Ya estas inscrito???????

No puedes faltar!!!!!!!!

Seminario Latino Americana 2019

Presentado por Sabino Martinez y Mike Gougler

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

rodolfo@poligrafia.com.mx

Reduciendo Resultados Inconclusos: Un Análisis Descriptivo de las Reglas de Decisión, Puntuaciones Electrodérmicas Ponderadas y Puntajes de Corte Multinomiales

Raymond Nelson y Mark Handler

Traductor

rodolfo@poligrafia.com.mx

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

Reduciendo Resultados Inconclusos: Un Análisis Descriptivo de las Reglas de Decisión, Puntuaciones Electrodérmicas Ponderadas y Puntajes de Corte Multinomiales

Raymond Nelson y Mark Handler[1]

Abstract

Se utilizó una muestra de archivo de N=100 exámenes poligráficos de campo confirmados para calcular estadísticas descriptivas, incluyendo estimaciones de puntos, para estudiar el efecto en los resultados inconclusos y otras métricas de precisión de la prueba. Los datos se analizaron en función de las diferentes reglas de decisión, la ponderación estructural de las puntuaciones de los sensores y los puntajes de corte. Las reglas de decisión incluyeron las reglas de zona federal, regla de gran total, regla de subtotal y reglas de dos etapas. Las puntuaciones numéricas se obtuvieron mediante un algoritmo de extracción de características automatizado. Los resultados se evaluaron tanto con puntuaciones numéricas de tres posiciones no ponderadas y con las puntuaciones enteras después de ponderar las respuestas electrodérmicas. Los resultados se muestran tanto para los puntajes de corte numéricos tradicionales como también para los puntajes de corte obtenidos en las distribuciones de referencia multinomiales para las puntuaciones de las pruebas con preguntas de comparación de tres posiciones con puntuaciones electrodérmicas tanto ponderadas como no ponderadas. El uso de las diferentes reglas de decisión tuvo menos efecto con los puntajes de corte multinomiales que con los puntajes de corte tradicionales. Los puntajes de EDA ponderados produjeron una reducción promedio del 49% en los resultados inconclusos en todas las reglas de decisión, y la combinación de los puntajes del EDA ponderados y los puntajes de corte multinomiales redujeron la aparición de resultados inconclusos en un promedio del 72%.

Introducción

Pocas cosas son más decepcionantes para los examinadores poligráficos de campo y para los agentes de referencia que los resultados inconclusos[2] en las pruebas. Poco se compara con la sensación de frustración cuando - después de todo el tiempo y el esfuerzo invertidos en la preparación de la prueba, la entrevista previa a la prueba, la selección de objetivos, la formulación de preguntas, la recopilación de datos de prueba y el análisis de datos de prueba - el resultado de la prueba no es estadísticamente significativo para engaño o veracidad.

1 Los autores están extremadamente agradecidos con el Sr. Don Krapohl y el Capitán Matt Hicks del DPS de Texas (quien funcionó como editor de acción para este proyecto) por revisar, comentar y editar borradores anteriores de este manuscrito.

2 Para mantener la coherencia con la terminología de las ciencias forenses, hemos elegido llamar "inconclusos" a los resultados de polígrafo indeterminados en lugar del neologismo "Sin opinión".

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

Este sentido de frustración es, a veces, compartido por algunos examinados - especialmente aquellos que son inocentes - quienes, habiendo aceptado realizar la prueba con la esperanza de producir datos de prueba que puedan servir como base de evidencia para apoyar conclusiones profesionales sobre su inocencia o veracidad, necesariamente se les debe informar que el resultado de prueba inconcluso no proporciona mejor información sobre la verdad o el engaño de la que ya estaba disponible antes de la prueba.

En años previos, en ausencia de una visión probabilística de los resultados de las pruebas poligráficas, pudo ser tentador vender las capacidades del polígrafo como virtualmente infalibles. Durante esa era, un resultado inconcluso a veces se consideraba como una indicación de un examinador no calificado. Detrás de esta actitud o creencia, es probable que existía un deseo sincero entre los examinadores poligráficos ser una ayuda real para los profesionales de referencia, para quienes un resultado de prueba inconcluso ofrecía poco o ningún valor práctico. Y prácticamente nadie quiere comprar los servicios de una prueba probabilística para la que podríamos ofrecer poco más que conjeturas sobre la fuerza de la conclusión. En ausencia de una habilidad para cuantificar de manera realista el nivel de confianza o el margen de incertidumbre para el resultado de la prueba, puede haber sido una cuestión de marketing profesional el aumentar o vender en exceso las capacidades de la prueba, incluyendo las expectativas poco realistas de infalibilidad y la ausencia de resultados inconclusos.

Puede ser útil recordar que los resultados inconclusos no son exclusivos de las pruebas poligráficas; todas las pruebas forenses están cargadas con una cierta proporción de resultados inconclusos. La razón declarada de ellos depende de la disciplina forense que realiza la prueba, pero en general se debe a que la señal / patrón / trazo / muestra / marcador / imagen es inadecuada o está contaminada. Lo mismo ocurre con el polígrafo. Incluso en condiciones ideales, con un esfuerzo heroico y con examinados, entrevistas previas a la prueba y gráficos perfectos, siempre habrá casos en los que los datos no permitirán una decisión confiable. Un objetivo importante en todas las disciplinas forenses es minimizar esas ocasiones.

Objetivamente, si la gente creyera que el polígrafo es infalible, entonces no sería muy difícil aceptar un resultado de prueba en sentido literal. Sin embargo, la mayoría de los examinadores de polígrafo, y otros profesionales que conocemos, dudarían en aceptar un resultado de prueba en un sentido literal sin analizar la administración de la prueba, los datos de la prueba y el resultado analítico. Además, ninguna persona pensante aceptará hoy la noción de que el polígrafo es infalible o determinista. Está bien establecido que todos los resultados de las pruebas científicas se basan en probabilidades y teoría de la probabilidad para cuantificar fenómenos importantes que no pueden ser objeto de observación determinista perfecta o medición física directa. Las tecnologías, metodologías y estándares poligráficos han evolucionado junto con este entendimiento.

En la actualidad, la mayoría de los profesionales del polígrafo, los agentes remitentes, los tribunales, los legisladores, los científicos, los miembros de la comunidad y las industrias de entretenimiento y noticias entienden razonablemente que la prueba del polígrafo es, como

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

otras pruebas científicas, simplemente un clasificador estadístico destinado a cuantificar un fenómeno que no puede ser objeto de observación determinista perfecta o de medición física directa. Todas las pruebas son fundamentalmente probabilísticas. Junto con cualquier uso de la estadística y de la teoría de la probabilidad, existe el potencial de error de prueba. A menudo, el propósito práctico del análisis de datos de prueba es lograr una clasificación categórica o un resultado de prueba categórica. Pero los expertos que poseen una comprensión más amplia y completa de la prueba científica son conscientes de que el propósito real del análisis de datos de prueba es cuantificar - de alguna manera reproducible - el nivel de confianza o el margen de incertidumbre que rodea el resultado de la prueba. Cualquier uso razonable e inteligente de la teoría estadística y de los métodos estadísticos incluirá el reconocimiento de cierto potencial de error de prueba, y cierto potencial de que los resultados de la prueba no alcancen un nivel requerido de significancia estadística. A pesar de esto, los resultados inconclusos de las pruebas siguen siendo la pesadilla de los examinadores poligráficos en todo el mundo.

Reducir la proporción de resultados inconclusos, potencialmente puede aumentar la efectividad de la prueba. Los resultados inconclusos están inversamente relacionados con la utilidad. Tasas más altas de resultados inconclusos corresponden con una tasa más baja de utilidad. Una estrategia común para reducir la aparición de resultados de pruebas inconclusos es aumentar el volumen de datos de pruebas disponibles, repitiendo la serie de preguntas relevantes y de comparación una o dos veces más que las tres repeticiones mínimas. Otra estrategia para reducir la aparición de resultados inconclusos es incluir información de diagnóstico independiente adicional, como un sensor vasomotor a la gama tradicional de sensores de respiración, cardio y electrodérmicos. También existen otras estrategias más sutiles para reducir la aparición de resultados de pruebas inconclusos, que pueden incluir la selección de objetivos, formulación de preguntas y habilidades de entrevista.

En este documento, utilizamos una muestra de archivo de polígrafos de investigación criminal confirmados para mostrar tres enfoques adicionales basados en evidencia de las prácticas de campo que pueden reducir la aparición de resultados de pruebas inconclusos al tiempo que aumentan la objetividad y el poder estadístico de la prueba. El primero de estos involucra las *reglas de decisión poligráficas*. La segunda área de práctica de campo involucra *la ponderación de las puntuaciones electrodérmicas* (EDA) cuando se asignan puntuaciones de tres posiciones tipo Likert (1932) a la respuesta fisiológica de las preguntas poligráficas. En tercer lugar, mostramos el efecto de *reemplazar los puntajes de corte tradicionales poligráficos* con puntajes de corte derivados de una distribución multinomial de las puntuaciones poligráficas basadas en la teoría analítica de la prueba poligráfica.

Método de Datos de Muestra

Se obtuvo una muestra de polígrafos de campo confirmados de un estudio anterior realizado por Krapohl (2005). Los datos de la muestra consistieron en N=100 exámenes poligráficos que se seleccionaron al azar de un archivo de casos de campo confirmados. Todos los exámenes fueron realizados por las agencias de procuración de la ley de EE. UU. utilizando el formato de Prueba de Comparación de Zona Federal (ZCT) (Light, 1999; Department of

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

Defense, 2006a), para el que la secuencia de preguntas calificadas, incluidas las tres preguntas relevantes y tres preguntas de comparación que se repitieron tres veces. Los casos de muestra consistieron en n=50 casos veraces confirmados y n=50 casos de engaño confirmados. Los datos para todos los casos constaron del registro de sensores para detectar cambios en la actividad respiratoria torácica y abdominal, la actividad electrodérmica y la actividad cardiovascular.

La Tabla 1 muestra un cálculo extendido de los resultados de la muestra informados por Krapohl (2005) utilizando la Regla de Zona Federal (FZR) y los puntajes de siete posiciones, incluyendo la tasa de sensibilidad de prueba o verdadero positivo (TP) y la tasa de especificidad o verdadera o negativa (TN), errores falso-positivo (FP) y falso-negativo (FN), resultados inconclusos (INC) y el promedio no ponderado de decisiones correctas y resultados inconclusos para casos confirmados de engaño y veracidad. La Tabla 1 también muestra el valor predictivo positivo (PPV; calculado como TP/(TP + FP) y el valor predictivo negativo (NPV; calculado como TN/(TN + FN). También se muestra en la Tabla 1 el coeficiente de eficiencia de detección (DEC; Kircher, Horowitz y Raskin, 1988), calculado como la correlación de Pearson entre el estatus del caso [-1, 1] y el resultado de la prueba [-1, 0, 1]. El DEC proporciona una métrica única que abarca clasificaciones correctas, errores y resultados inconclusos para casos de muestra tanto de engaño como veraces.

Krapohl (2005) reportó los resultados utilizando "reglas de investigación", para las cuales se realizaron clasificaciones de engaño si el puntaje del gran total igualó o superó un puntaje de corte de -6, o cualquier puntaje subtotal igual o que excediera el -3. Las clasificaciones veraces se realizaron solo cuando el gran total fue igual o superior a +6 y todos los puntajes subtotales superaron el +1. Todas las demás condiciones fueron clasificadas como inconclusas. Estos procedimientos pueden encontrarse en publicaciones del Department of Defense (2006a; 2006b) y se refiere como la Regla de Zona Federal (FZR) como recordatorio de este manuscrito.

Krapohl (2005) también reportó resultados usando "reglas evidenciarias", para las cuales se hicieron clasificaciones de engaño si el puntaje del gran total era igual o superior a -6, mientras que se hicieron clasificaciones de veracidad si el puntaje del gran total era igual o superior a +4. En los casos en que el gran total estuvo en el rango entre -5 a +3, se utilizaron puntuaciones subtotales para hacer clasificaciones de engaño si algún subtotal igualó o superó el -3. Todas las demás condiciones fueron clasificadas como inconclusas. Senter y Dollins (2003) describieron por primera vez el proceso de usar la puntuación de gran total y posteriormente usar las puntuaciones subtotales si la suma de gran total no era inconclusa, y se refiere como la regla de dos etapas (TSR) para el resto de este manuscrito

Tabla 1. Resultados de la muestra (n=100) reportados por Krapohl (2005) para las puntuaciones de siete posiciones.

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

	Reglas de Investigación / FZR	Reglas Evidenciarias / TSR
Sensibilidad (engaño)	.78	.81
Especificidad (veracidad)	.62	.80
Errores falso negativo	.07	.10
Errores falso positivo	.09	.09
Culpables inconclusos	.15	.09
Inocentes inconclusos	.29	.11
Precisión no ponderada	.90	.90
Inconclusos no ponderados	.22	.10
Valor predictivo positivo	.90	.90
Valor predictivo negativo	.90	.89
Coeficiente de Detección de Eficiencia	.67	.74

Los resultados de Krapohl (2005), mostrados en la Tabla 1, son consistentes con otros estudios utilizando el FZR y mostraron que los resultados inconclusos se cargan hacia las personas inocentes. Krapohl demostró que el uso de las puntuaciones de siete posiciones con el TSR y con los puntajes de corte evidenciarios dio como resultado una precisión de clasificación similar, aunque con una especificidad de prueba mejorada y una reducción de la tasa de inconclusos en un 55%. Es importante destacar que Krapohl utilizó diferentes puntajes de corte para el FZR y TSR, y no se sabe qué porción de la diferencia observada se puede atribuir a las reglas de decisión y/o a los diferentes puntajes de corte. El presente análisis es un intento de proporcionar más información acerca de las diferencias observadas en la precisión de la prueba en función de las tasas de inconclusos para las reglas de decisión, puntajes de corte y otros factores.

Extracción de Características y Reducción de Datos

Los casos de muestra se calificaron utilizando el método de puntuación numérica de tres posiciones (Bradley y Janisse, 1981; Departament of Defense, 2006b; van Herk, 1991). Para cada iteración de cada pregunta relevante y para cada sensor de registro, se asignaron puntuaciones numéricas de tres posiciones a través de un algoritmo de extracción de características automatizado que se desarrolló utilizando el lenguaje de cálculo estadístico R (R Core Team, 2018). El método de puntuación de tres posiciones es un sistema de codificación tipo Likert, basado en la teoría analítica de la prueba del polígrafo (Nelson, 2015). Las puntuaciones de +1, 0 y -1 se asignaron a los pares de preguntas relevantes y de comparación - a los que los examinadores poligráficos de campo los denominan *análisis subtotales o "spots"*. Las puntuaciones negativas indicaron mayores cambios en la actividad fisiológica en respuesta a las preguntas relevantes, mientras que las puntuaciones positivas indicaron mayores cambios en la actividad fisiológica en respuesta a las preguntas de comparación. Se asignaron puntuaciones de 0 cuando hubo poca o ninguna diferencia en la respuesta a las preguntas relevantes y de comparación.

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 algoritmo automatizado fue diseñado para extraer información relacionada con la amplitud relativa del aumento de la actividad electrodérmica, el aumento relativo de la presión arterial y la supresión relativa o la reducción de la actividad respiratoria. Se ha demostrado que estas respuestas se correlacionan con las diferencias en la respuesta fisiológica a las preguntas poligráficas relevantes y de comparación basándose en la teoría analítica de la prueba poligráfica con preguntas de comparación (Kircher y Raskin, 1988; Kubis, 1962; Summers, 1939) [ver Nelson (2016) para su discusión]. El algoritmo automatizado de extracción de características realizó casi todas las tareas de análisis tradicionales, incluyendo la identificación del inicio de la respuesta y el pico de respuesta y el cálculo de las diferencias numéricas entre el pico de respuesta y el inicio de la respuesta.

El algoritmo automatizado de extracción de características también seleccionó pares de preguntas relevantes y de comparación, de manera que para cada repetición de las secuencias de preguntas para los casos ZCT en los datos de la muestra, las respuestas fisiológicas a la segunda y tercera pregunta relevante se combinaron con la respuesta fisiológica de la pregunta de comparación anterior. Para cada sensor de registro, el algoritmo automatizado apareó las respuestas fisiológicas de la primera pregunta relevante con la respuesta fisiológica de la pregunta de comparación previa o posterior seleccionando la pregunta de comparación que produjo el mayor cambio en la actividad fisiológica.

Previo a la extracción de características, el algoritmo automatizado escaló los datos fisiológicos registrados para su visualización y realizó una identificación y rechazo de datos con artefactos como las respiraciones profundas en la actividad respiratoria, movimientos físicos en los datos de cardio y respuestas electrodérmicas lábiles que no estaban relacionadas o no eran oportunas con la pregunta de prueba.

Se obtuvieron mediciones numéricas sin dimensiones a partir de los datos fisiológicos escalados desde el inicio hasta el final de la respuesta utilizando una ventana de evaluación de 15 segundos (EW). La excursión de la respiración se midió con la media de intervalos de 1 segundo de 0 a 14 segundos a una tasa de datos de 30 muestras por segundo, excluyendo 2 segundos antes y después de la respuesta verbal. Las respuestas electrodérmicas y cardiovasculares se midieron por su diferencia adimensional de la diferencia máxima entre el inicio de un segmento de pendiente positiva que comenzó durante la ROW y un pico subsecuente de un segmento de pendiente positiva. Los puntos máximos de respuesta de EDA y cardio se identificaron en la EW y se incluyeron en las mediciones extraídas si ocurrieron después de la EW y si el segmento de pendiente positiva comenzó durante el ROW.

Los segmentos medidos para las preguntas relevantes y de comparación se combinaron por la proporción R/C, que es la medición de la pregunta relevante extraída dividida por la medición de la pregunta de comparación extraída. Las relaciones R/C se registraron para que produjeran valores que fueran simétricos en torno a una media de cero.

Se asignaron las puntuaciones de enteros no paramétricos a las respuestas fisiológicas de los puntos de análisis (es decir, pares de preguntas relevantes y de comparación) utilizando una escala tipo Likert (1932) de tres posiciones [+1, 0 -1]. Se asignaron puntuaciones enteras de valor de signo positivo cuando el cambio en fisiológico fue mayor en la pregunta de

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

comparación seleccionada que en la pregunta relevante. Se asignaron puntuaciones enteras de valor de signo negativo cuando el cambio en la fisiología fue mayor en la pregunta relevante que en la pregunta de comparación seleccionada. Las puntuaciones con signo cero (cero numérico) se asignaron cuando los datos se distorsionaron por artefacto de movimiento/ actividad o fueron insuficientes para una puntuación debido a que no hubo respuesta o debido a un inicio de respuesta antes del inicio de la pregunta, sin inicio de respuesta durante una ventana de inicio de respuesta definida (ROW) desde el inicio de la pregunta hasta cinco segundos después de la respuesta verbal. Las respuestas electrodérmicas y cardiovasculares no se utilizaron si comenzaron durante un período de latencia de .5 segundos desde el inicio de la pregunta.

Las puntuaciones no paramétricas están destinadas a capturar y expresar información general sobre las diferencias en la respuesta ante preguntas relevantes y de comparación, pero no proporcionan información que pueda estar sujeta a supuestos lineales sobre el grado de diferencia en la respuesta. Sin embargo, se utilizaron algunas restricciones de umbral para evitar la asignación de puntajes numéricos en segmentos de datos que produjeron muy poca diferencia o diferencias extremas en las respuestas a las preguntas relevantes y de comparación. Las restricciones de umbral se determinaron a partir de experimentos de optimización con otra muestra de casos confirmados utilizada por Nelson (2018a) y se muestran en la Tabla 2. Las puntuaciones de enteros se asignaron cuando la relación registrada R/C estaba dentro de estas restricciones, y no se asignó ninguna puntuación cuando el valor era menor o mayor que estos umbrales de restricción.

Tabla 2. Restricciones de umbral para puntuaciones no paramétricas

-	Límite inferior		Límite superior	
Sensor	Proporción	Proporción registrada	Proporción	Proporción registrada
Respiratorio	1.05/1.25*	.049/.223	1.5	.406
Electrodérmico	1.05	.049	1000	6.908
Cardio	1.05	.049	1000	6.908

* Se utilizó una restricción asimétrica de 1.25 para el límite inferior de las puntuaciones respiratorias cuando se asignaron puntuaciones con signo de valor +. Los estudios de optimización con otros datos de muestreo indican que es más probable que las puntuaciones + estén correlacionadas negativamente con la veracidad sin la restricción asimétrica.

Reducción de datos

La reducción de los datos para cada uno de los casos de muestra se logró sumando las puntuaciones numéricas para los sensores de respiración, electrodérmicos y de cardio para todas las presentaciones de cada una de las preguntas de prueba relevantes. De esta manera, se obtuvo una puntuación numérica para cada una de las preguntas relevantes. Luego se sumaron las puntuaciones subtotales para obtener una puntuación de gran total para cada

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

uno de los casos de la muestra. Las clasificaciones de engaño y de veracidad se hicieron con puntajes de gran total y/o subtotales de acuerdo con las reglas de decisión poligráficas establecidas.

Análisis

Los casos de muestra se analizaron para determinar las decisiones correctas, los errores y los resultados inconclusos para los casos confirmados de campo y se calculó la sensibilidad de la prueba, la especificidad, los errores falsos negativos y los falsos positivos, y los resultados inconclusos para los subgrupos de culpables e inocentes. Además, se calculó la proporción de decisiones correctas para los casos de culpables e inocentes después de excluir resultados inconclusos, junto con el promedio no ponderado de precisión de decisión para los dos grupos. También se calculó el valor predictivo positivo y los valores predictivos negativos, así como los coeficientes de eficiencia de detección. Los índices de precisión se calcularon para varias condiciones, incluidas las puntuaciones de tres posiciones utilizando puntuaciones de corte numérico tradicionales, después de ponderar las puntuaciones EDA y utilizando puntuaciones de corte seleccionadas de una distribución de referencia multinomial de puntuaciones de tres posiciones ponderadas y no ponderadas (Nelson, 2017 , Nelson, 2018b).

Resultados

Los resultados se tabularon para puntajes de tres posiciones utilizando diferentes reglas de decisión y puntajes de corte numéricos tradicionales. Los resultados también se tabularon después de duplicar el valor de todas las puntuaciones electrodérmicas. Finalmente, los resultados se tabularon usando puntajes de corte que se seleccionaron de distribuciones de referencia multinomiales para puntajes de EDA ponderados y no ponderados de tres posiciones.

Reglas de Decisión

Estudios previos de Senter & Dollins (2003) han sugerido que la elección de las reglas de decisión puede desempeñar un papel importante en la efectividad de las clasificaciones poligráficas de engaño y veracidad. [Ver Nelson (2018c) para una discusión de las diferentes reglas de decisión]. La Tabla 3 muestra los resultados de las puntuaciones numéricas de tres posiciones para los n=100 casos confirmados de campo. Los resultados se muestran utilizando la Regla Federal de ZCT (FZR) con puntajes de corte numéricos tradicionales (descritos anteriormente). Los resultados para este estudio también se calcularon utilizando la Regla del Gran Total (GTR). El uso del GTR es una cuestión de sumar todos los puntajes numéricos y comparar el resultado con los puntajes numéricos de engaño o veracidad (tradicionalmente +6 y -6). Los resultados también se calcularon utilizando la regla de puntuación subtotal (SSR), para la cual el subtotal de la pregunta más baja (la indicación más fuerte de engaño) se

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

compara con las puntuaciones de corte numéricas de veracidad (tradicionalmente +3 o más para todos los subtotales) o engaño (tradicionalmente -3 o inferior para cualquier subtotal). También se muestran en la Tabla 3 los casos de muestra que utilizan la extracción de características automatizada y la asignación de puntajes utilizando el TSR con los puntajes de corte numéricos tradicionales y los puntajes de corte evidenciarios propuestos por Krapohl (2005), como se describió anteriormente.

Tabla 3. Resultados de muestra para las reglas de decisión con puntajes de tres posiciones y puntajes de corte tradicionales (n = 100).

	FZR	GTR	SSR	TSR*	EDR/TSR**
Sensibilidad (engaño)	.80	.56	.80	.80	.80
Especificidad (veracidad)	.32	.32	.06	.32	.48
Errores falso negativo	<.01	<.01	<.01	<.01	<.01
Errores falso positivo	.10	<.01	.10	.10	.10
Culpables inconclusos	.20	.44	.20	.20	.20
Inocentes inconclusos	.58	.68	.84	.58	.42
Precisión no ponderada	.88	>.99	.69	.88	.91
Tasa de Inconclusos no ponderados	.39	.56	.52	.39	.31
Valor predictivo positivo	.89	>.99	.89	.89	.89
Valor predictivo negativo	>.99	>.99	.86	>.99	>.99
Coeficiente de Detección de Eficiencia	.79	.67	.79	.76	.82

* Los resultados con el TSR se muestran utilizando los puntajes de corte numéricos tradicionales.

** En comparación con la Tabla 1, los resultados de estas reglas de decisión evidenciaria (EDR) con el TSR, se muestran utilizando puntajes de corte asimétricos utilizados por Krapohl (2005).

La sensibilidad de la prueba en la Tabla 3 es similar a la de la Tabla 1, aunque las tasas de especificidad son más bajas para todas las reglas de decisión. El PPV y el NPV para las puntuaciones de tres posiciones fueron mayores que los resultados en la Tabla 1, junto con tasas generalmente más altas de resultados no concluyentes. Los errores de FN para las puntuaciones automatizadas de tres posiciones fueron menores que en la Tabla 1. El DEC también se incrementó para el FZR y el TSR, y esto se puede atribuir a la reducción observada en los errores de FN. Es posible que estas diferencias se deban al uso de puntuaciones numéricas de tres posiciones frente a las siete posiciones, aunque se desconoce qué diferencias pueden deberse al uso de la extracción automática de características frente a la extracción de características visuales/subjetivas.

Los PPV y el NPV fueron los más altos para el GTR, aunque la tasa de inconclusos también fue mayor para esta regla de decisión. Light (1999) argumentó que la tasa de casos no inconclusos para el GTR no era aceptable cuando se usa en procuración de la ley. Sin

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

embargo, el estudio de Light tiene un alcance limitado, que involucra solo casos de culpabilidad confirmada, y la Tabla 3 muestra que las tasas de inconclusos se cargan hacia los casos de inocentes. Además, Light no incluyó una evaluación de errores PF o la efectividad de diferentes puntajes numéricos. Los resultados inconclusos para los casos de culpabilidad en la Tabla 3 son mayores que en la Tabla 1, y esto es más probable que se atribuya a las diferencias en las puntuaciones de siete y tres posiciones.

En el presente estudio, utilizando puntajes de tres posiciones y extracción de características automatizada, no está claro si el TSR proporciona alguna ventaja real - en términos de precisión de clasificación - sobre del FZR. Los resultados que se muestran en la Tabla 3 indican una reducción de los resultados inconclusos para casos de inocentes como resultado de mejores puntuaciones de corte numéricas con el TSR. DEC fue mayor para el TSR con las puntuaciones de corte numéricas mejoradas. Esto sugiere la posibilidad de que los puntajes de corte numéricos tradicionales para los puntajes de gran total son ineficientes para la escala reducida de tres posiciones.

Puntuaciones electrodérmicas ponderadas

Varias publicaciones han sugerido que los datos deL EDA representan una mayor porción de la varianza diagnóstica de los puntajes de pruebas (Ansley y Krapohl, 2000; Harris, Horner y McQuarrie, 2000; Harris y Olson, 1994; Kircher, 1981, 1983; Kircher, Kristjianson, Gardner & Webb, 2005; Kircher & Raskin, 1988; Krapohl & McManus, 1999; Nelson, Krapohl & Handler, 2008; Raskin, Kircher, Honts & Horowitz, 1988) y contribuyen con más información de conclusiones efectivas sobre el engaño o veracidad en comparación con los demás sensores de registro. Para observar las diferencias en los resultados de la muestra con las puntuaciones de tres posiciones, las puntuaciones EDA se duplicaron en valor de la manera descrita anteriormente por Krapohl y McManus (1999). La tabla 4 muestra los resultados de la muestra después de ponderar las puntuaciones electrodérmicas.

Tabla 4. Resultados de la muestra para las reglas de decisión con n=100 casos de campo confirmados con puntajes de corte tradicionales y puntajes electrodérmicos ponderados.

Especificidad	FZR	GTR	SSR	TSR*	EDR/TSR**
Sensibilidad (engaño)	.92	.66	.92	.92	.92
(veracidad)	.52	.54	.16	.54	.68
Errores Falso negativo	<.01	<.01	<.01	<.01	.02
Errores Falso positive	.24	<.01	.24	.22	.14
Culpables inconclusos	.08	.34	.08	.08	.06
Inocentes inconclusos	.24	.46	.60	.24	.18
Precisión no ponderada	.84	>.99	.70	.86	.90
Inconclusos no ponderados	.16	.40	.34	.16	.12
Valor Predictivo Positivo	.79	>.99	.79	.81	.87

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

Valor Predictivo Negativo	>.99	>.99	>.99	>.99	.97
Coeficiente de Eficiencia de Decisión	.87	.78	.86	.88	.90

* Los resultados con el TSR se muestran utilizando los puntajes de corte numéricos tradicionales.

** En comparación con la Tabla 1, los resultados de estas reglas de decisión probatoria (EDR) con el TSR se muestran usando puntajes de corte asimétricos utilizados por Krapohl (2005).

Al ponderar las puntuaciones del EDA más que las puntuaciones de otro sensor, se aumentó la sensibilidad y especificidad de la prueba y se redujo la aparición de resultados inconclusos en un 59% para el FZR, 29% para el GTR y 35% para el SSR. La reducción de los resultados inconclusos fue del 59% para el TSR con puntajes de corte tradicionales y del 61% para el TSR utilizando puntajes de corte que se sugirieron como óptimos para los exámenes evidenciarios. La reducción de los resultados inconclusos fue mayor para los casos inocentes que los culpables para todas las reglas de decisión.

El GTR proporcionó la mayor precisión general de clasificación, aunque los errores FN fueron bajos para todas las reglas de decisión. El efecto más obvio de la ponderación de las puntuaciones del EDA en la Tabla 4, en comparación con la Tabla 3, fue una reducción de los resultados inconclusos, junto con aumentos en la sensibilidad de la prueba al engaño y la especificidad a la veracidad. La reducción de los resultados inconclusos fue mayor en los casos inocentes confirmados. Curiosamente, el PPV se redujo para todos los modelos, excepto el GTR, que produjo tasas de error FN y FP más bajas que otras reglas de decisión. El DEC mejoró para todas las reglas de decisión y fue mayor para el TSR con mejores puntajes de corte. Esto sugiere que la selección de puntajes de corte puede ser una consideración importante en el manejo y la reducción de errores de clasificación o de resultados inconclusos.

Puntuaciones de corte multinomial

Puntuaciones de corte multinomial con puntuaciones ponderadas de tres posiciones

Las puntuaciones de tres posiciones ponderadas se evaluaron utilizando puntuaciones de corte seleccionadas de las distribuciones de referencia multinomiales descritas por Nelson (2017). [Vea Nelson (2018b) para una discusión sobre cómo usar las distribuciones de referencia multinomiales.] Los puntajes de corte multinomiales para tres posiciones ponderadas de polígrafos de eventos específicos con tres preguntas relevantes fueron los siguientes: total general >= +3 para clasificaciones veraces o <= -3 para clasificaciones de engaño. Las clasificaciones de engaño se realizaron utilizando las puntuaciones subtotales cuando la puntuación total fue inconclusa, si alguna puntuación subtotal <= -7. La puntuación de corte para las puntuaciones subtotales se determinó mediante una corrección estadística de los efectos de multiplicidad para evitar la posible inflación de los errores PF cuando se utilizan puntuaciones subtotales múltiples para clasificaciones de engaño. Los resultados se muestran en la Tabla 5 para la combinación de puntuaciones de EDA ponderadas y puntuaciones de corte multinomiales.

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

Tabla 5. Resultados de la muestra (n=100) que utilizan puntuaciones de tres posiciones ponderadas con puntuaciones de corte multinomiales.

	FZR	GTR	SSR	TSR
Sensibilidad (engaño)	.92	.88	.92	.92
Especificidad (veracidad)	.80	.80	.34	.80
Errores Falso negative	.04	.04	.02	.04
Errores Falso positive	.08	.06	.24	.08
Culpables inconclusos	.04	.08	.06	.04
Inocentes inconclusos	.12	.14	.42	.12
Precisión no ponderada	.93	.94	.78	.93
Inconclusos no ponderados	.08	.11	.24	.08
Valor Predictivo Positivo	.92	.94	.79	.92
Valor Predictivo Negativo	.95	.95	.94	.95
Coeficiente de Eficiencia de Detección	.94	.94	.85	.94

El uso de puntuaciones de corte multinomiales mejoró aún más la sensibilidad de la prueba al engaño para el GTR, y mejoró la especificidad de la prueba a la veracidad para el FZR, GTR, SSR y TSR. En comparación con el uso de puntajes electrodérmicos ponderados con puntajes de corte numéricos tradicionales, los puntajes de corte multinomiales produjeron una reducción en los resultados inconclusos en un 50% para el FZR, 73% para el GTR, 29% para el SSR y 33% para el TSR. Los DECs para las puntuaciones de corte multinomiales fueron consistentemente mayores que para las puntuaciones de corte numéricas tradicionales.

No sorprende que la SSR haya mostrado una alta sensibilidad al engaño, aunque no mayor que cualquiera de las otras reglas de decisión, junto con una especificidad más débil a la veracidad y una tasa de inconclusos más alta cargada hacia los casos inocentes. La precisión de la decisión general y el DEC para la SSR fue más bajo y las tasas de inconclusos fueron más altas que en las otras reglas de decisión que incluyeron el uso del puntaje de gran total. Esta diferencia puede atribuirse a la multiplicidad inherente cuando se usan puntuaciones subtotales, y también al menor volumen de información disponible para respaldar las decisiones basadas en las puntuaciones de subtotales individuales.

Puntuaciones de Corte Multinomial con Puntuaciones de Tres Posiciones no Ponderadas

Las puntuaciones de tres posiciones también se evaluaron utilizando puntuaciones de corte obtenidas de una distribución de referencia multinomial para exámenes poligráficos con tres preguntas relevantes y tres a cinco gráficos. Las puntuaciones de corte multinomiales para puntuaciones de tres posiciones no ponderadas fueron las siguientes: gran total >= +2 para clasificaciones veraces o <= -2 para clasificaciones de engaño. Las clasificaciones de engaño se realizaron utilizando las puntuaciones subtotales cuando la puntuación de gran total fue inconclusa, si alguna puntuación subtotal era <= -6. La puntuación de corte para las

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

puntuaciones subtotales se determinó mediante una corrección estadística de los efectos de multiplicidad para evitar la posible inflación de los errores PF cuando se usan puntuaciones de subtotales múltiples para clasificaciones de engaño. Los resultados se muestran en la Tabla 6 para la combinación de puntuaciones de EDA ponderadas y puntuaciones de corte multinomiales.

Tabla 6. Resultados de la muestra (n = 100) que utilizan puntuaciones de tres posiciones no ponderadas con puntuaciones de corte multinomiales.

	FZR	GTR	SSR	TSR
Sensibilidad (engaño)	.88	.86	.80	.88
Especificidad (veracidad)	.78	.76	.30	.78
Errores Falso negativo	.04	.04	.02	.04
Errores Falso positivo	.04	.04	.10	.04
Culpables inconclusos	.08	.10	.18	.08
Inocentes inconclusos	.20	.20	.60	.20
Precisión no ponderada	.95	.95	.86	.95
Inconclusos no ponderados	.14	.15	.39	.14
Valor Predictivo Positivo	.96	.96	.89	.96
Valor Predictivo Negativo	.95	.95	.94	.95
Coeficiente de Eficiencia de Detección	.91	.90	.79	.91

El uso de puntuaciones de corte multinomiales mejoró la efectividad de las clasificaciones con las puntuaciones de tres posiciones no ponderadas. La precisión de la clasificación con las puntuaciones de tres posiciones utilizando puntuaciones de corte multinomiales fue similar a la del modelo multinomial ponderado. Sin embargo, tanto la sensibilidad como la especificidad se redujeron ligeramente para las puntuaciones de tres posiciones no ponderadas. En comparación con el modelo ponderado, la tasa de inconclusos para puntuaciones de tres posiciones no ponderadas aumentó en promedio en un 62% para todas las reglas de decisión. El aumento de resultados inconclusos se cargó en casos de inocentes. Las puntuaciones de tres posiciones no ponderadas también produjeron menos errores falsos positivos. Estas diferencias pueden parecer sustanciales cuando se describen por el porcentaje de cambio. El SSR tuvo un relativo rendimiento inferior con respecto a otras reglas de decisión, con una especificidad de prueba más débil y resultados inconclusos más altos. La precisión general de detección para el modelo multinomial de tres posiciones fue alta, aunque no fue igual a la efectividad del modelo ponderado de tres posiciones.

Resumen

Este proyecto involucró el cálculo de estadísticas descriptivas para la precisión de la prueba, error y tasas de inconclusos en función de las diferentes reglas de decisión, de la

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

ponderación estructural de las puntuaciones de los sensores y de las puntuaciones de corte. Los resultados muestran que una serie de decisiones de procedimientos y de prácticas de campo pueden tener un impacto importante en el criterio de precisión de los resultados de las pruebas poligráficas. Aunque el PPV y el NPV fueron consistentemente altos para la mayoría de las condiciones, se pueden observar diferencias en la tasa de resultados inconclusos y esto puede afectar directamente el poder de la prueba en términos de sensibilidad, especificidad y tasas de error de prueba.

Los resultados para las puntuaciones de tres posiciones, que se muestran en la Tabla 3, fueron similares para el FZR y el TSR. Sin embargo, hubo una pequeña reducción de los resultados inconclusos, al tiempo de un aumento correspondiente en la especificidad de la prueba al usar el TSR con puntajes de corte que se sugirieron como óptimos para las pruebas de poligráficas evidenciarias. La ponderación de los puntajes EDA, más que los puntajes de otros sensores, que se muestran en la Tabla 4, incrementó la sensibilidad y especificidad de la prueba, y redujo la aparición de resultados inconclusos en un promedio del 49%, en comparación con los resultados de los puntajes de tres posiciones no ponderados, para todas las reglas de decisión. El uso de puntuaciones de corte multinomiales produjo reducciones adicionales en la aparición de resultados inconclusos junto con aumentos adicionales en los DECs para todas las reglas de decisión.

La combinación de puntajes de EDA ponderados y puntajes de corte multinomiales redujo la aparición de resultados inconclusos en un promedio del 72% en todas las reglas de decisión, en comparación con los puntajes de tres posiciones no ponderados y puntajes de corte numéricos tradicionales. La reducción en los resultados inconclusos fue mayor para el GTR, que no utiliza puntuaciones subtotales, y para el cual la tasa de resultados inconclusos con puntuaciones de corte multinomiales fue más cercana a la de otras reglas de decisión.

> Los profesionales de campo han proporcionado información anecdótica que sugiere que sus tasas observadas de las tasas de inconclusos son inconsistentes y más bajas que las de los estudios publicados. Esto es comprensible porque los profesionales de campo, que trabajan a nivel de casos individuales, pueden estar éticamente justificados al participar en prácticas destinadas a resolver o reducir la ocurrencia de resultados inconclusos (por ejemplo, realizar repeticiones adicionales de la secuencia de preguntas o repetir un examen). En contraste, los investigadores que trabajan con muestras de casos serían vulnerables a las sugerencias de manipular el resultado de una investigación si se comprometieran en tales acciones a nivel de algunos casos individuales, aunque no de todos. El resultado es que las tasas de inconclusos en la práctica de campo pueden seguir siendo más bajas que las informadas en estudios publicados.

El modelo más efectivo en este análisis - ilustrado por sensibilidad, inconclusos, y DECs - en la Tabla 5, fue con el TSR utilizando EDA ponderado y puntajes de corte multinomiales. Curiosamente, las precisiones para el FZR y el TSR fueron efectivamente idénticas para las puntuaciones ponderadas de tres posiciones con puntajes de corte multinomiales, lo que sugiere que la selección de los puntajes de corte puede ser más importante que las reglas de

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

decisión. La similitud del DEC para el GTR, FZR y TSR proporciona una indicación adicional de esto, y sugiere que algunas conclusiones previamente informadas sobre el GTR pudieron estar indebidamente influenciadas por la dependencia en los puntajes de corte numéricos tradicionales que eran ineficientes para los puntajes de gran total.

Este proyecto, como todos los proyectos, no está exento de algunas limitaciones. La limitación más obvia es el tamaño pequeño de la muestra (N=100). Aunque de tamaño moderado para un proyecto de este tipo, es axiomático que los tamaños de muestra más grandes se verían más fácilmente como una aproximación más cómoda de la población. Sin embargo, el tamaño de la muestra no es la única consideración principal al tratar de comprender la representatividad de una muestra, para la cual la selección aleatoria puede ser más importante. Este proyecto, que involucró una muestra de archivo, está necesariamente excluido de cualquier influencia debido a que la metodología de muestreo no está presentemente expresada en el muestreo de los datos. También es necesariamente dependiente de los supuestos, que los datos de muestreo son de alguna manera informativos.

Otra limitación importante de este proyecto es que no se completaron pruebas de significancia estadística. Esto fue así por diseño, ya que se esperaba que un enfoque descriptivo del análisis estadístico pudiera ser de mayor valor práctico para los examinadores de campo y directores de programas poligráficos que pueden estar más familiarizados y versados con las decisiones de las políticas de práctica de campo, que el ANOVA múltiple. Investigaciones futuras deben incluir un análisis más completo de la varianza de los tamaños del efecto relacionados con las reglas de decisión poligráficas, los coeficientes de ponderación estructural para las puntuaciones de los sensores y los puntajes de corte. Además, no se incluyeron intervalos de confianza estadísticos en este documento, aunque los lectores informados pueden usar fácilmente una serie de métodos para calcular los intervalos de confianza de interés.

Este proyecto involucró solo una muestra de campo de polígrafos de investigación criminal confirmados (eventos específicos) y no incluyó una muestra de polígrafos exploratorios de asuntos múltiples. Sugerimos que una generalización cautelosa de estos resultados todavía está recomendada. Esto se debe a la diferencia práctica e importante entre los polígrafos de diagnóstico de eventos específicos y los polígrafos exploratorios que involucran suposiciones sobre la independencia de las preguntas exploratorias de asuntos múltiples. Estas suposiciones son las mejores *suposiciones de conveniencia* porque asumen independencia en el hecho de que diferentes elementos de prueba no tienen una fuente compartida de varianza de respuesta - todo lo que podría influir en las respuestas de un elemento podría no haber afectado a ningún otro elemento. Como sucede, todas las preguntas poligráficas dentro de cualquier examen siempre tendrán una fuente compartida de varianza de respuesta, en la forma de atención del examinado. Tanto los polígrafos diagnósticos de eventos específicos como de exploración de asuntos múltiples también se verán influenciados por los efectos de la multiplicidad estadística determinados por la selección de las reglas de decisión poligráficas. Por estas razones, se puede esperar un patrón similar de resultados para los exámenes de poligráficos exploratorios, como se observa con esta muestra de datos.

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

Los lectores astutos notarán que este proyecto no intenta discutir todos los métodos posibles para reducir las tasas de inconclusos y mejorar la efectividad de la prueba poligráfica. Algunos de esos otros métodos pueden incluir: enfoques de entrevista, actividades de control de calidad, mayor uso de la automatización, uso del sensor vasomotor, registro de gráficos adicionales, aclaración de definiciones operacionales, uso de mapas de ruta en la entrevista, selección de objetivos refinada y/o mejoras en la formulación de preguntas. Todos estos deben permanecer como áreas de investigación y desarrollo continuos.

Los resultados de este estudio apuntan claramente al hecho de que los puntajes de corte numéricos tradicionales son efectivos para producir una tasa de FN baja, pero se cargan hacia una especificidad de prueba innecesariamente débil para veracidad y tasas innecesariamente altas de resultados inconclusos. Se puede hacer una observación interesante acerca de estos resultados y es que no hubo ninguna ventaja en el uso de la SSR en términos de una mayor sensibilidad de la prueba al engaño, en comparación con las otras reglas de decisión. La tasa de FN para el SSR fue igual a la de las otras reglas de decisión cuando se utilizaron las puntuaciones de corte tradicionales y se redujo la del TSR y del FZR en un 50% (.02 / .04), mientras que la tasa de FP aumentó en un factor de 3 (.24 / .08) para el SSR. La tasa de inconclusos fue mayor para el SSR que para el FZR y el TSR y se cargó en casos inocentes. Como se muestra en la Tabla 3, el SSR fue especialmente débil con puntuaciones de tres posiciones no ponderadas y puntajes de corte numéricos tradicionales. La implicación práctica de estas observaciones es que puede ser difícil justificar el uso del SSR fuera de los contextos exploratorios del polígrafo, donde es posible que sea deseable o se pretenda un exceso de predicción - y es difícil justificar el uso del SSR cuando se usa un puntaje no ponderado y puntajes tradicionales.

Estos resultados muestran claramente que la optimización de las prácticas de campo en cada una de estas áreas - reglas de decisión, ponderación de los puntajes del EDA y la selección de puntajes de corte - puede brindar importantes ventajas a muchos, incluyendo a los examinadores poligráficos de campo, directores de programas, tribunales, legisladores e investigadores y poligrafistas. Se necesita más exploración para comprender mejor las funciones de utilidad en términos de valores económicos y costos operativos asociados con la sensibilidad, especificidad, errores de FN y FP, y tasas de resultados inconclusos. Una mayor dependencia de la teoría de la medición estadística puede permitir a los programas de polígrafo refinar sus políticas para un mejor logro de sus objetivos y metas de misión.

Es probable que los resultados inconclusos de las pruebas persistan como una pesadilla para los examinadores de campo del polígrafo, los directores de programas y de otros, incluidos los examinados del polígrafo. La disponibilidad de soluciones de procedimiento basadas en evidencia que pueden reducir de manera confiable la aparición de resultados de pruebas no concluyentes parece merecer una mayor atención y consideración.

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

Referencias

Ansley, N. & Krapohl, D.J. (2000). The frequency of appearance of evaluative criteria in field polygraph charts. *Polygraph , 29,* 169-176.

Bradley, M. T. & Janisse, M.P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18,* 307-315.

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].

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.

Harris, J. C. & Olsen, D.E. (1994). Polygraph Automated Scoring System. Patent Number: 5,327,899.
U.S. Patent and Trademark Office.

Kircher, J. C. (1981). *Computerized chart evaluation in the detection of deception.* University of Utah. [Thesis],

Kircher, J. C. (1983). *Computerized decision making and patterns of activation in the detection of deception.* Doctoral dissertation, University of Utah, Salt Lake City. Dissertation Abstracts International, 44, 345.

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

Kircher, J. C., Kristjiansson, S. D., Gardner, M. K. & Webb, A. (2005). *Human and computer decision- making in the psychophysiological detection of deception.* University of Utah.

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

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

Krapohl, D. J. (2005). Polygraph decision rules for evidentiary and paired testing (Marin protocol) applications. *Polygraph, 34,* 184-192.

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

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

Light, G. D. (1999). Numerical evaluation of the Army zone comparison test. *Polygraph, 28,* 37-45.

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

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

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

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

Nelson, R. (2018b). Guide for how to use the ESS-multinomial reference tables in four steps. *APA Magazine, 51(2),* 78-89.

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

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

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.

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.

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

Summers, W. G. (1939). Science can get the confession. *Fordham Law Review, 8,* 334-354.

Van Herk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. *The Newsletter of the Canadian Association of Police Polygraphists, 7,* 28-47. Reprinted in Polygraph, 20(2), 70-7

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