exámenes DLC también han demostrado que mantienen su eficacia en diferentes idiomas y culturas (Nelson, Handler & Morgan, 2012)

Además de las preguntas PLC y DLC, se han sugerido y argumentado otras variantes de las preguntas de comparación, incluidas las preguntas de comparación excluyentes y las preguntas de comparación no excluyentes (es decir, inclusivas). Los estudios han demostrado que todas estas variantes de CQ se comportan con efectividad similar, para la que la precisión no difiere en un nivel estadísticamente significativo (Amsel, 1999; Honts y Reavy, 2009; Horvath y Palmatier, 2008; Horvath, 1988; Palmatier, 1991). Una investigación meta-analítica reciente (American Polygraph Association, 2011b) ha solidificado aún más esta conclusión, lo que demuestra que las mismas técnicas de polígrafo se desempeñan con eficacia equivalente y no hubo diferencias significativas en las distribuciones de muestreo de criterio de calificaciones de engaño y criterio de veracidad, cuando las técnicas son empleadas con preguntas PLC o DLC. Los supuestos científicos subyacentes de los modelos de calificación para técnicas PLC y DLC, asumen solamente que los examinados responderán de manera diferente a los estímulos relevantes y de comparación en función del engaño en respuesta a las RQs.

Todas las técnicas poligráficas podrían incluir otras preguntas de procedimiento que no son calificadas numéricamente. Las preguntas de procedimiento diseñadas para otros fines técnicos de examinación, no han sido avaladas por estudios científicos, incluidas las preguntas de verdad general (Abrams, 1984; Hilliard, 1979), preguntas de asuntos externos - referidas también como preguntas "sintomáticas" - que intentan cuestionar la interferencia de algo externo en el ámbito de las preguntas de examinación (Honts, Amato & Gordon, 2004 ; Krapohl & Ryan, 2001), preguntas de complejo de culpa (Podlesny, Raskin y Barland, 1976), y preguntas relevantes de sacrificio, relacionadas con la intención de un examinado a contestar con la verdad (Capps, 1991; Horvath, 1994). La falta de evidencia para apoyar su validez ha llevado al abandono en la utilización de la mayoría de estas preguntas técnicas.

Solamente dos preguntas técnicas no calificables se mantienen con amplio uso hoy en día, y ellas solo se utilizan en un sentido procesal. Aunque no se califican numéricamente y no se incluyen en los modelos de decisión estructurados o estadísticos, las preguntas de asuntos externos se han retenido como parte estructural y procesal de algunos formatos de prueba. Igualmente, aunque

nunca se calificaron numéricamente y no están incluidas en los modelos de decisión estructurados estadísticos, las preguntas relevantes de sacrificio son valoradas por el pretendido propósito de absorber y descargar la respuesta inicial del evaluado a la primera pregunta que describe el objetivo de la investigación. Estas preguntas tampoco son calificadas numéricamente y no se incluyen en los modelos de decisión estructurados o estadísticos. Las preguntas de sacrificio no-calificables se incluyen en prácticamente todas las técnicas modernas de polígrafo en uso hoy en día.

Un principio básico de la medición y la evaluación es obtener varias mediciones por cada elemento de preocupación. Esto se logra durante la prueba de polígrafo mediante el uso de diferentes sensores componentes, cada uno de los cuales está diseñado para monitorear incrementos o cambios en la actividad del sistema nervioso autónomo, y mediante la práctica estandarizada de agregar o combinar las respuestas de varias presentaciones de cada estímulo de prueba (Kircher y Raskin, 1988; Bell, Raskin, Honts, y Kircher 1999; Raskin, Kircher, Honts, y Horowitz, 1988; Reid, 1947; Research Division Staff, 1995a; 1995b). Los procedimientos poligráficos de campo (Kircher y Raskin, 1988; Handler & Nelson, 2008, Department of Defense, 2006a) requieren que los estímulos de prueba sean presentados al menos tres veces y un máximo de cinco. El método común es repetir la serie completa de preguntas de prueba, mediante una pausa en la grabación y desinflado del sensor de cardio entre repeticiones. Algunos protocolos de examen (Department of Defense, 1995a, 1995b; Handler, Nelson & Blalock, 2008) obtienen diferentes repeticiones de las preguntas del examen sin detener la examinación.

Análisis de Datos de Prueba - calificación de exámenes poligráficos.

Antes de informar al examinado o a alguien más de los resultados del examen poligráfico, el examinador debe analizar los datos de prueba. Los procedimientos para el análisis de datos de prueba están diseñados para dividir y comparar las fuentes de varianza de respuesta: varianza en respuesta a las RQs y varianza en respuesta a las CQs. Las respuestas se codifican numéricamente y el resultado se compara en contra de las calificaciones de corte que representan las expectativas normativas para personas con engaño o veraces. La teoría general de la prueba poligráfica es que las respuestas a las RQs y CQs varían significativamente en función del engaño y veracidad en respuesta a las RQs.

La premisa básica de la calificación numérica de los exámenes poligráficos fue descrita inicialmente por Kubis (1962) con un método similar al de Likert (1932), quien mostró cómo reducir la subjetividad utilizando la codificación numérica de datos de respuesta no lineales ordinales. La calificación numérica fue popularizada en la profesión poligráfica por Backster (1963) como el sistema de calificación de siete posiciones, y que ha sido objeto de un desarrollo y refinamiento posterior a través del estudio empírico. La construcción del procedimiento para la evaluación de las diferencias en la reacción a las RQs y las CQs se puede remontar a Summers (1939), quien utilizó una secuencia de preguntas que constaba de tres preguntas objetivo relevantes y tres preguntas de respuesta comparativa repetidas tres veces. Se encontró que los datos resultantes producían diferentes distribuciones para los diferentes grupos de criterio, y estas distribuciones pueden ser utilizadas para clasificar otras observaciones de casos. Las variantes de este modelo se observan tanto en la teoría de la detección de la señal y de la discriminación de la señal (Wickens, 1991; 2002).

Hay tres variantes de uso común del sistema de calificación de siete-posiciones hoy en día, incluyendo el modelo desarrollado por el Gobierno de los Estados Unidos (Department of Defense, 2006a; 2006b), el modelo publicado por ASTM International (2002), y el desarrollado por investigadores de la Universidad de Utah (Bell, Raskin, Honts, y Kircher, 1999; Kircher y Raskin, 1988; Raskin y Hare, 1978) y descrita por Handler (2006) y Handler y Nelson (2008). Las diferencias entre estos métodos de siete-posiciones son procesales y podrían ser intrascendentes en términos de la precisión de la prueba (American Polygraph Association, 2011b).

Una modificación comúnmente utilizada del modelo de calificación de siete-posiciones es el sistema de tres-posiciones definida por el Department of Defense (2006a, 2006b). Los modelos de calificación de tres-posiciones son favorecidos por algunos examinadores por su simplicidad y confiabilidad, aunque hay un aumento conocido en los resultados inconclusos de prueba (American Polygraph Association, 2011b) cuando se utilizan cortes de calificación numéricos destinados para puntuaciones de siete-posiciones. El Sistema de Puntuación Empírica (Nelson, Krapohl y Handler, 2008; Nelson & Handler, 2010; Nelson et al., 2011) es una modificación estadísticamente referenciada, estandarizada y basada en evidencia de los modelos de calificación de siete-posiciones y tres-posiciones, con una precisión de prueba comparable con otros modelos de calificación, aunque sin el incremento de resultados inconclusos.

Como una cuestión teórica, la calificación de los exámenes poligráficos no es diferente a la evaluación de otras pruebas científicas en medicina, psicología y forenses, e involucra cuatro preocupaciones básicas: 1) la identificación de criterios medibles y observables (es decir, características de calificación), 2) transformación de características de calificación a valores numéricos, y la reducción de los valores numéricos a índices de gran total para la examinación como un todo y de índices subtotales para los elementos individuales del examen 3) distribuciones de referencia estadísticas para calcular los clasificadores estadísticos y cortes numéricos, y 4) políticas de decisión estructuradas.

Las discusiones sobre los métodos de calificación poligráficos son inseparables de las discusiones acerca de la teoría de la prueba, incluyendo tanto a la teoría de decisión como a la teoría de la detección de señales. La teoría de la decisión (Greenberg, 1982; Lehmann, 1950; Lieblich Ben-Shakhar, Kugelmass, & Cohen, 1978; Pratt, Raiffa, & Schlaifer, 1995; Wald, 1939;), así como la teoría del aprendizaje estadístico (Hastie, Tibshirani & Friedman, 2001) se preocupan de la toma de decisiones óptimas. La teoría de la detección de la señal se ocupa de identificar y separar la información útil, del ruido de fondo o de la información aleatoria Green, & Swets. 1966; Marcum, 1947; Schonhoff, & Giordano, 2006; Swets, 1964; Swets, 1996; Tanner, Wilson, & Swets, 1954). La teoría de detección de señal incluye dos modelos fundamentales: la detección de la señal (por ejemplo, Sí o No) (Wickens, 2002) y los modelos de discriminación de señales (por ejemplo, A o B) (Wickens, 1991). La prueba de polígrafo CQT representa la segunda forma de estas dos, ya que las decisiones del polígrafo intentan lograr precisión diagnóstica al colocar o predecir la pertenencia de un examinado individual en categorías criterio de engaño y veracidad. Esfuerzos recientes (Nelson, Krapohl y Handler, 2008; Nelson et al, 2011;. Nelson & Handler, 2010) han comenzado a utilizar de

forma más extensa la teoría de la decisión estadística para cuantificar la probabilidad de un resultado de prueba poligráfica erróneo.

Características de reacción fisiológica.

La calificación de exámenes poligráficos comienza con la identificación de respuestas fisiológicas observables o medibles que estén correlacionadas con el engaño y que se puedan combinar en un modelo de diagnóstico eficiente y efectivo. Una serie de estudios - en gran parte financiados por el Departamento de Defensa de Estados Unidos y llevados a cabo en la Universidad de Utah y la Universidad Johns Hopkins - han descrito investigaciones sobre la identificación y extracción de características de calificación poligráficas (Bell, Raskin, Honts, y Kircher, 1999; Harris, Horner, y McQuarrie, 2000; Kircher, Kristjansson, Gardner, y Webb, 2005; Kircher y Raskin, 1988; Raskin et al., 1988). Estos esfuerzos se reflejan en los estándares de práctica de campo publicados por el Departamento de Defensa (2006a; 2006b), por ASTM International (2002), y en publicaciones acerca del Sistema de Calificación Empírica (Nelson & Handler, 2010; Nelson et al., 2011).

Un pequeño número de indicadores fisiológicos han demostrado repetidamente estar correlacionados con el engaño, en modelos de decisión estructurales que se utilizan actualmente en los programas de polígrafo de campo. Ellos son: la respiración - observada por la longitud de la línea respiratoria (Kircher y Raskin, 1988), la longitud de la excursión respiratoria (Kircher y Raskin, 2002), o ya sea por disminuciones sostenidas en la amplitud respiratoria durante tres o más ciclos respiratorios, desaceleración de la tasa respiratoria durante tres o más ciclos, o aumento temporal de la línea base respiratoria por tres o más ciclos, y apnea; la actividad electrodérmica - observada o medida por el incremento en la conductancia de la piel (decremento de resistencia), el incremento en la duración de la respuesta, y por respuestas múltiples; y la actividad cardiovascular - en forma de incremento de la presión sanguínea relativa, incremento en la duración de la respuesta, la desaceleración de la frecuencia cardíaca y la disminución en el volumen de pulso de sangre del dedo.

Las características de calificación se han descrito como primarias o secundarias (Bell, et al., 1999; Department of Defense, 2006a; 2006b). Las características primarias son aquellas que capturan el mayor grado de varianza entre las respuestas de engaño y de veracidad en las RQs y las CQs dentro en cada uno de los canales fisiológicos registrados. Las características secundarias están correlacionadas con las diferencias entre las respuestas de engaño y de veracidad a un nivel estadísticamente significativo, pero tienen coeficientes de correlación más débiles en comparación con las características primarias. Además, las características secundarias proporcionan información que se encuentran tan fuertemente correlacionada con sus contrapartes primarias, que la información adicional es en gran medida redundante y podría no ser un agregado a la eficiencia de algunos modelos estructurales. Algunos algoritmos de calificación computarizada (Honts y Devitt, 1992; Kircher y Raskin, 1988; Krapohl, 2002; Krapohl y McManus, 1999; MacLaren y Krapohl, 2003; Nelson, Krapohl y Handler 2008; Raskin et al., 1988;) en uso hoy en día, y el Sistema de Calificación Empírico basado en evidencia (Nelson & Handler, 2010; Nelson et al., 2011) han sido diseñados para utilizar sólo características primarias, renunciando a las características de reacción consideradas de importancia secundaria. Las características primarias, que a veces son denominadas como "Características Kircher" son las siguientes: respiración - observada por la longitud de excursión o patrones correlacionados, la actividad electrodérmica - observada por la amplitud del aumento

vertical, y la actividad cardiovascular – observada por la amplitud del aumento vertical de la presión sanguínea relativa.

Transformaciones numéricas. Se asignan puntuaciones numéricas, en forma de números enteros no paramétricos de valor positivo o negativo, a cada presentación de cada RQs mediante la comparación de la fuerza de reacción de cada RQ contra la fuerza de reacción en las CQs presentadas en la secuencia con las RQs. Un supuesto fundamental durante las pruebas con pregunta de comparación, es que tanto los examinados veraces como los que engañan pueden presentar algún grado de reacción a los estímulos de preguntas relevantes. De hecho, Ansley (1999), Ansley y Krapohl (2000) y Offe y Offe (2007) han demostrado empíricamente que no es la presencia o ausencia de una respuesta, ni la magnitud lineal de respuesta a las preguntas relevantes lo que discrimina entre engaño y veracidad. En lugar de esto, la magnitud relativa simple o el grado de respuesta a las CQs, en relación con el grado de respuesta a las RQs, es la característica que diferencia a los examinados que engañan de los veraces.

Los examinados que engañan, generalmente exhiben un cambio de mayor magnitud en la actividad autónoma en respuesta a estímulos relevantes que a los estímulos de comparación, mientras que los examinados veraces generalmente exhibirán cambios de mayor magnitud a los estímulos de comparación que a los estímulos relevantes. Los puntajes de engaño se asignan cuando la magnitud de los cambios ante los estímulos de las preguntas relevantes es mayor que a los estímulos de las preguntas de comparación. Por el contrario, las puntuaciones veraces se asignan cada vez que el grado de cambio en la respuesta ante los estímulos de pregunta de comparación es mayor que las respuestas ante los estímulos de pregunta relevante. Numerosas revisiones científicas de innumerables estudios científicos han afirmado la validez del constructo operacional de que las respuestas a los estímulos relevantes y de comparación varían en función del engaño o veracidad con respecto a un comportamiento pasado (American Polygraph Association, 2011; Ansley, 1983; 1990; Abrams, 1973; 1977; 1989; National Research Council/National Academy of Science, 2003; Nelson & Handler, 2013; Office of Technology Assessment, 1983; Podlesny & Raskin, 1978; Raskin, Honts & Kircher, 2014).

Puntajes de corte de decisión y distribuciones de referencia.

Los resultados numéricos de las pruebas se traducen en resultados categóricos de prueba, mediante la comparación de los puntajes de la prueba contra los puntajes de corte que están anclados a las distribuciones de referencia que describen la densidad estadística o probabilidad de obtener cada puntuación en particular dentro del rango de los puntajes posibles de las pruebas. Debido a que todos los datos científicos son una combinación tanto de varianza diagnóstica (es decir, varianza explicada) como de varianza no explicada (es decir, varianza de error, varianza aleatoria o varianza no controlada), se puede esperar que las puntuaciones individuales podrían variar de alguna manera dentro de cada examen. Por esta razón, se han encontrado que la sumatoria de los puntajes de prueba, proporcionan la mayor eficiencia diagnóstica. Esto se presenta a menudo en la forma de una puntuación de gran total, aunque los puntajes subtotales también se utilizan con algunas técnicas poligráficas. Los puntajes de gran total y los subtotales se comparan contra los puntajes de corte y contra las distribuciones de referencia estadística para determinar la probabilidad de que un puntaje de prueba observado ocurrió debido simplemente a la varianza de error no controlada o el azar.

Los puntajes de corte probabilísticos son una expresión de nuestra tolerancia a la incertidumbre o el error, expresado como una probabilidad estadística de error, a menudo utilizando la letra Griega, declarada antes de la conducción del examen. El puntaje de corte probabilístico común en poligrafía y en otras disciplinas científicas es .05, con el objetivo de limitar la proporción de errores al 5% o menos, al intentar proporcionar un nivel de confianza mínimo del 95% para el resultado categórico de prueba. Los límites de probabilidad alternativos de .10 e incluso .01, representan niveles de confianza previstos al 90% y 99% y a veces se utilizan cuando los objetivos de prueba indican una necesidad de un menor número de resultados de prueba inconclusos o no resueltos (.10) o para un menor número de errores (.01).

Los puntajes de corte también se determinaron utilizando las curvas de rendimiento (Bell et al., 1999) y mediante la experiencia heurística. Independientemente del método utilizado para determinar los puntajes de corte numéricos, todos los puntajes de corte de decisión tendrán alguna información estadística asociada para describir el nivel de significancia o probabilidad de error. La relación entre las puntuaciones numéricas y las distribuciones de referencia estadísticas asociadas se puede calcular matemáticamente, y también se pueden determinar convenientemente utilizando tablas de referencia publicadas.

Las distribuciones de referencia se han resumido (American Polygraph Association, 2011; Nelson & Handler 2015) en la forma de una estadística descriptiva que nos informa acerca de la ubicación (es decir, media o promedio), la dispersión (es decir, varianza o desviación estándar) y la forma de la distribución de los puntajes observados en los datos de muestreo para las personas con criterio de engaño o criterio de veracidad. Las distribuciones de referencia publicadas pueden ser utilizadas para calcular el margen de incertidumbre, en la forma de un nivel de significancia estadística, proporciones extrañas, niveles de confianza o probabilidad de error asociado con cualquier calificación de prueba posible. Se dice que los resultados de prueba son *estadísticamente significativos* cuando la probabilidad de error es menor o igual a un puntaje de corte con probabilidad declarada o nivel alfa (es decir, $p \leq$). Esto es equivalente a la condición cuando una calificación de prueba es igual o superior a un puntaje de corte.

Reglas de decisión.

Las reglas de decisión son la comparación práctica y metodológica de los resultados de prueba, ya sea con puntajes de corte tradicional Bell, *et al.*, 1999; Department of Defense, 2006a; 2006b; Kircher & Raskin, 1988; Raskin *et al.*, 1988) o con puntajes de corte seleccionados por su nivel de significancia estadística y probabilidad de error utilizando distribuciones de referencia publicados (American Polygraph Association, 2011, Krapohl & McManus, 1999; Krapohl, 2002; Nelson, Krapohl & Handler, 2008; Nelson & Handler, 2010; 2015; Nelson *et al.*, 2011). Procesalmente, después de la asignación de puntajes numéricos, todos los puntajes se agregan mediante la sumatoria de puntajes subtotaes para todas las presentaciones de cada estímulo RQ. Entonces, los puntajes subtotaes se suman para obtener una puntuación de gran-total para los exámenes diagnósticos de evento-específico. Las reglas de decisión procesales se construyen en relación con los supuestos de criterio de varianza independiente y no independiente de las RQs en los exámenes de diagnóstico de evento-específico y en los exámenes exploratorios multi-tema.

Las reglas de decisión procesales para los exámenes de diagnóstico de evento específico, para los que se asume que el criterio de varianza de las diferentes RQs no son independientes o son dependientes (es decir, todas las RQs se dirigen a un solo evento para el cual el estado de criterio de los diferentes estímulos de prueba podría estar fuertemente relacionado), hará uso de la puntuación de gran total para realizar la clasificación más precisa posible con respecto a la prueba como un todo. Algunas reglas de decisión, como las utilizadas por el Department of Defense (2006a; 2006b), según lo descrito por Light (1999), o aquellas desarrolladas por Senter y Dollins (2002) también podrían hacer uso de las puntuaciones subtotales en un intento por reducir los resultados inconclusos, incrementar la sensibilidad de la prueba hacia el engaño, o reducir los errores falsos-negativos. Los resultados de las decisiones que involucran el gran total, que aquí se refieren como regla de gran-total (GTR), se logran mediante la comparación de la puntuación de gran total con los puntajes de corte de las clasificaciones de engaño y veracidad. Cuando se utiliza la GTR, los resultados de las pruebas son estadísticamente significativos y se realiza la conclusión categórica si es que la puntuación de gran total es igual o superior a uno de los puntajes de corte.

Una modificación en dos-etapas de la GTR, referida aquí como la regla de dos etapas (TSR) fue descrita por Senter (2003) y por Senter y Dollins (2002; 2008). El TSR permite el uso de las puntuaciones subtotales para lograr una conclusión categórica cuando la puntuación de gran total no es estadísticamente significativa (es decir, inconclusa). Cuando se usan, las puntuaciones subtotales deben compararse con los puntajes de corte que están estadísticamente corregidos para la inflación conocida del alfa, y del incremento potencial asociado de errores falsos positivos, que resulta del uso comparaciones estadísticas múltiples (Abdi, 2007, Nelson y Handler, 2010; Nelson et al., 2011; Nelson, Krapohl y Handler, 2008). El uso de una corrección estadística simple, referida como la *corrección de Bonferroni*, puede evitar un incremento de errores falsos positivos cuando se utilizan puntuaciones subtotales de exámenes de diagnóstico de evento-específico.

Por definición, se asume que los estados de criterio de las RQs en exámenes exploratorios de asuntos-múltiples varían de forma independiente. Por esta razón, generalmente no se utilizan los puntajes de gran total con los exámenes exploratorios de asunto-múltiple, para los que con mayor frecuencia se utilizan los puntajes subtotales. Los puntajes para polígrafos exploratorios de asuntos-múltiples comúnmente se evalúan utilizando las puntuaciones subtotales individuales (Department of Defense, 2006a; 2006b; Nelson and Handler, 2010; Nelson *et al.*, 2011; Nelson, Krapohl & Handler, 2008), utilizando una regla que se refiere aquí como la regla de puntuación por subtotal (SSR). El SSR se ejecuta mediante la comparación de cada puntaje subtotal contra los puntajes de corte derivados de distribuciones estadísticas de los puntajes de los exámenes construidos por preguntas para las que se asume que el criterio de estado varía de forma independiente (American Polygraph Association, 2011; Nelson and Handler, 2010; Nelson *et al.*, 2011).

Aunque la SSR involucra el uso de puntuaciones subtotales individuales, investigaciones previas (Barland, Honts y Barger, 1989; Podlesney y Truselov, 1993; Raskin, Honts y Kircher, 2014; Raskin, Kircher, Honts y Horowitz, 1988) ha demostrado que aunque las pruebas de polígrafo con preguntas de comparación son eficientes para diferenciar entre individuos que son veraces o que engañan, estas pruebas no son tan eficientes para determinar la pregunta o las preguntas exactas dentro de una serie, a la que un individuo mintió o dijo la verdad. Las razones de esto, podrían tener

que ver tanto con las demandas psicológicas y de atención a los objetivos de estímulos independientes múltiples, y también por la complejidad matemática y estadística que resulta de la suma de las tasas de sensibilidad, especificidad, falsos positivos y falsos negativos de resultados múltiples independientes. Las preguntas de prueba de los exámenes de asuntos múltiples también comparten una fuente de varianza de respuesta, no independiente en la persona examinada. Por estas razones, la clasificación final de los resultados de examen, como pertenecientes ya sea al grupo de personas con engaño o con veracidad, siempre se determina al nivel de la prueba como un todo.

Cuando se utiliza el SSR, el resultado de la prueba se clasifica como de engaño, si *cualquier* pregunta independiente produce un resultado que sea estadísticamente significativo para el engaño, mientras que las clasificaciones veraces requieren que los resultados de *todas* las preguntas independientes sean estadísticamente significativas de veracidad. Las prácticas de campo (American Polygraph Association, 2009a; Department of Defense, 2006a; 2006b) no dan soporte a la interpretación de las respuestas en algunas preguntas como veraces y en otras respuestas como de engaño dentro de la misma examinación. Por supuesto, los métodos estadísticos relacionados con la regresión, la varianza y la covarianza podrían proporcionar capacidades no disponibles dentro de la rúbrica procesal simple de la RSS.

Idealmente, los puntajes de corte subtotal para la veracidad deberían determinarse utilizando procedimientos que corrijan estadísticamente la reducción potencial de la especificidad de la prueba cuando se requieren múltiples puntuaciones veraces estadísticamente significativas antes de realizar una clasificación de veracidad. Una solución común para esta corrección en las ciencias estadísticas y matemáticas, como se ha descrito en los procedimientos para la calificación del polígrafo (Nelson y Handler, 2010; Nelson et al, 2011), es la *corrección* Šidák, utilizada cuando se requieren eventos de probabilidad independiente estadísticamente significativos múltiples (Abdi, 2007). Algunos procedimientos involucran el uso de las puntuaciones subtotales con puntuaciones de corte tradicional que no se derivaron por distribuciones de referencia estadística, sino que se basan en las curvas de rendimiento de la clasificación o experiencia heurística (Bell *et al.*, 1999; Department of Defense, 2006a; 2006b).

Bases fisiológicas del polígrafo.

Aunque está más allá del alcance de este artículo, una descripción minuciosa y detallada de las respuestas fisiológicas registradas por el polígrafo y una descripción práctica de la fisiología del polígrafo está inextricablemente ligada a la necesidad de traducir los cambios en los datos registrados a una forma de puntaje y resultado de prueba. A diferencia de los modelos previos para el análisis de datos de prueba que dependían en gran medida del reconocimiento de patrones como un medio para el monitoreo y observación de la actividad fisiológica (Department of Defense, 2004), los modelos basados en evidencia en uso hoy en día, emplean solamente aquellas características fisiológicas que son susceptibles de medición y que han demostrado, a través de estudios científicos publicados, replicados y revisado por pares, que se correlaciona a niveles estadísticamente significativos con las diferencias en respuesta ante los diferentes tipos de estímulos de prueba que se producen en función del engaño o veracidad acerca de un comportamiento pasado (Bell, et al., 1999; Harris, Horner y McQuarrie, 2000; Kircher, Krisjansson, Gardner y Webb, 2005; Kircher y Raskin, 1988; Podlesny y Truslow, 1993, Raskin, Kircher, Honts y Horiwitz, 1988). La instrumentación de registro poligráfico ha tendido a centrarse en la adquisición de datos de la respuesta fisiológica que es de uso práctico

para la tarea de calificar e interpretar los resultados de la prueba poligráfica, con pocas capacidades más allá de este objetivo. Por ejemplo: la instrumentación del polígrafo no se utiliza para evaluar la salud cardiovascular o respiratoria.

La instrumentación poligráfica consta como mínimo de tres sensores componentes: dos sensores neumográficos (torácica y abdominal) para registrar la actividad de movimiento respiratorio; sensores eléctricos para registrar la actividad autónoma en las regiones palmar o distales (Handler, Nelson, Krapohl y Honts, 2010); y los sensores cardiovasculares que registran los cambios relativos en de la presión sanguínea (American Polygraph Association, 2009a, 2009b; Department of Defense 2006a). Los sensores vasomotores (Kircher y Raskin, 1988; Bell et al, 1999) se consideran como componentes opcionales del instrumento poligráfico. Desde el 2007, los protocolos de pruebas de campo, han recomendado para ayudar en la detección, del uso de sensores de actividad de contramedidas, y son ahora requeridos por la American Polygraph Association, (2011a) desde enero 1 del 2012. Esta combinación requerida de sensores principales ha sido estudiada por muchas décadas, y ha sido empíricamente demostrado que produce resultados numéricos que están estructuralmente correlacionados con las categorías de criterio de engaño y veracidad en distribuciones de referencia estadísticas de muestras desarrolladas y validadas tanto en estudios de campo como de laboratorio (American Polygraph Association, 2011; Bell, Raskin, Honts & Kircher, 1999; Harris & Olsen, 1994; Harris, Horner & McQuarrie, 2000; Horowitz, Kircher, Honts & Raskin, 1997; Kircher & Raskin, 1988; Kircher, Kristjansson, Gardner & Webb, 2005; MacLaren & Krapohl, 2003; Offe & Offe, 2007; Olsen, Harris & Chiu, 1994; Raskin, Kircher, Honts & Horowitz, 1988).

El mecanismo fisiológico de las respuestas poligráficas durante las pruebas con preguntas de comparación, ocurren en el contexto del sistema nervioso autónomo (ANS) que incluye componentes tanto simpático (S/ANS) como parasimpático (PS/ANS) (Bear, Barry, y Paradiso, 2007; Costanzo, 2007 Maton et al, 1993; Paradiso, Bear, & Connors, 2007; Silverthorn, 2009; Standring, 2005). El ANS regula los procesos involuntarios incluyendo el ritmo cardíaco, respiración, salivación, sudoración, y otras formas de excitación. La actividad del S/ANS es responsable de la estimulación de los órganos internos en respuesta de las demandas de actividad.

La actividad PS/ANS sirve para reducir la activación fisiológica al nivel mínimo necesario para garantizar tanto la longevidad como las respuestas adecuadas a demandas situacionales. La actividad del PS/ANS y la del S/ANS están por lo tanto, en equilibrio homeostático con respecto a las demandas reales o percibidas. La alternativa al equilibrio homeostático es un estado general de enfermedad que eventualmente podría causar la muerte. Por esta razón, cada forma de cambio en el ANS puede ser pensado como un intento de mantener la homeostasis y la supervivencia.

Los examinadores poligráficos, están principalmente interesados en la grabación y observación de la actividad del S/ANS, pero es importante entender que cierta actividad, como son algunos datos cardiovasculares y algunas respuestas respiratorias de interés para los poligrafistas, podrían ser en realidad el resultado de cambios en la actividad PS/ANS. (Para obtener más información acerca de la inervación dual del sistema nervioso autónomo se remite al lector a las obras más completas de Jänig (2006), Porges (2014), Handler y Richerter (2014)). El proceso de cambio, con el objetivo de mantener la homeostasis, es conocido como alostasis (Sterling y Eyer,

1988; Berntson y Cacioppo, 2007). Los cambios registrados durante la prueba poligráfica pueden considerarse como cambios alostáticos (Handler, Rovner y Nelson, 2008) que se producen en un intento de alcanzar o mantener la homeostasis.

Los cambios fisiológicos observables y medibles de la actividad fisiológica que están estructuralmente correlacionados con el engaño y la veracidad durante las pruebas con preguntas de comparación, incluyen las tres siguientes características: 1) supresión respiratoria temporal y sutil (ejemplo, la supresión o reducción de los movimientos respiratorios, 2) magnitud relativa o actividad electrodérmica, indicativa del incremento de la actividad del S/ANS, 3) magnitud relativa en el movimiento promedio de la presión sanguínea relativa. Estas reacciones medibles se han descrito en varias publicaciones (ASTM International, 2002; Bell *et al.*, 1999; Department of Defense, 2006a, 2006b; Harris, Horner & McQuarrie, 2000; Kircher & Raskin, 1988; Kircher *et al.*, 2005; Krapohl & McManus, 1999; Raskin & Hare, 1978; Raskin *et al.*, 1988). Las respuestas fisiológicas a estos tres sensores primarios (supresión respiratoria, actividad electrodérmica y actividad cardiovascular) son fácilmente observadas y registradas.

De entre los conceptos erróneos comunes acerca del polígrafo, se incluye la noción de que el polígrafo mide la respiración profunda o acelerada, palmas sudorosas o actividad de sudoración y actividad rápida o mayor en la frecuencia cardíaca. De estos, los dos primeros, el incremento en la actividad respiratoria y la actividad de la sudoración, se sabe que son modelos explicativos inexactos y poco satisfactorios de las reacciones de polígrafo. Sólo la tercera, la tasa cardíaca, se ha incluido como clasificador estadístico validado para el engaño y la veracidad. Sin embargo, es el *decremento* del ritmo cardíaco, y no el aumento, lo que se correlaciona con el engaño (Raskin y Hare, 1978; Kircher *et al.*, 2005). La frecuencia del pulso se incluye en el modelo estadístico algorítmico PolyScore (Blackwell, 1998; Dollins, Krapohl y Dutton, 1999; Dollins, Krapohl y Dutton, 2000; Harris & Olsen, 1994; Olsen *et al.*, 1991; Olsen *et al.*, 1994; Olsen, Harris, Capps y Ansley, 1997). Los cambios en la tasa cardíaca no están incluidos en otros modelos estadísticos validados de calificación para pruebas con preguntas comparativas. Por lo tanto, la actividad en la tasa del pulso rara vez se incluye en las decisiones poligráficas en ambientes de campo.

El análisis manual de la magnitud relativa o absoluta del cambio de respuesta en la actividad de la respuesta fásica, se logra fácilmente tanto en la actividad electrodérmica como en la actividad cardiovascular. Se ha demostrado que los datos electrodérmicos son el indicador más fuerte de excitación en el S/ANS (Boucsein, 2012), y que contribuyen de forma más robusta y confiable en la puntuación final y en la clasificación resultante en los resultados de pruebas poligráficas con preguntas comparativas Ansley & Krapohl, 2000; Harris & Olsen, 1994; Kircher, 1981; 1983; Kircher & Raskin, 2002; Kircher & Raskin, 1988; Kircher *et al.*, 2005; Krapohl & McManus, 1999; Nelson, Krapohl & Handler, 2008; Olsen *et al.*, 1997; Raskin *et al.*, 1988).

Las reacciones ante los estímulos de prueba pueden ser evaluadas a través tanto de la observación no-paramétrica, o mediante una medición lineal. Sin embargo, los fabricantes de instrumentos poligráficos no han estandarizado completamente el procesamiento de señales y los métodos de extracción de características, por lo cual los datos obtenidos durante la prueba poligráfica tienen que ser anclados a cambios lineales en la actividad fisiológica. Como resultado de esto, las mediciones de respuestas ante los estímulos poligráficos se utilizan sólo en modelos de clasificación

estadística automatizados dentro del paradigma de la prueba del polígrafo y podrían no estar relacionados directamente con las mismas mediciones de la actividad fisiológica utilizadas en el campo médico. Muchos paradigmas de calificación poligráfica serán métodos de extracción de características no paramétricas.

Las respuestas cardiovasculares durante las pruebas con preguntas comparación han demostrado estar correlacionadas con las categorías de criterio de engaño y veracidad a niveles estadísticamente significativos (Bell et al, 1999;. Harris et al, 2000;. Kircher y Raskin, 1988; Kircher et al ., 2005; Nelson et al, 2008; Raskin et al, 1988). La correlación estructural para la actividad de respuesta cardiovascular, ha demostrado ser más débil que los datos de la respuesta electrodérmica, aunque más fuerte que los datos de la respuesta respiratoria. Las características de diagnóstico y la interpretación de la actividad en la respuesta cardiovascular fueron descritas por Handler y Reicherter (2008) y Handler, Geddes y Reicherter (2007). Algunos examinadores poligráficos en campo hacen uso de los datos del pletismógrafo fotoeléctrico, una forma de registro cardiovascular cuya información ha sido descrita por Handler y Krapohl (2007), Geddes (1974) y Honts, Handler, Shaw y Gougler (2015).

De los tres sensores fisiológicos, los datos respiratorios han demostrado ser los más susceptibles a la desorganización por la actividad voluntaria durante la prueba poligráfica. Los datos respiratorios tienen los coeficientes estructurales más débiles de los sensores requeridos por el polígrafo (Harris & Olsen, 1994; Harris et al, 2000;. Kircher y Raskin, 1988; Kircher et al, 2005;. Nelson, y Krapohl Handler, 2008; Olsen et al, 1997; Raskin et al, 1988). Sin embargo, los examinadores de campo han aprendido a evaluar los datos respiratorios como indicadores de cooperación o no cooperación durante las pruebas, además de evaluar estos datos de respiración como indicadores de engaño y veracidad. Cierta investigación (Kircher *et al.*, 2005) sugirió que los datos pneumográficos son de menor diagnóstico durante la conducción de pruebas con preguntas comparativas utilizando exámenes DLC, aunque otros hallazgos han mostrado que los datos respiratorios del DLC contienen información diagnóstica útil (Honts & Handler, 2014).

La supresión respiratoria, aunque es medida de forma exacta por la distancia curvilínea (Kircher y Raskin, 1988; Raskin et al, 1988;. Timm, 1982) o por la suma de la magnitud absoluta del cambio en la excursión del eje-y (Kircher y Raskin, 2002), no es fácilmente medible sin dispositivos mecánicos. A los examinadores de campo se les enseña a evaluar los datos registrados por la presencia o ausencia de patrones de reacción que han sido descritos como correlacionados con las categorías de criterio de engaño y veracidad (Raskin y Hare, 1978; Bell et al, 1999;. Harris et al., 2000; Kircher y Raskin, 1988; Kircher et al, 2005;. Raskin et al, 1988). La características de patrones han demostrado que se correlacionan con la supresión respiratoria y las categorías criterio del CQT son pocas, e incluyen las siguientes: 1) una reducción sutil y temporal del volumen de corriente de inhalación que resulta en una reducción de la magnitud del eje Y (vertical) de los trazos respiratorios en ciclos respiratorios múltiples después de la aparición de la pregunta estímulo de prueba, 2) una disminución sutil y temporal de la tasa respiratoria por ciclos respiratorios múltiples después de la aparición de la pregunta estímulo de la prueba, y 3) una elevación sutil y temporal de la línea de base de exhalación o del volumen residual por ciclos respiratorios múltiples después de la aparición de estímulo. La Apnea también se encuentra correlacionada con las diferencias en engaño y veracidad (Bell *et al.*, 1999; Kircher & Raskin, 1988), pero puede ser fácilmente fingida.

Aunque los sensores del polígrafo, son capaces de registrar la respuesta simpática autónoma ante los estímulos de prueba, no son robustos frente a la actividad física o somática disruptiva, que a veces no se observa fácilmente. En respuesta a la preocupación acerca del potencial de intentos de fingimiento durante la prueba (por ejemplo, contramedidas), han sido desarrollados sensores de actividad somática. Hay indicios en la literatura de que los sensores de actividad somática pueden incrementar la habilidad del examinador para observar y detectar estos intentos (Ogilvie y Dutton, 2008; Stephenson & Barry, 1986). En la ausencia de datos grabados o alterados, de calidad extraña o no interpretable, indicativa de actividad física abierta o encubierta, los examinadores de campo supondrán que las respuestas registradas por los sensores de respiración, electrodérmicos y cardiovasculares tienen su origen en el ANS y no están alterados o contaminados por actividad encubierta.

Bases psicológicas del polígrafo.

Una teoría psicológica satisfactoria dará cuenta de forma parsimoniosa y holística de la variedad de fenómenos conocidos y observados asociados con la prueba poligráfica. Tal teoría explicará las respuestas electrodérmicas, las respuestas cardiovasculares y las respuestas respiratorias, tanto de las preguntas PLC como de las DLC, y contribuirá a nuestra comprensión de la exactitud de la prueba tanto con personas psicópatas y no psicópatas. Más aún, una comprensión sólida de las bases psicológicas de la prueba poligráfica nos permitirá entender mejor los problemas de idoneidad o falta de idoneidad a la prueba (es decir, para quienes la prueba podría o no funcionar). Un entendimiento teórico comprensible de las bases psicológicas de las respuestas a diferentes paradigmas de prueba como el CQT y otros paradigmas de detección de mentira poligráficos - como la prueba de información oculta (CIT) -, que utiliza señales fisiológicas de registro similares como base para cálculos ipsativos de la significancia estadística de las diferencias en las respuestas ante diferentes estímulos de prueba. Finalmente, una teoría psicológica satisfactoria de la prueba poligráfica, logrará una integración coherente de los conocimientos científicos del polígrafo con los conocimientos existentes en campos afines de la ciencia incluyendo la psicología cognitiva, social y del comportamiento, la psicofisiología, la teoría de la detección de señales, la teoría de la decisión, la teoría del aprendizaje estadístico entre otras.

Aunque una discusión comprensible de las bases psicológicas de la prueba poligráfica está más allá del alcance de este artículo, una breve explicación sostiene que las bases psicológicas de las respuestas ante los estímulos de la prueba poligráfica involucran una constelación de mecanismos psicológicos simples incluyendo la cognición, la emoción y el condicionamiento conductual (Handler & Nelson, 2007; Handler, Shaw y Gougler, 2010; Kahn, Nelson, y Handler, 2009; Senter, Weatherman, Krapohl, y Horvath, 2010). Recientemente, la teoría del proceso preliminar, relacionada con la teoría de la orientación (Barry, 1996) se ha sido sugerido como una explicación potencialmente parsimoniosa de las diferencias observadas en las respuesta ante los diferentes estímulos de prueba (Palmatier y Rovner, 2014), aunque se necesita una mayor discusión para entender completamente las ventajas y limitaciones de esta teoría aplicable a la prueba poligráfica. Hasta que se describa una evidencia más detallada, un constructo general sugiere que todas las respuestas a los estímulos de prueba son el resultado de una combinación de la actividad mental, emocional y de condicionamiento conductual. Todos estos podrían desempeñar un papel en las reacciones fisiológicas que son una carga diferencial para los diferentes tipos de estímulos de prueba poligráfica (es decir, preguntas

relevantes y de comparación) en función del engaño o veracidad en respuesta a los estímulos relevantes que describen el asunto conductual de preocupación. Será importante abstenerse en el intento de definir cuál es esa emoción individual, o definir el foco exacto de atención y cognición en el examinado hasta que exista evidencia que verifique una descripción más detallada.

Históricamente, los examinadores de campo han tendido a simplificar la explicación de la psicología del polígrafo a un nivel mínimo que los satisface a sí mismos y a sus evaluados. Esto se hizo utilizando a menudo la explicación no satisfactoria científicamente del "set psicológico" (ver Handler & Nelson, 2007) que se relaciona con la respuesta de ataque o huida que se ha atribuido a Cannon (1929). Aunque ahora se considera como un modelo inadecuado tanto para las respuestas poligráficas y como de las respuestas ante el estrés (Bracha, et al, 2004;.. Taylor et al, 2000), la aplicación de esta hipótesis problemática sostiene que los examinados centrarán su atención y respuesta fisiológica ante la pregunta o asunto que presente la mayor amenaza inmediata para su supervivencia y bienestar. La evidencia más obvia de las limitaciones de la hipótesis del "set psicológico" es que no puede dar cuenta de la efectividad de las DLC y no explica adecuadamente la eficacia de la prueba con psicópatas de quienes se ha demostrado que tienen bajos niveles de condicionamiento ante el miedo (Birbaumer, et al., 2005). Adicionalmente, el "set psicológico" requiere del supuesto de que los sensores de polígrafo puede identificar diferentes tipos de emociones, aunque la literatura no es compatible con esta noción (Kahn, Nelson, y Handler, 2009). Más aún, esta explicación sufre de una vulnerabilidad fundamental ante la sugerencia de que es psudocientífico ya que no puede satisfacer el requerimiento científico fundamental de la falsabilidad (Popper, 1959).

Handler y Nelson (2007) describieron los orígenes problemáticos del término "set psicológico", que no aparece en la literatura psicológica científica en la forma empleada por los examinadores de polígrafo. Se ha sugerido que la saliencia diferencial, es una teoría psicológica más completa, que es más consistente con el campo de la psicología científica, incluyendo la emoción, la cognición y el aprendizaje condicionado como base de respuesta a los estímulos del polígrafo (Senter, Weatherman, Krapohl y Horvath, 2010).

Las respuestas poligráficas también podrían ser explicadas mediante el uso del encuadre conceptual del condicionamiento conductual, descrito por primera vez por Pavlov (1927), y de la teoría del aprendizaje, incluyendo los conceptos de sensibilización y habituación (Domjan, 2010, Groves y Thompson, 1970). Un modelo del aprendizaje condicionado para la respuesta ante los estímulos del polígrafo sostiene que la participación en transgresiones graves se acumula en una forma de condicionamiento conductual de juicio-único, con preguntas de prueba que funcionan como estímulo condicionado. La teoría de la entrevista poligráfica sostiene que mediante una entrevista pre-test eficaz, se dará al examinado veraz la oportunidad de habituarse a las preguntas de la prueba, al tiempo que permite a un examinado con engaño sensibilizarse a las preguntas de prueba como estímulo condicionado.

La teoría cognitiva-conductual, que incluye la cognición, emoción y el aprendizaje conductual/experimental como base de la respuesta fisiológica, también se ha sugerido como hipótesis explicativa de la variedad de los fenómenos conocidos en el polígrafo (Kahn, Nelson & Handler, 2009), y este

modelo es consistente con la hipótesis de la saliencia descrita por Handler y Nelson (2007) y Senter et al., (2010). Una generalización del modelo cognitivo-conductual para las reacciones de polígrafo sugiere que la veracidad presenta una demanda de tareas cognitivas y emocionales más simples que el engaño.

Un modelo cognitivo-conductual y de saliencia diferencial sostendría que las respuestas fisiológicas ante una secuencia repetida de estímulos de prueba poligráfica se cargarán con diferentes tipos de estímulos de prueba en función del engaño y la veracidad con respecto a los asuntos objetivo la investigación, y que la base de las respuestas observadas pueden considerarse como generadas por la cognición, la memoria, la emoción y la experiencia condicionada en relación con los estímulos de la prueba. Las diferencias relativas en respuesta a los diferentes tipos de estímulos de prueba pueden ser comparadas con las distribuciones de referencia estadística y evaluadas por su nivel de significancia estadística para cuantificar el margen de incertidumbre con respecto a una conclusión categórica de engaño o veracidad. Esta forma de explicación teórica es fundamentalmente comprobable y por lo tanto fundamentalmente científica (Popper, 1959).

Precisión de exámenes poligráficos.

Los resultados de varias décadas de estudios científicos han soportado consistentemente la validez de la hipótesis de que la combinación de registros instrumentales y del modelado estadístico puede discriminar entre engaño y veracidad en tasas significativamente mayores que el azar. Las revisiones científicas de estudios de polígrafo revisados por pares han encontrado esto en repetidas ocasiones. Abrams (1989) estudió la literatura publicada y reportó un nivel de precisión de .89. Honts y Peterson (1997), Raskin (2002) y Raskin y Podlesny (1979) reportaron la precisión de los estudios del polígrafo, excediendo el .90. La revisión sistemática realizada por la Office of Technology Assessment (1983) sugirió que los estudios de laboratorio tienen una precisión media no ponderada de 0.83, con una precisión ligeramente mayor, .85, en comparación con estudios de campo de ese momento. Crewson (2001) reportó una tasa de exactitud de 0.88 en polígrafos diagnósticos en comparación con las pruebas médicas y psicológicas. El National Research Council (2003) concluyó con reserva que el polígrafo diferencia el engaño de la veracidad en tasas significativamente mayores que el azar, aunque menores a la perfección y reportó una mediana ROC de .89 para estudios de campo y de .86 para estudios de laboratorio.

Los diferentes tipos de estudios ofrecen diferentes ventajas y desventajas. Los estudios de campo ofrecen validez ecológica asumida, pero se acompañan de una falta de control experimental, por una confirmación inconsistente de casos y por una selección no aleatoria de casos – realizando una generalización problemática o imposible de algunos resultados de estos estudios de campo-. Los estudios de laboratorio ofrecen el potencial de un muestreo aleatorio y un control experimental suficiente para estudiar cuestiones de causalidad, pero teniendo una validez ecológica desconocida. La tendencia general en la investigación psicológica ha sido un alto grado de correspondencia entre estudios de campo y de laboratorio (Anderson, Lindsay y Bushman, 1999), y este hecho subraya la necesidad de evitar el confundir la validez ecológica con la validez externa.

La validez externa y la validez ecológica no son sinónimas. La validez externa - capacidad de generalizar los resultados de los ambientes de campo - a menudo se logra mediante estudios científicos en ambientes de laboratorio con validez ecológica imperfecta. Estudios previos realizados por la Office of Technology Assessment (1983), the National Research Council (2003), the American Polygraph Association (2011), y Pollina *et al.*, (2004) no mostraron diferencias estadísticamente significativas entre la exactitud de los resultados de la prueba poligráfica de estudios de campo y de laboratorio.

La revisión científica más reciente de las técnicas de preguntas de comparación poligráficas en uso actual (American Polygraph Association, 2011) reportó una precisión media de .89 para polígrafos diagnósticos de evento-específico, con algunos métodos basados en evidencia que han mostrado que generan una precisión media con niveles que exceden el .90. Los polígrafo de asuntos-múltiples, del tipo utilizado en seguridad operacional, pre-empleo en para procuración de la ley y los programas exploratorios post-convicción, han mostrado tener una tasa de precisión media de .85. Más importante que las estadísticas de precisión media son los rangos de confianza al 95% que rodean esos puntajes estimados, descritos más adelante en este reporte, especialmente en los límites-bajos de precisión de la prueba.

Las revisiones científicas publicadas previamente (Abrams, 1973; Ansley, 1983; Ansley, 1990) han reportado tasas más altas de precisión de prueba, a menudo arriba de los 90s. Se piensa ahora que los resultados de estas revisiones sistemáticas previas se conformaron por metodologías de muestreo que pueden haber sobre-enfaticado el auto-reporte del examinador, posiblemente sin acceso del investigador a los datos fisiológicos registrados o a las calificaciones numéricas, y se pudo haber exagerado el uso de la información de la confesión como criterio de confirmación y selección de caso. Potencialmente, estos factores pueden incorporar criterios de selección no aleatorios y no representativos de casos que pueden excluir sistemáticamente tanto los casos con error falso negativos y falsos positivos para los que es probable que no se obtuvo una confesión. La investigación en defensa de técnicas poligráficas de propiedad parece que también ha dado lugar a la producción de estimaciones de precisión exageradas - a menudo cercanas a la perfección -, algunas veces involucrando al investigador principal como examinador, calificador y desarrollador/propietario de la técnica. Estudios individuales que reportan una precisión casi-perfecta han descrito defectos metodológicos severos (American Polygraph Association, 2011).

Como en todas las formas de prueba científica, las pruebas de diagnóstico conducidas con respecto a un solo asunto de preocupación, para el que existe evidencia de un problema, proporcionarán una mayor precisión global que las pruebas exploratorias (American Polygraph Association, 2011; Crewson, 2001), que intentan evaluar simultáneamente varios asuntos para los que se asume que el criterio de varianza es independiente (es decir, una persona podría mentir a una o más de las preguntas de investigación al tiempo que no miente a otros objetivos de investigación). Las exámenes de asuntos-múltiples incluyen más decisiones probabilísticas y estadísticas. Otras causas de las diferencias de precisión entre los exámenes diagnósticos y exploratorios podrían involucrar las demandas de atención que compiten entre múltiples estímulos objetivos de prueba y el potencial de que los formatos de exámenes exploratorios pueden ocasionalmente estar sistemáticamente sesgados por la sensibilidad de la prueba - con la meta de sobre-predecir ligeramente los problemas que podrían ser resueltos mediante investigaciones posteriores - . Es

también posible que algunos estudios exploratorios fueran completados utilizando reglas de decisión sub-óptimas y puntajes de corte que no fueron derivados mediante análisis científico.

La tasa de sensibilidad agregada para el engaño en polígrafos diagnósticos fue reportado al .84 (95% de rango de confianza de .73 al .93) y las tasas de especificidad agregadas de la veracidad en polígrafos diagnósticos fue reportado al .77 (95% de rango de confianza de .65 a .85) (American Polygraph Association, 2011). La evidencia indica que las personas con engaño tienen una probabilidad estadística significativamente mayor de fallar una prueba de polígrafo mientras que las personas veraces tienen una probabilidad estadísticamente significativamente mayor de pasar un polígrafo.

Las estimaciones de error agregadas para las pruebas de polígrafo de diagnóstico fueron calculadas en el estudio meta-analítico más reciente que utilizó 24 estudios científicos revisados por pares que incluían 8,975 calificaciones confirmadas de estudios de campo y laboratorio, reportaron un .08 de errores falsos negativos (es decir, personas no veraces que aprobaron el polígrafo) y .12 de errores falso positivos (es decir, personas veraces que no pasaron el polígrafo). Las tasas de inconclusos se reportaron con un .09 para personas con engaño y .13 para personas veraces. El 95% de rango de confianza para la precisión de la decisión de polígrafos de diagnóstico de evento-específico fue de .83 a .95 (American Polygraph Association, 2011). Es obvio, que las desviaciones de los protocolos de prueba validados empíricamente podrían decrementar la precisión de prueba esperada, y por supuesto, estos estimados de precisión asumen que cada examinación se conduce en un examinado idóneo.

La tasa de sensibilidad agregada para el engaño durante polígrafos de asunto-múltiple, del tipo empleado en los programas de polígrafo exploratorio, fue reportado al .77 (95% de rango de confianza de .60 a .90) y las tasas de especificidad agregadas para veracidad en polígrafos de temas-múltiples con criterio de independencia fue reportado al .72 (95% de rango de confianza del .63 al .81) (American Polygraph Association, 2011). La evidencia indica que las personas con engaño, tienen una probabilidad estadísticamente significativa mayor que el azar de fallar una prueba construida con preguntas de asuntos múltiples independientes, mientras que las personas veraces tienen una probabilidad estadísticamente significativamente mayor que de pasar un polígrafo construido de manera similar.

Las estimaciones de error agregadas para las pruebas de polígrafo construidas a partir de preguntas de prueba para las que se asume un criterio de varianza independiente, fueron calculadas en 14 estudios revisados por pares que involucraron 1,194 calificaciones confirmadas en campo y laboratorio y fueron reportados con .11 para errores falsos negativos y .14 para errores falso positivos. Las tasas de inconclusos fueron también reportadas al .11 en personas no veraces y al .14 en personas veraces. El 95% de intervalo de confianza de la tasa de precisión del promedio no ponderado para examinaciones poligráficas construidas con preguntas de prueba para las que se asume que el criterio de varianza de las preguntas relevantes es independiente fue reportado con .77 al .93 (American Polygraph Association, 2011).

Para los polígrafos de diagnóstico, el Valor Predictivo Positivo (PPV), o probabilidad de que un resultado no aprobatorio de polígrafo sea correcto, se reportó al .89 (95% de rango de confianza de .81 a .99) mientras que el Valor Predictivo Negativo (NPV) o probabilidad de que el resultado de polígrafo aprobatorio sea correcto, fue reportado al .91 (95% de rango de confianza del .82 al .99) (American Polygraph Association, 2011). Para los polígrafos exploratorios, el PPV fue reportado al .83 (95% de rango de confianza de .71 a .94), mientras el NPV fue reportado al .88 (95% de rango de confianza de .78 a .97).

Los juicios conservadores requieren de la selección del extremo inferior del límite de confianza como el límite en el que se puede confiar que la precisión del polígrafo será superada. Por lo tanto se puede suponer que los polígrafos diagnósticos proporcionan una precisión superior al .81 para los resultados de engaño, y superior al .82 para los resultados veraces, mientras que puede asumirse que los polígrafos exploratorios proporcionan una precisión superior a .71 para los resultados de engaño y superior a .88 para los resultados veraces. Sin embargo, tanto el PPV y el NPV, no son resistentes a las diferencias en las tasas base, y los resultados reportados en el presente documento aplican solamente con exámenes poligráficos con grupos balanceados. Las estimaciones inferenciales de la precisión de la prueba (por ejemplo, sensibilidad, especificidad, inconclusos y tasas de error) son resistentes a las diferencias en las tasas base y pueden ser más útiles cuando se interpreta el significado del resultado de un examen único, como lo hace una corte en la evaluación de casos individuales.

Amenazas para la exactitud del polígrafo.

Debido a que las pruebas poligráficas – como todas las pruebas – son inherentemente probabilísticas (es decir, no existe observación determinística ni medición física), no son perfectas. La prueba no probabilística es inmune a un error potencial o a las amenazas a la precisión de la prueba. Aunque el National Research Council, (2003) no encontró evidencia científica alguna de que el tipo de personalidad o factores endógenos afectan significativamente la exactitud de la prueba de polígrafo, se entiende comúnmente que la precisión de la prueba de polígrafo puede verse comprometida o reducida por la salud, el nivel de funcionamiento o idoneidad del examinado.

Abrams (1975) mostró que la precisión de la prueba de polígrafo se redujo significativamente por el nivel de madurez funcional en delinquentes juveniles. En otros estudios, Abrams y Weinstein (1974) mostraron que no se puede esperar que el polígrafo sea preciso en personas que tienen diagnóstico de salud mental crónico dentro del espectro del trastorno psicótico, y mostró posteriormente que la precisión del polígrafo es inestable en personas cuyas habilidades intelectuales se encuentran por debajo del nivel límite del rango normal. (Abrams, 1974)

Aunque los problemas del desarrollo, el bajo funcionamiento intelectual, la baja madurez funcional, y la psicosis pueden afectar adversamente la precisión del polígrafo, no existe evidencia de que los aspectos de la personalidad psicopática puedan afectar adversamente la precisión de la prueba poligráfica. Barland y Raskin (1975) estudiaron sospechosos criminales con altos puntajes de

desviación psicopática en la prueba MMPI y no se encontraron diferencias significativas en la capacidad para detectar el engaño. Patrick y Lacono, (1989) también mostraron diferencias no significativas en la detección del engaño entre reclusos psicópatas y no psicópatas. Raskin y Hare (1978) reportaron la misma conclusión con una muestra diferente de sujetos reclusos. Balloun y Holmes (1979) mostraron que la precisión de la prueba utilizando un paradigma de prueba de conocimiento de culpabilidad no fue significativamente diferente para estudiantes universitarios con puntuaciones altas y bajas en desviación psicopática en la prueba MMPI.

Aunque tanto la Oficina de Technology Assessment (1983) y el National Research Council (2003), expresaron su preocupación por la idea de que la precisión de la prueba poligráfica podría ser menor para las personas con perfiles de personalidad peligrosa, ambos reportaron que la evidencia científica publicada no apoya, y consistentemente refuta, la hipótesis de que los psicópatas creen sus mentiras, y que por lo tanto pueden derrotar la prueba poligráfica. En resumen, se puede suponer que la prueba del polígrafo con personas psicopáticas es similar - tan exacta como inexacta - como lo es con las personas no psicopáticas. En cualquier caso, las reacciones del público y los medios de comunicación tienden a asumir de manera simplista que una persona ha "derrotado" el polígrafo cuando se observa un error de la prueba. No deberían de sorprender cierta proporción de errores de prueba a menos que esta proporción de errores se muestre excediendo el intervalo de confianza del 95% para las tasas de error normalmente esperadas.

En respuesta a las preocupaciones de que los exámenes conducidos bajo circunstancias *amistosas*, como las llevadas a cabo bajo el privilegio abogado-cliente, tienen menor validez que los realizados por los examinadores de procuración de justicia, Honts y Peterson (1997) describieron la lógica defectuosa y la dependencia en una hipótesis falsa como la base de esta preocupación, y que se resumió en los hallazgos reportados por Honts (1997) quien investigó la hipótesis mediante la lógica, análisis de caso, y meta-análisis. En la actualidad no existe una base de evidencia para apoyar, y la evidencia disponible contradice, la noción de que los exámenes realizados bajo el privilegio abogado-cliente ofrecen reducción de la precisión o validez. Estos resultados subrayan el valor de los métodos cuantitativos estructurados y la importancia de la objetividad y reproducibilidad cuando se analizan datos de prueba.

Una última preocupación respecto a la exactitud de la prueba de polígrafo involucra la posibilidad de que las contramedidas (es decir, falsificación) podrían ser eficaces para alterar el resultado de la prueba. Esta hipótesis representa preocupación importante para los programas de seguridad operacional del gobierno de los Estados Unidos, así como para el pre-empleo exploratorio en procuración de la ley, y en la supervisión posterior a la condena de las personas en libertad condicional. A pesar de la importancia de esta preocupación, solo han sido publicados un pequeño número de estudios sobre este tópico de la falsificación y la precisión del polígrafo.

Rovner (1979, 1986) y Rovner, Raskin y Kircher (1979) mostraron que el tener acceso a información acerca de la técnica poligráfica era insuficiente para alterar significativamente la precisión de la prueba. Un hallazgo preocupante en estos estudios fue que los sujetos veraces que trataron de emplear contramedidas, en un intento de aumentar la seguridad de pasar la prueba, de

hecho aumentaron su probabilidad de ser clasificados con engaño. Estos hallazgos llevaron al National Research Council (2003) a concluir que no era aconsejable el uso de contramedidas por parte de examinados veraces.

Timm (1991) reportó que la sugestión post-hipnótica es ineficiente como contramedida poligráfica, mientras Ben-Shakhar y Dolev (1996), junto con Elaad y Ben-Shakhar (1991) sugirieron que los esfuerzos mentales podrían tener un efecto principalmente en el canal electrodérmico. Los estudios de Iacono, Boisvenu y Fleming (1984) y Iacono, Cerri, Patrick y Fleming (1992) mostraron que tanto las benzodiazepinas como los medicamentos estimulantes son ineficientes como contramedidas, aunque un estudio previo de Waid, Orne y Orne (1981) indicó que el meprobamato podría afectar negativamente los resultados del polígrafo. Una limitación inherente para nuestra adquisición adicional de conocimientos en esta área sería la obtención de una aprobación del comité de ética para explorar plenamente los efectos de drogas o medicamentos psiquiátricos en los resultados de pruebas de polígrafo lo que sería difícil o improbable.

La preocupación relativa a las contramedidas poligráficas han sido descritos por Honts (1987), Honts, Amato & Gordon, (2004); Honts and Hodes (1983), Honts, Hodes and Raskin (1985), Honts, Raskin and Kircher (1987), and by Honts, Raskin, Kircher and Hodes (1988) cuyo trabajo colectivo comenzó a sugerir que las personas expertas en polígrafo no son tan eficientes como dicen para diferenciar el uso de contramedidas de entre otros artefactos. Además Honts et al., (1988) y Honts y Reavy (2009) encontraron que el uso espontáneo de contramedidas no era raro tanto entre los examinados veraces y no veraces. Raskin y Kircher (1990) reportaron que el entrenamiento en contramedidas físicas puede reducir la precisión de la prueba poligráfica y Honts, Raskin y Kircher, (1994) reportaron a las contramedidas físicas y mentales como igualmente efectivas. En un estudio diferente, Honts, Winbush y Devitt (1994) reportaron que las contramedidas mentales se pueden utilizar para derrotar pruebas de conocimiento de culpabilidad. En otro estudio, Honts y Amato (2001) nuevamente reportaron que las contramedidas intentadas por sujetos veraces resultaron nuevamente en una mayor producción de resultados de pruebas de engaño. Los sensores de actividad somática y los procedimientos de prueba destinados a esclarecer contramedidas tanto mentales como físicas, serían utilizados más ampliamente siguiendo estos estudios, y es necesaria la réplica de estos estudios utilizando una instrumentación y metodología de prueba contemporánea.

Los sensores de actividad están diseñados para ser sensibles a la actividad del sistema nervioso somático/comportamiento sin dejar de ser robustos frente al registro de los efectos de la actividad del ANS de interés para los modelos analíticos de datos de la prueba poligráfica. La justificación del uso de sensores de actividad involucra el hecho de que los sensores componentes del polígrafo, al tiempo que intentan ser sensibles a la actividad simpática del sistema nervioso autónomo, no son robustos también para registrar también los efectos de la actividad somática. Stephenson y Barry, (1986) junto con Ogilvie y Dutton, (2008) demostraron que al agregar un sensor de actividad puede incrementar la detección de actividad somática, y que esto podría reducir la aparición de acusaciones falsas sobre el uso de contramedidas. Se asume que la actividad observable indica que los datos poligráficamente grabados, probablemente son un compuesto adulterado tanto de actividad autónoma y somática. Correspondientemente, la ausencia de actividad somática indicaría que los datos autónomos grabados es más probable que no estén alterados y sean auténticos.

Los métodos estadísticos aún no han sido ampliamente explotados para la detección de contramedidas. Sin embargo, el algoritmo OSS-3 (Nelson et al., 2008), incluye un requisito procesal para revisar los datos por su calidad interpretable y marcar los segmentos de datos que están alterados por movimientos o por otra actividad problemática. El algoritmo OSS-3 calculará el número y la ubicación de los eventos indicados como artefactos y calculará la probabilidad estadística de que los artefactos observados hubieran ocurrido debido a causas aleatorias. El algoritmo alertará al examinador de la posibilidad de que el examinado pudo tratar de alterar sistemáticamente o de forma intencionada los datos fisiológicos registrados siempre que la probabilidad de falsedad esté debajo de los límites de un alfa establecida para ser significativos estadísticamente. Los métodos estadísticos aún no han sido estudiados exhaustivamente, se requieren investigaciones adicionales con respecto a la aplicación de métodos estadísticos y computacionales para detectar intentos de contramedidas.

Se han descrito estrategias que involucran tanto a la actividad muscular/física encubierta como las actividades mentales. Una estrategia alternativa de contramedida social, será el intento de convencer al examinador para que ignore los indicadores registrados por engaño como el resultado de una causa alternativa.

Otras formas de contramedidas potenciales involucran el uso de medicamentos o drogas, privación del sueño o cansancio físico, y el uso de esfuerzos mentales como la hipnosis, la meditación o la actividad mental. En general, se puede esperar que las contramedidas poligráficas intenten ya sea moderar o exagerar las respuestas a todo el conjunto de preguntas de prueba, lo que resultará en un aumento de la probabilidad de un resultado inconcluso en la prueba. Alternativamente, los examinados podrían intentar alterar estratégicamente las respuestas a los estímulos de prueba relevantes o comparativos. Debido a que es poco aconsejable exagerar las respuestas a los estímulos relevantes, y debido a que la supresión confiable de las respuestas en sólo algunas de las preguntas de la prueba presenta complicaciones no triviales para el examinado, puede ser más probable que las estrategias de contramedidas poligráficas involucren intentos ya sea de moderar o alterar la prueba en su totalidad, o el de aumentar o incrementar las respuestas a los estímulos de preguntas de comparación.

El objetivo de las contramedidas o intentos de fingimiento, en términos del análisis de los datos de prueba, es el de alterar sustancialmente la varianza de diagnóstico y de error contenido en los datos registrados de manera tal que no sean interpretables o de que no exista correlación entre las diferencias de estado de criterio de engaño o veracidad y entre las diferencias de reacción que ocurren en respuesta a los diferentes tipos de preguntas de estímulo de prueba. Esto produciría un resultado de prueba inconcluso ya sea porque los datos no son interpretables o no son estadísticamente significativos.

El objetivo de una contramedida sofisticada se propondría alterar los datos de prueba registrados de manera tal que los métodos matemáticos, estadísticos y computacionales de partición y comparación de las fuentes de varianza, resultaran en una inversión directa de la valencia de los

coeficientes de correlación para los criterios de estado de engaño y veracidad. Los intentos de contramedidas exitosos de este tipo requerirían también que la adulteración tanto de la varianza de diagnóstico y de error en los datos fisiológicos registrados se lleven a cabo de manera que convengan a los examinadores entrenados de la calidad auténtica de los datos registrados. La dificultad inherente de este desafío se hace más difícil por el hecho de que las pruebas poligráficas, como en otras formas de prueba, pueden incluir mecanismos y procedimientos analíticos diseñados para cuantificar la probabilidad de que una persona ha intentado entablar contramedidas.

En respuesta a la complejidad de estas cuestiones, las afirmaciones y las conclusiones alrededor del potencial de las contramedidas durante la prueba del polígrafo, el National Research Council (2003) escribió lo siguiente:

"Debido a que la aplicación efectiva de las contramedidas mentales o físicas por parte de los examinados requiere de la habilidad para distinguir entre las preguntas relevantes y de comparación, habilidad para regular la respuesta fisiológica, y la habilidad para ocultar las contramedidas de los examinadores capacitados, las pretensiones de que es fácil de entrenar examinados para "vencer" tanto al polígrafo como a los examinadores capacitados requieren evidencia de soporte científico para ser creíbles. Sin embargo, no tenemos conocimiento de cualquier tipo de tal investigación." (P.147).

La revisión bibliográfica del National Research Council (2003) no pudo presentar evidencias que apoyaran la hipótesis de que las contramedidas poligráficas e intentos de falsificación son efectivas para ayudar a las personas con engaño a derrotar los polígrafos con preguntas de comparación como actualmente se utilizan en ambientes de campo. Sin embargo, esto no debe interpretarse como un soporte a la infalibilidad del polígrafo o como una afirmación de que nadie ha pasado alguna vez la prueba del polígrafo con error.

Aunque todos los paradigmas de prueba podrían tener algún potencial de vulnerabilidad para ser explotado, los datos en este momento sugieren que estos esfuerzos sistemáticos para alterar el resultado de la prueba no están bien sustentados por la evidencia. Tampoco lo son las afirmaciones de que la prueba poligráfica es infalible. Los resultados de las pruebas poligráficas siguen siendo un estimado probabilístico del grado de incertidumbre que rodea a una conclusión categórica. Lo que se sabe en este momento es que prácticamente todas las personas culpables o con engaño que aceptan someterse a la prueba del polígrafo podrían tratar de involucrarse en algún tipo de actividad con intento de lograr un resultado de prueba negativo.

El polígrafo ha demostrado ser una herramienta eficaz, aunque imperfecta, para discriminar engaño y veracidad, incluso en estudios de campo que incluyen personas sospechosas de crímenes reales, a veces graves - algunos de los cuales se puede suponer que han intentado involucrarse en algún tipo de contramedida para intentar pasar la prueba cuando están mintiendo. Sin embargo, aún un potencial remoto del uso de una contramedida eficaz - y la relación de esta con la precisión del polígrafo en la seguridad operacional gubernamental, pre-empleo en fuerzas del orden, y en la supervisión posterior a la condena de transgresores convictos en entornos comunitarios - significará

que la interacción entre los intentos de contramedidas y la autenticidad de los datos de pruebas registrados seguirá siendo un área importante de preocupación. Independientemente de la eficiencia o ineficiencia de las contramedidas y de los intentos de fingimiento, se necesita investigación adicional y continua para comprender de mejor manera la vulnerabilidad a las contramedidas en los procedimientos de prueba de polígrafo contemporáneos.

Contribución de los resultados del polígrafo a las decisiones profesionales.

En un intento de cuantificar el valor y contribución de los resultados de polígrafo para los juicios profesionales, Honts y Schweinle (2009) describieron el uso del Índice de Ganancia de Información (IGI, Wells & Olson, 2002) en polígrafos diagnósticos y exploratorios. IGI es una medición del aumento en la precisión de la decisión que resulta del método, y ofrece la ventaja de proporcionar información a través del espectro previo a la tasa-base, - algo diferente que la métrica estadística falló en realizar eficientemente. Este método compara el incremento en la precisión de la decisión proporcionado por los resultados poligráficos, en comparación con juicios sin apoyo profesional - Vrij mostró que los oficiales de policía alcanzan una precisión de decisión del 56% cuando intentan determinar engaño o veracidad. La importancia de la estadística IGI es que provee información acerca del incremento en la precisión de la decisión a través de un completo rango de tasas-base posibles.

Honts y Schweinle (2009) demostraron que la precisión de decisión del diagnóstico de polígrafo aumenta de forma estadísticamente significativa tanto en examinados con engaño como en veraces, alcanzando un máximo de 27 veces de precisión en comparación con decisiones no asistidas, bajo el supuesto de condiciones de tasa-base alta, con incrementos estadísticamente significativos de precisión de decisión, en comparación con la opinión de expertos sin ayuda desde la tasa-base de .01 hasta .97 para conclusiones de engaño y de .03 hasta .99 para conclusiones de veracidad.

Las pruebas exploratorias, normalmente aplicadas bajo condiciones de tasa-base bajas, muestran un incremento estadísticamente significativo en la precisión de decisión con respecto a las decisiones de engaño, pero ningún incremento significativo en la precisión de las decisiones con respecto a las decisiones de veracidad. El incremento en la precisión de la decisión para los resultados de engaño fue estadísticamente significativa, en comparación con la detección de mentiras no-asistida, desde la tasa base de .02 hasta .83, y fue estadísticamente significativa para los resultados de veracidad desde la tasa base de .1 hasta .99.

Handler, Honts and Nelson (2013) evaluaron posteriormente la estadística IGI utilizando pruebas poligráficas exploratorias del tipo utilizadas en seguridad operacional, exploratorias de procuración de la ley, y programas de supervisión de trasgresores sexuales post-condena. Handler *et al.*, mostraron incrementos estadísticamente significativos en la precisión de las decisiones exploratorias, en comparación con la detección de mentiras no asistida, desde las tasas base a .01 hasta .94 para las detecciones de engaño y desde las tasas-base de .07 hasta .99 para las decisiones de veracidad. Esto sugiere que la efectividad del uso de la prueba poligráfica, aunque probabilística e imperfecta, tiene el potencial de incrementar la efectividad de la toma de decisiones profesional.

Conclusiones

Existe evidencia para apoyar la validez científica de las pruebas poligráficas tanto en el contexto diagnóstico como exploratorio, y existe evidencia suficiente para garantizar el interés continuo tanto de investigación como práctico en la discriminación instrumental y estadística del engaño y la veracidad tanto en programas forenses como exploratorios. La evidencia disponible puede describir y dar cuenta virtualmente de todos los aspectos de la prueba poligráfica, incluyendo la teoría de prueba, procedimientos de prueba, teoría de la decisión, y teoría de detección de señales, precisión de la prueba, y vulnerabilidad ante contramedidas o fingimiento, junto con las bases psicológicas y fisiológicas de las respuestas ante los estímulos del polígrafo.

La prueba poligráfica, al igual que otras pruebas científicas, es una prueba probabilística que involucra el registro de respuestas fisiológicas ante estímulos y utiliza la teoría de la decisión estadística para cuantificar el margen de error o el nivel de significancia estadística - o, alternativamente, lo extraño o el nivel de nivel confianza - asociado con los resultados de prueba (Nelson, 2014A; 2014b; 2014c, 2014d; 2014e). La prueba poligráfica logra sus objetivos a través de la combinación estructural de respuestas fisiológicas que han demostrado ser proxies confiables que están correlacionadas a niveles estadísticamente significativos con las diferencias en las respuestas que se cargan en diferentes tipos de estímulos de prueba (es decir, RQ y CQS) como una función del engaño y la veracidad con respecto a un comportamiento pasado.

La necesidad de una prueba que pueda discriminar entre el engaño y la veracidad surge del hecho de que podría no existir una evidencia, o podría no haberse descubierto aún para permitir una conclusión determinista o una medición física. Una prueba probabilística del engaño - con una precisión significativamente superior al azar y a la detección de mentiras no asistida - es la alternativa científica a las tasas de precisión cercanas al azar de la inferencia humana sin ayuda. Se necesitan pruebas científicas cuando no es posible realizar una observación determinista perfecta.

A menudo se necesitan pruebas para llegar a conclusiones informadas acerca de eventos en el pasado, que ya no pueden ser observados directa o deterministamente, y también para entender el potencial de eventos futuros que aún no han ocurrido y que por lo tanto no se pueden observar directa o deterministamente. Los resultados de la prueba poligráfica se refieren tanto a la posibilidad de que hubiera ocurrido un comportamiento en el pasado, y del futuro potencial de que la información o la evidencia sea descubierta para confirmar una conclusión.

También se necesitan pruebas científicas cuando existe el deseo de medir un fenómeno amorfo que no puede ser sometido a una medición física. No son necesarias las pruebas cuando es posible una medición mecánica o lineal directa; en cuyo caso simplemente se mide el elemento de interés. La medición, en comparación con las pruebas, involucra un error de medición mecánica. Las pruebas de cualquier tipo pueden involucrar fuentes humanas de varianza y otras fuentes aleatorias o no controladas de error de varianza, además de la varianza de diagnóstico contenida y expresada por los

datos de la prueba.

Todos los datos de prueba, incluyendo los datos de prueba poligráfica, son una combinación de la varianza de diagnóstico (también conocida como varianza explicada, varianza controlada, o señal) y la varianza de error (también conocida como la varianza aleatoria, varianza no explicada, varianza no controlada, o ruido). Idealmente, los datos de prueba incluyen una gran porción de varianza de diagnóstico y una pequeña porción de varianza de error, pero ninguna prueba es perfecta. Se espera que las pruebas científicas solamente cuantifiquen y contabilicen el margen de error que rodea a un resultado de prueba, y que den cuenta de las bases de los supuestos relacionados con los procedimientos de prueba.

Debido a que las pruebas científicas se utilizan para evaluar fenómenos amorfos que no pueden ser sometidos a una observación determinista o medición física, todos los resultados de la prueba son probabilísticos - incluso cuando son simplificados en resultados de prueba categóricos. La cuantificación del margen de error o del nivel de significancia estadística asociada con un resultado de prueba permitirá a los profesionales de referencia y a los consumidores de los resultados de las pruebas poder establecer conclusiones mejor informadas acerca del significado y de la utilidad de un resultado de prueba.

Las preocupaciones acerca de la ética de la prueba poligráfica, y especialmente de los programas de polígrafo exploratorios, algunas veces han apuntado hacia la falta de perfección y a la tasa de error de falsos positivos como el argumento base en contra del uso del polígrafo. Las expectativas de la perfección determinista - en la que los resultados de prueba no se ven afectados por la varianza no controlada, la conducta humana o el error aleatorio - no son realistas, y la frustración o decepción por la falta de perfección determinista no se justifica en el contexto de las pruebas científicas.

Es importante recordar que una de las metas operacionales de las pruebas - en la medicina, psicología, forenses, en los programas de polígrafo y en otros contextos de prueba - es la reducción de los daños resultantes por los errores tanto falsos positivos como falsos negativos. Cualquier método de prueba práctico que alcanza una precisión significativamente mayor que el azar tiene el potencial de reducir ese daño si se utiliza eficientemente. También es importante observar que algunos tipos de pruebas exploratorias pueden estar destinadas a sobre-predecir ligeramente la presencia de problemas - con el objetivo de corregir los errores con pruebas de diagnóstico subsecuentes. Mientras que los errores falsos positivos pueden ser identificados y corregidos con pruebas adicionales y con investigación, la identificación de errores falsos negativos a veces no es posible hasta que el problema ha escalado a un grado en el que a veces podría afectar permanentemente vidas y futuros individuales. Es igualmente importante recordar que ni los resultados de la prueba poligráfica, como ningún tipo de resultado de prueba, debe utilizarse aisladamente como base para la toma de decisiones que afecten los derechos y libertades de los individuos. No existen políticas o estándares de práctica publicadas para polígrafos exploratorios que recomienden o requieran el uso aislado de los resultados de pruebas poligráficas como una base suficiente para la toma de decisiones profesionales.

Como siempre, es deseable mayor información e investigación en algunas áreas, perteneciente tanto a las construcciones teóricas y las preocupaciones prácticas. Las teorías son satisfactorias sólo en la medida que den cuenta de fenómenos conocidos y observados. La aparición de cualquier evidencia o fenómeno que no se pueda explicar mediante nuestras teorías actuales, debe tomarse como un indicador de la necesidad y la obligación de seguir revisando nuestro conocimiento y supuestos en respuesta a la información nueva. El no hacer revisiones a las teorías de trabajo es un indicador de la estasis, y es una característica de los proyectos pseudocientíficos. Los científicos actualizan continuamente todas las teorías de trabajo en respuesta a un volumen siempre en incremento de fenómenos conocidos y observados. Se espera que todas las teorías en el campo de la ciencia evolucionen con el tiempo y avancen hacia un marco integrado con teorías de otros campos de la ciencia. Por esta razón, será importante para la profesión poligráfica el hacer uso continuo de información nueva de los campos de ciencia relacionados.

Las áreas prácticas en las que hace falta mayor información incluyen la precisión tanto de polígrafos diagnósticos como exploratorios, la contribución de los resultados de los polígrafos diagnósticos a las conclusiones de una investigación, y la contribución de los polígrafos exploratorios a las mediciones de resultados de casos y programas como el de violaciones de reglamentos, corrupción, negligencia y reincidencia. Mientras que algunas personas se sientan motivadas a involucrarse en el engaño, y mientras otros traten de utilizar tecnologías científicas para detectar el engaño, habrá otros que estén interesados en el desarrollo de estrategias de contramedidas para evadir la detección. Por esta razón, habrá un interés continuo en el desarrollo de conocimiento acerca de las vulnerabilidades potenciales de la prueba poligráfica, y de qué métodos adicionales se pueden aplicar para contrarrestar esas vulnerabilidades.

A pesar de que se conoce coloquialmente como una prueba de "detección de mentiras" como un término de conveniencia, la ciencia y la razón científica no suponen que el polígrafo realmente mida mentiras *per se*. Todos los resultados de prueba son declaraciones de probabilidad. La alternativa de una comprensión probabilística de los resultados de prueba poligráfica es alentar la expectativa falsa y frustración de que las mentiras por sí mismas pueden ser sometidas de alguna manera a la observación determinista o medición física. En realidad, los principios de fisiología y psicología son tan complejos y variables que es necesario e inevitable un modelo probabilístico.

Debido a que las mentiras *per se* son eventos temporales amorfos, la detección de mentiras probablemente permanecerá como una tarea probabilística e imperfecta. Es importante seguir siendo conscientes de que los objetivos de las pruebas científicas son a menudo para cuantificar y medir fenómenos que no pueden ser sometidos a observación o medición mecánica determinista. Los resultados de las pruebas poligráficas son una medición de la incertidumbre que rodea a una conclusión categórica, basándose en las diferencias de las respuestas que se cargan en los diferentes tipos de estímulos de prueba como una función del engaño o la veracidad con respecto a un comportamiento de preocupación.

Referencias

- Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. Sage.
- Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Sciences*, 18, 313-326.
- Abrams, S. (1974). The validity of the polygraph with schizophrenics. *Polygraph*, 3, 328-337.
- Abrams, S. (1975). A response to Lykken on the polygraph. *American Psychologist*, 30, 709-711.
- Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.
- Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington Books.
- Abrams, S. (. (1984). The question of the intent question. *Polygraph*, 13, 326-332.
- Abrams, S. & Weinstein, E. (1974). The validity of the polygraph with retardates. *Journal of Police Science and Administration*, 2, 397-66.
- American Polygraph Association (2009a). *Model Policy for Post-conviction Sex Offender Testing*. [Electronic version] Retrieved January 25, 2012, from [http://www .polygraph.org](http://www.polygraph.org).
- American Polygraph Association (2009b). *Model Policy for Law Enforcement Public Services Pre-Employment Screening Examinations* [Electronic version]1/25/2012, from <http://www .polygraph.org>.
- American Polygraph Association (2011a). *By-laws: American Polygraph Association, effective 1-1-2012*. [Electronic version] Retrieved January 6, 2012, from <http://www .polygraph.org>.
- American Polygraph Association (2011b). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40(4), 196-305.
- Amsel, T. T. (1999). Exclusive or nonexclusive comparison questions: A comparative field study. *Polygraph*, 28, 273-283.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions In Psychological Science*, 8, 3-9.
- Ansley, N. (1983). A compendium on polygraph validity. *Polygraph*, 12, 53-61.
- Ansley, N. (1990). Law notes: Civil and criminal cases. *Polygraph*, 19, 72-102.
- Ansley, N. (1999, July). *The frequency of appearance of evaluative criteria in polygraph charts*. Defense Personnel Research Center.
- Ansley, N. & Krapohl, D.J. (2000). The frequency of appearance of evaluative criteria in field polygraph charts. *Polygraph*, 29, 169-176.
- ASTM (2002). *Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data* (E 2229-02). ASTM International.
- Backster, C. (1963). *Standardized polygraph notepack and technique guide: Backster zone*

comparison technique. Cleve Backster: New York.

- Balloun, K. D. & Holmes, D.S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology*, 64, 316-322.
- Barland, G. H. (1981). *A Validity and Reliability Study of Counterintelligence Screening Test*. Fort George G. Meade, Maryland: Security Support Battalion, 902d Military Intelligence Group. [Reprinted in *Polygraph* 41 (1) 1-27].
- Barland, G. H., Honts, C. R. & Barger, S.D. (1989). *Studies of the accuracy of security screening polygraph examinations*. DTIC AD Number A304654. Department of Defense Polygraph Institute.
- Barland, G. H. & Raskin, D.C. (1975a). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12 (3), 321-330.
- Barland, G. H. & Raskin, D.C. (1975a). Psychopathy and detection of deception in criminal suspects. *Psychophysiology*, 12, 224.
- Barry, R. J. (1996). Preliminary process theory: Towards an integrated account of the psychophysiology of cognitive processes. *Acta Neurobiologiae Experimentalis*, 56(1), 469-484.
- Bear, M. F., Barry, W. C. & Paradiso, M.A. (2007). *Neuroscience: Exploring the brain: Third Edition*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28 (1), 1-9.
- Ben-Shakhar, G. & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, 81, 273-281.
- Berntson, G. G., & Cacioppo, J. T. (2007). Integrative physiology: Homeostasis, allostasis and the orchestration of systemic physiology, in Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (Eds.) *Handbook of Psychophysiology*, 3rd edition (pp. 433-449). New York: Cambridge University Press.
- Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W. & Flor, H. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Archives of general psychiatry*, 62, 799-805.
- Blackwell, J. N. (1998). *PolyScore 33 and psychophysiological detection of deception examiner rates of accuracy when scoring examination from actual criminal investigations*. Available at the Defense Technical Information Center. DTIC AD Number A355504/PAA. [Reprinted in *Polygraph*, 28 (2), 149-175].
- Blalock, B., Nelson, R., Handler, M. & Shaw, P. (2011). A position paper on the use of directed lie comparison questions in diagnostic and screening polygraphs. *Police Polygraph Digest*, (2011), 2-5.
- Blalock, B., Nelson, R., Handler, M. & Shaw, P. (2012). The empirical basis for the use of directed lie comparison questions in diagnostic and screening polygraphs. *APA Magazine*, 45(1), 36-39.
- Bracha, H. S., Ralston, T. C., Matsukawa, J. M., Williams, A. E. & Bracha, A.S. (2004). "Does "Fight or Flight" need updating? *Psychosomatics*, 45 (5), 448-449.

- Bradley, M. T. & Janisse, M.P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology*, 18, 307-315.
- Boucsein, W. (2012). *Electrodermal activity*. New York: Plenum Press.
- Campion, M., Campion, J. & Hudson, J. (1994). Structured Interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79, 998-1002.
- Cannon, W. B. (1929). *Bodily changes in pain, hunger, fear, and rage*. New York: Appleton-Century-Croft
- Capps, M. H. (1991). Predictive value of the sacrifice relevant. *Polygraph*, 20, 1-6.
- Costanzo, L. (2007). *Physiology*. Hagerstown, MD: Lippincott Williams & Wilkins. Department of Defense (2004a). *Federal psychophysiological detection of deception examiner handbook*. [Retrieved from <http://www.antipolygraph.org> on 2-23-2008].
- Crewson, P. E. (2001). *A comparative analysis of polygraph with other screening and diagnostic tools*. Research Support Service. Report No. DoDPI01-R-0003. Reprinted in *Polygraph* 32 (57-85).
- Department of Defense (2004b). *Test data analysis: DoDPI numerical evaluation scoring system*. [Retrieved from <http://www.antipolygraph.org> on 6-28-2007].
- Department of Defense (2006a). *Federal psychophysiological detection of deception examiner handbook*. Reprinted in *Polygraph*, 40 (1), 2-66.
- Department of Defense (2006b). *Test data analysis: DoDPI numerical evaluation scoring system*. [Retrieved from <http://www.antipolygraph.org> on 3-31-2007].
- Department of Defense Polygraph Institute (2002). *Department of Defense Polygraph Institute Law Enforcement Pre-employment Test*. [Retrieved from <http://www.antipolygraph.org> on 6-2-2011].
- Dollins, A. B., Krapohl, D. J. & Dutton, D. (1999). *A comparison of computer programs designed to evaluate psychophysiological detection of deception examinations: Bakeoff 1*. Department of Defense Polygraph Institute. Reprinted in *Polygraph* 29 (237-247).
- Domjan, M. (2010). *Principles of learning and behavior, 6th edition*. Cengage/Wadsworth.
- Drever, E. (1995). *Using Semi-Structured Interviews in Small-Scale Research. A Teacher's Guide*. Scottish Council for Research in Education, Edinburgh.
- Elaad, E. & Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology*, 11, 99-108.
- Furedy, J. J. (1989). The North American CQT polygraph and the legal profession: a case of Canadian credulity and a cause for cultural concern. *The Criminal Law Quarterly*, 31, 431-451.
- General Accounting Office (1991). *Using structured interviewing techniques*. Program Evaluation and Methodology Division [Retrieved online http://www.gao.gov/policy/10_1_5.pdf on 10-

20-2009].

- Green, D. M. & Swets. J A (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Greenberg, I. (1982). The role of deception in decision theory. *Journal of Conflict Resolution*, 26, 139-156.
- Groves, P. M. & Thompson, R.F. (1970). Habituation: A dual-process theory. *Psychological Review*, 77, 419-450.
- Handler M. (2006). The Utah PLC. *Polygraph*, 35, 139-148. Handler, M. & Nelson, R. (2007). Polygraph terms for the 21st Century. *Polygraph*, 36, 157-164.
- Handler, M., Geddes, L., Reicherter, J. (2007). A discussion of two diagnostic features of the cardiovascular channel. *Polygraph*, 36 (2), 70-83.
- Handler, M., Honts, C., Krapohl, D., Nelson, R. & Griffen, S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Police and Criminal Psychology*, 24, 69-86.
- Handler, M. & Krapohl, D. (2007). The use and benefits of the photoelectric plethysmograph in polygraph testing. *Polygraph*, 36, 18-25.
- Handler, M. & Nelson, R. (2007). Polygraph terms for the 21st Century. *Polygraph*, 36, 157-164.
- Handler, M. & Nelson, R. (2008). Utah approach to comparison question polygraph testing. *European Polygraph*, 2, 83-110.
- Handler, M. Honts, C. & Nelson, R. (2013). Information gain of the Directed Lie Screening Test. *Polygraph*, 42, 192-202.
- Handler, M., Nelson, R. & and Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph*, 37 (2), 100-111.
- Handler, M., Nelson, R., Krapohl, J. & Honts, C. (2010). An EDA Primer for Polygraph Examiners. *Polygraph*, 39, 68-108.
- Handler, M., & Reicherter, J. (2008). Respiratory blood pressure fluctuations observed during polygraph examinations. *Polygraph*, 37 (4), 256-262.
- Handler, M., Rovner, L. & Nelson, R. (2008). The concept of allostasis in polygraph testing. *Polygraph*, 38, 228-233.
- Handler, M., Shaw, P. & Gougler, M. (2010). Some thoughts about feelings: A study of the role of cognition and emotion in polygraph testing. *Polygraph*, 39, 139-154.
- Harris, J. C. & Olsen, D.E. (1994). Polygraph Automated Scoring System. Patent Number: 5,327,899. U.S. Patent and Trademark Office.
- 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.
- Hastie, T., Tibshirani, R. & Friedman, J. (2001). *Elements of statistical learning*. Springer.
- Hilliard, D. L. (1979). A cross analysis between relevant questions and a generalized intent to

answer truthfully question. *Polygraph*, 8, 73-77.

- Hodgson, P. (1987). *A practical guide to successful interviewing*. Maidenhead: McGraw-Hill.
- Honts, C. R. (1987). Interpreting research on polygraph countermeasures. *Journal of Police Science and Administration*, 15, 204-209.
- Honts, C. R. (1997). *Is it time to reject the friendly polygraph examiner hypothesis (FPEH)?* Paper presented at the American Psychological Society at the 9th annual meeting, Washington, D. C., May 23-26, 1997.
- Honts, C. R. & Alloway, W.R. (2007). Information does not affect the validity of a comparison question test. *Legal and Criminological Psychology*, 12, 311-320.
- Honts, C. R. & Amato, S. (2001). Psychophysiological credibility assessment. *Journal of Forensic Psychology Practice*, 1, 87-99.
- Honts, C., Amato, S. & Gordon, A. (2004). Effects of outside issues on the comparison question test. *Journal of General Psychology*, 131 (1), 53-74.
- Honts, C. R. & Devitt, M.K. (1992). *Bootstrap decision making for polygraph examinations*. Department of Defense Polygraph Institute report No DoDPI92-R-0002.
- Honts, C. & Handler, M. (2014). Scoring respiration when using directed lie comparison questions. *Polygraph*, 43 (3) 71-78.
- Honts, C. Handler, M. Shaw, P. & Gougler, M. (2015). The vasomotor response in the comparison question test. *Polygraph* 44 (1) 62-78.
- Honts, C. R. & Hodes, R.L. (1983). The detection of physical countermeasures. *Polygraph*, 12, 7-17.
- Honts, C. R., Hodes, R. L. & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, 70, 177-187.
- Honts, C. R. & Peterson, C.F. (1997). *Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer*. Available from the author.
- Honts, C. R. & Raskin, D. C. (1988). A field study of the validity of the directed-lie control question. *Journal of Police Science and Administration*, 16, 56-61.
- Honts, C. R., Raskin, D. C. & Kircher, J.C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Psychophysiology*, 1, 241-247.
- Honts, C. R., Raskin, D. C. & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252-259.
- Honts, C. R., Raskin, D. C., Kircher, J. C. & Hodes, R.L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. *Journal of Police Science and Administration*, 16, 91-94.
- Honts, C. R. & Reavy, R. (2009). *Effects of comparison question type and between test stimulation on the validity of comparison question test. Final progress report on contract No.W911Nf-07-1-0670*, submitted to the Defense Academy of Credibility Assessment (DACA). Boise State University.

- Honts, C. R., Winbush, M. & Devitt, M.K. (1994). Physical and mental countermeasures can be used to defeat guilty knowledge tests. *Psychophysiology*, 31, S57.
- Horneman, C. J. & O'Gorman, J.G. (1985). Detectability in the card test as a function of the subject's verbal response. *Psychophysiology*, 22, 330-333.
- Horowitz, S. W., Kircher, J. C. & Raskin, D.C. (1986). Does stimulation test accuracy predict accuracy of polygraph test?. *Psychophysiology*, 23, 442.
- Horowitz, S. W., Kircher, J. C., Honts, C. R. & Raskin, D.C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.
- Horvath, F. & Palmatier, J. (2008). Effect of two types of control questions and two question formats on the outcomes of polygraph examinations. *Journal of Forensic Sciences*, 53(4), 1-11.
- Horvath, F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16, 198-209.
- Horvath, F. S. (1994). The value and effectiveness of the sacrifice relevant question: An empirical assessment. *Polygraph*, 23, 261-279.
- Iacono, W. G., Boisvenu, G. A. & Fleming, J.A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology*, 69, 289-299.
- Iacono, W. G., Cerri, A. M., Patrick, C. J. & Fleming, J.A. (1992). Use of antianxiety drugs as countermeasures in the detection of guilty knowledge. *The Journal of applied psychology*, 77, 60-4.
- Janig, W. (2006). *The Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis*. Cambridge University Press.
- Kahn, R. L. and C. F. Cannell (1957). *The psychological basis of the interview. The dynamics of interviewing: theory, technique, and cases*. New York, John Wiley & Sons: 22-64.
- Kahn, J., Nelson, R. & Handler, M. (2009). An exploration of emotion and cognition during polygraph testing. *Polygraph*, 38, 184-197.
- Kirby, S. L. (1981). The comparison of two stimulus tests and their effect on the polygraph technique. *Polygraph*, 10, 63-76.
- Kircher, J. C. (1981). *Computerized chart evaluation in the detection of deception*. Masters thesis: University of Utah.
- Kircher, J. C. (1983). *Computerized decision making and patterns of activation in the detection of deception*. Dissertation Abstracts International, 44, 345.
- Kircher, J. & Raskin, D. (2002). Computer methods for the psychophysiological detection of deception. In Murray Kleiner (Ed.), *Handbook of Polygraph Testing*. San Diego: Academic Press.
- Kircher, J. C., Kristjansson, S. D., Gardner, M. K. & Webb, A. (2005). *Human and computer decision-making in the psychophysiological detection of deception*. University of Utah.

- Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*, 291-302.
- Kircher, J. C., Packard, T., Bell, B. G. & Bernhardt, P. C., (2001). *Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Probable-Lie and Directed-lie Polygraph Tests*. Final report to the U. S. Department of Defense Polygraph Institute, Ft. Jackson, SC. Salt Lake City: University of Utah, Department of Educational Psychology.
- Krapohl, D. J. (2002). Short report: Update for the objective scoring system. *Polygraph, 31*, 298-302.
- Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph, 28*, 209-222.
- Krapohl, D. J. & Ryan, A.H. (2001). Final Comment on the bleated look at symptomatic questions. *Polygraph, 30*, 218-219.
- Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing*. Sage Publications.
- Lehmann, E. L. (1950). Some principles of the theory of testing hypotheses. *Annals of Mathematical Statistics, 21* (1), 1-26.
- Lieblich, I., Ben-Shakhar, G., Kugelmass, S. & Cohen, Y. (1978). Decision theory approach to the problem of polygraph interrogation. *Journal of Applied Psychology, 63*, 489-498.
- Light, G. D. (1999). Numerical evaluation of the Army zone comparison test. *Polygraph, 28*, 37-45.
- Lindlof, T. R. & Taylor, B.C. (2002). *Qualitative communication research methods*. Thousand Oaks: Sage.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-55.
- Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector*. New York:McGraw-Hill.
- MacLaren, V. & Krapohl, D. (2003). Objective Assessment of Comparison Question Polygraphy. *Polygraph, 32*, 107-126.
- Marcum, J. I. (1947). *A statistical theory of target detection by pulsed radar*. Rand Corporation. [Retrieved online <http://www.rand.org> 2011-11-2].
- Maton, A., Hopkins, J., McLaughlin, C., Johnson, S., Quon Warner, M., LHart, D. & Wright, J. (1993). *Human biology and health*. Englewood Cliffs, New Jersey, USA: Prentice Hall.
- National Research Council (2003). *The polygraph and lie detection*. Washington, D.C.: National Academy of Sciences.
- Nelson, R. (2014a). What does the polygraph measure? *APA Magazine, 47*(2) 39-47.
- Nelson, R. (2014b). Redux: What does the polygraph measure? (in 600 words or less). *APA Magazine, 47* (3), 36-37.
- Nelson, R. (2014c). Take 3: What does the polygraph measure? (in 250 words or less). *APA Magazine, 47* (4), 60.

- Nelson, R. (2014d). Short Answer: What does the polygraph measure? (in 150 words or less). *APA Magazine*, 47 (5), 27.
- Nelson, R. (2014e). Sound-bite: What does the polygraph measure? (in 50 words or less). *APA Magazine*, 47 (6), 29.
- Nelson, R. & Handler, M. (2010). *Empirical scoring system: NPC quick reference*. Lafayette Instrument Company. Lafayette, IN.
- Nelson, R. & Handler, M. (2012). Monte Carlo study of criterion validity of the directed lie screening test using the Empirical Scoring System and the Objective Scoring System version 3. *Polygraph*, 41, 145-155.
- Nelson, R. & Handler, M. (2013). A brief history of scientific reviews of polygraph accuracy research. *APA Magazine*, 47 (6), 22-28.
- Nelson, R. & Handler, M. (2015,). Statistical reference distributions for comparison question polygraphs. *Polygraph*, 44 (1), 91-114.
- Nelson, R., Handler, M., Blalock, B. & Hernandez, N. (2012). Replication and extension study of Directed Lie Screening Tests: criterion validity with the seven and three Position models and the Empirical Scoring System. *Polygraph* 41 (3), 186-198.
- Nelson, R., Handler, M., & Morgan, C. (2012). Criterion validity of the directed lie screening test and the empirical scoring system with inexperienced examiners and non-naive examinees in a laboratory setting. *Polygraph* 41 (3), 176-185.
- 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.
- Nelson, R., Krapohl, D. & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37, 185-215.
- Offe, H. & Offe, S. (2007). The comparison question test: does it work and if so how?. *Law and human behavior*, 31, 291-303.
- Office of Technology Assessment (1983). *The validity of polygraph testing: A research review and evaluation*. Printed in *Polygraph*, 12, 198-319.
- Ogilvie, J. & Dutton, D. (2008). Improving the detection of physical countermeasures with chair sensors. *Polygraph*, 37 (2), 136-148.
- Olsen, D. E., Ansley, N., Feldberg, I. E., Harris, J. C. & Cristion, J.A. (1991). Recent developments in polygraph technology. *Johns Hopkins APL Technical Digest*, 12, 347-357.
- Olsen, D. E., Harris, J. C. & Chiu, W.W. (1994). The development of a physiological detection of deception scoring algorithm. *Psychophysiology*, 31, S11.
- Olsen, D. E., Harris, J. C., Capps, M. H. & Ansley, N. (1997). Computerized polygraph scoring system. *Journal of Forensic Sciences*, 42, 61-70.
- Palmatier, J. J. (1991). *Analysis of two variations of control question polygraph testing utilizing exclusive and nonexclusive controls*. Unpublished doctoral dissertation.

- Palmatier, J., Rovner, L (2014). Credibility assessment: Preliminary process theory, the polygraph process, and construct validity. *International Journal of Psychophysiology*, [available online: <http://www.ncbi.nlm.nih.gov/pubmed/24933412>].
- Paradiso, M. A., Bear, M. F. & Connors, B.W. (2007). *Neuroscience: Exploring the Brain*. Hagerstwon, MD: Lippincott Williams & Wilkins.
- Patrick, C. J. & Iacono, W.G. (1989). Psychopathy, threat and polygraph test accuracy. *Journal of Applied Psychology*, 74, 347-355.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Sage Publications Inc. 2009.
- Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-359.
- Podlesny, J., Raskin, D. & Barland, G. (1976). *Effectiveness of techniques and physiological measures in the detection of deception*. Report No. 76-5, Contract 75-N1-99-001 LEAA (available through Department of Psychology, University of Utah, Salt Lake City).
- Popper, K. R. (1959). *The logic of scientific discovery*. Routledge.
- Porges, S. (2011). *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, and Self-regulation*. Norton.
- Pratt, J., Raiffa, H. & Schlaifer, R. (1995). *Introduction to statistical decision theory*. Cumberland, RI: MIT Press.
- Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J. & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of applied psychology*, 89, 1099-105.
- Popper, K. R., (1959). *The Logic of Scientific Discovery*. The German version is currently in print by Mohr Siebeck (ISBN 3-16-148410-X), the English one by Routledge publishers (ISBN 0-415-27844-9).
- Powell, M. B. & Snow, P.C. (2007). Guide to questioning children during the free-narrative phase of an investigative interview. *Australian Psychologist*, 42 (1), 57-65.
- Raffle, A. E. & Muir Gray, J.A. (2007). *Screening*. Oxford University Press.
- Raskin D. C., Honts. C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of Polygraph Testing*. San Diego:Academic Press.
- Raskin, D. C., Honts, C. R., Kircher, J. C. (2014). *Credibility assessment: Scientific research and applications*. Academic Press.
- Raskin, D. C. & Hare, R.D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126-136.
- Raskin, D. C. & Kircher, J.C. (1990, May 2). *Development of a computerized polygraph system and physiological measures for detection of deception and countermeasures: A pilot study*. Scientific Assessment Technologies, Inc.

- Raskin, D. C. & Podlesny, J.A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin*, 86, 54-59.
- Raskin, D., Kircher, J. C., Honts, C. R. & Horowitz, S.W. (1988). *A study of the validity of polygraph examinations in criminal investigations*. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542-547. Reprinted in *Polygraph* 11, 17-21.
- Research Division Staff (1995a). *Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage*. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 27, (3), 171-180.
- Research Division Staff (1995b). *A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats*. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 26 (2), 79-106.
- Richerter J. & Handler, M. (2014). *A physiology manual for PDD lifelong learners of the science*. Available from the authors. Retrieved online at [http://www.polygraph.org/files/apa_psychophysiology_study_guide.pdf] on 11-30-2014.
- Rovner, L. I. (1979). *The effects of information and practice on the accuracy of physiological detection of deception*. Unpublished doctoral dissertation: University of Utah.
- Rovner, L. I. (1986). Accuracy of physiological detection of deception for subjects with prior knowledge. *Polygraph*, 15 (1), 1-39.
- Rovner, L. I., Raskin, D. C. & Kircher, J.C. (1979). Effects of information and practice on detection of deception. *Psychophysiology*, 16, 197-198 (abstract).
- Standring, S. (2005). *Gray's anatomy* (39th ed.). Elsevier Churchill Livingstone.
- Saxe, L. (1991). Science and the CQT polygraph: A theoretical critique. *Integrative Physiological and Behavioral Science*, 26, 223-231.
- Schonhoff, T. A. & Giordano, A.A. (2006). *Detection and estimation theory and its applications*. New Jersey: Pearson Education.
- Senter, S. M. (2003). Modified general question test decision rule exploration. *Polygraph*, 32, 251-263.
- Senter, S. M. & Dollins, A.B. (2002). *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.
- Senter, S. M. & Dollins, A.B. (2008). Optimal decision rules for evaluating psychophysiological detection of deception data: an exploration. *Polygraph*, 37 (2), 112-124.
- Senter, S., Weatherman, D., Krapohl, D. & Horvath, F. (2010). Psychological set or differential salience: A proposal for reconciling theory and terminology in polygraph testing. *Polygraph*, 39 (20), 109-117.

- Silverthorn, D. U. (2009). *Human physiology: An integrated approach* (4 ed.). Pearson/Benjamin Cummings.
- Stephenson, M. & Barry, G. (1986). Use of a motion chair in the detection of physical countermeasures. [Reprinted in *Polygraph*, 17, 21-27].
- Sterling, P. and Eyer, J. (1988) Allostatics: a new paradigm to explain arousal pathology. In: Fisher, S., Reason, J. (Eds.) *Handbook of Life Stress, Cognition and Health*. New York: Wiley and Sons.
- Summers, W. G. (1939). Science can get the confession. *Fordham Law Review*, 8, 334-354.
- Swets, J. A. (1964). Signal detection and recognition by human observers. New York: Wiley.
- Swets, J. A. (1996). Signal detection theory and ROC analysis in psychology and diagnosis. Mahwah, NJ: Erlbaum.
- Tanner, J., Wilson, P. & Swets, J. (1954). A decision-making theory of visual detection. *Psychological Review*, 61 (6), 401-409.
- Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. & Updegraff, J.A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review*, 107, 411-429.
- Timm, H. W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10, 47-51.
- Timm, H. W. (1991). Effect of posthypnotic suggestions on the accuracy of preemployment polygraph testing. *Journal of Forensic Sciences*, 36, 1521-1535.
- Vrij, A. (2008). *Detecting Lies and Deceit, Pitfalls and Opportunities*, 2nd edition. West Sussex, England: Wiley & Sons.
- Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10 (4), 299-326.
- Waid, W. M., Orne, E. C. & Orne, M.T. (1981). Selective memory for social information, alertness, and physiological arousal in the detection of deception. *Journal of Applied Psychology*, 66, 224-232.
- Wells, G. L. & Olson, E.A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, 8, 155-167.

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

- Wickens, T. D. (1991). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, 36, 213-234.
- Wickens, T. D. (2002). Elementary signal detection theory. New York: Oxford.
- Widup, R. & Barland, G.H. (1994). *Effect of the location of the numbers test on examiner decision rates in criminal psychophysiological detection of deception tests*. Department of Defense Polygraph Institute.
- Wilson, J. & Jungner, G. (1968). Principles and practice of screening for disease. *WHO Chronicle*
Geneva: World Health Organization, 22 (11):473.