

Título: FORMATO DIAGNÓSTICO EXPERIMENTAL EN UNA SOLA SECUENCIA

Rodolfo Prado
International Polygraph Training Center

Carlos Grajales

Práctica Privada

Raymond Nelson

Práctica Privada

Estudio de Laboratorio de Polígrafos Diagnósticos con Repeticiones
Múltiples en una Sola Secuencia

Nos gustaría agradecer a los examinadores que participaron en este estudio; Julissa Gómez Varela, Mitzy Betzabeth Torres Paredes, Ricardo Reyes Monsalvo, Juan Carlos Martínez Carvajal, Miguel Ángel Suárez Espitia, Raúl Dirceu García Delgado, Juan Carlos Ortiz Sierra, Aurora Santillán González, Daniel Arturo Saucedo Torres y Jesús Castelán Ramírez.

Comunicaciones relacionadas con este artículo deberán ser dirigidas a Rodolfo Prado, contacto: 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

ABSTRACT

Un estudio de crimen simulado aplicado con población Mexicana, puso a prueba la efectividad de una técnica poligráfica experimental diagnóstica de evento-específico con una sola secuencia, con dos preguntas relevantes. El protocolo experimental tuvo una precisión no ponderada de 87%, una tasa de inconclusos de 18%, una sensibilidad de 80% y una especificidad de 93%. Su confiabilidad vía el estadístico Kappa fue de 0.73. Los resultados del estudio sugieren una precisión mayor que el azar, que es consistente con otras técnicas existentes. Aunque no se observaron ventajas en este formato experimental, estos resultados agregan soporte a la efectividad de los formatos poligráficos de una sola secuencia y a la efectividad similar de las exámenes poligráficas conducidas en español con aquellas realizadas en inglés.

Palabras clave: polígrafo, detección de mentiras, exámenes de diagnóstico, preguntas de comparación de mentira dirigida,

Estudio de Laboratorio de Polígrafos Diagnósticos con Repeticiones Múltiples en una Sola Secuencia

Este proyecto se diseñó para estimar la precisión diagnóstica de un formato experimental de prueba de evento-específico con preguntas de comparación (CQT) en un ambiente de laboratorio. Los formatos CQT han sido utilizados tanto en aplicaciones diagnósticas como exploratorias. Las pruebas diagnósticas de evento-específico (asunto único) involucran formatos de preguntas para los que el criterio de estado de las múltiples preguntas relevantes se interpreta bajo el supuesto de no-independencia. Esto significa que todos los estímulos objetivo tienen una fuente compartida de varianza, ya que se refieren a un solo incidente conocido o supuesto. Las pruebas exploratorias, que se conducen en ausencia de un incidente conocido o supuesto, a menudo se formulan utilizando técnicas poligráficas diseñadas para ser interpretadas bajo la suposición de que el criterio de estado de las múltiples preguntas relevantes variará de forma independiente. Los exámenes de asuntos múltiples son más complejos en términos de la teoría de probabilidad y de las demandas atencionales depositadas en el sujeto; y se ha encontrado que son menos precisas que las pruebas en las que las preguntas objetivo no son independientes. Debido a que el formato experimental fue diseñado para exámenes diagnósticos, el diseño experimental fue designado para la interpretación de múltiples preguntas relevantes dentro de cada examen, bajo el supuesto de no-independencia.

Una característica de la mayoría de los formatos CQT es que las respuestas a las repeticiones múltiples de la secuencia de estímulos de prueba, se obtienen en diferentes secuencias de registro. El presente estudio fue destinado para investigar el uso de un formato de una sola secuencia, para polígrafos diagnósticos de evento específico. La ventaja potencial de un formato de una sola secuencia es la reducción de fuentes no controladas de varianza, asociadas con los procedimientos de repetición de la secuencia de preguntas; con la esperanza de que esto podría generar a una mayor precisión de los resultados de prueba. Actualmente, los formatos CQT de una sola secuencia se utilizan eficientemente en pruebas poligráficas exploratorias (APA, 2011; Department of Defense, 2006; Handler, Nelson & Blalock, 2008; Research Division Staff, 1995a; 1995b). No se encontró información publicada acerca de la eficiencia de los

formatos poligráficos diagnósticos que utilizan repeticiones múltiples de los estímulos de prueba en un solo registro de secuencia.

En ocasiones, incorrectamente se hace referencia del polígrafo como un "detector de mentiras", con el malentendido de que el instrumento mostrará reacciones únicas cuando un sujeto mienta. Por sí mismos, los instrumentos poligráficos no miden mentiras. En su lugar, miden las diferencias en la reacción fisiológica ante dos diferentes tipos de preguntas de estímulo de prueba: las preguntas relevantes (RQ) y las preguntas de comparación (CQ). Se ha demostrado que las diferencias en la reacción varían como una función del engaño en respuesta a las RQs que describen el posible involucramiento conductual del sujeto en un asunto de interés (APA, 2011; Honts y Reavy, 2009; Kircher y Raskin, 1988; Kircher, Kristjansson, Gardner y Webb, 2005; National Research Council, 2003). Generalmente, las RQs evitan asuntos relacionados con la memoria, la intención y la motivación (APA, 2009a).

Las preguntas de comparación pueden presentarse como Comparativas de Mentira Probable (PLC) en las que el sujeto es manejado hacia una respuesta verbal en la que se asume que el sujeto probablemente está mintiendo, y también se pueden presentar como preguntas Comparativas de Mentira Dirigida (DLC) en las que el sujeto recibe instrucciones procesales por parte del examinador, para que responda a la preguntas DLC con mentira (Raskin y Honts, 2002). Las PLCs y DLCs han sido descritas con eficiencia similar, aunque se ha sugerido que las DLCs son éticamente menos cuestionables ya que su uso no implica la manipulación psicológica del sujeto. (Honts y Reavy, 2009; Kircher, Packard Bell, y Bernhardt, 2001). Este estudio incluyó solamente el uso de preguntas DLC dentro del formato de prueba experimental.

El diseño de la técnica experimental fue un formato con preguntas comparativas de evento-específico de asunto-único, que consta de dos RQs, ambas abordando el mismo asunto. Aun cuando las preguntas fueron redactadas con pequeñas diferencias, el significado de ambas preguntas fue el mismo, y ambas preguntas describían el mismo comportamiento. Dentro de la secuencia de preguntas se encontraban incluidas dos preguntas DLC, una pregunta neutral y una pregunta relevante de sacrificio. Todas las preguntas fueron presentadas a los sujetos en una sola secuencia, con cuatro repeticiones para cada RQ. Esto es diferente a la mayoría de las técnicas poligráficas, en las que el

examinador recolecta datos de prueba durante varias repeticiones de la secuencia de estímulos de prueba, mediante la detención de la grabación fisiológica y desinflando la manga de cardio entre cada secuencia.

Las posibles ventajas de un formato poligráfico diagnóstico en una sola secuencia, involucran la posibilidad de incrementar la precisión de la prueba, como resultado de la reducción de las fuentes de varianza no controlada en los datos registrados que podría presentarse como resultado de las prácticas tradicionales, que involucran secuencias de registro por separado para cada repetición de los estímulos de prueba. Las categorías de resultados de prueba, que a menudo se reportan utilizando los términos "Engaño Indicado" o "No se Indica Engaño," se basan en la existencia de una respuesta estadísticamente significativa al comparar el puntaje de la prueba con los puntajes de corte que fueron establecidos previamente, utilizando tablas calculadas para las distribuciones de referencia estadística de los sujetos culpables e inocentes dentro de los datos de muestreo normativos.

Métodos

Los participantes del estudio fueron voluntarios mexicanos quienes aceptaron participar al dar respuesta a un anuncio en un periódico mexicano muy conocido. Los voluntarios serían recompensados con \$ 200.00 pesos mexicanos (MXN) por su participación. Esto era equivalente a 3 días de salario mínimo, a cambio de participar en un estudio que tomaría de 3 a 4 horas. Los interesados llamaron un número telefónico y entonces recibieron una cita para la selección inicial con un monitor de examinados, cuya función era facilitar los participantes durante el estudio. Los criterios de inclusión para participar, requirieron cuatro condiciones mínimas. Se pidió a los participantes:

1. ser mayores de edad (18 años).
2. no estar bajo la influencia de alcohol o drogas.
3. no encontrarse excesivamente cansados al tiempo de la prueba y
4. no estar padeciendo hambre al tiempo de la prueba.

Todos los participantes recibieron una hoja de consentimiento, informándoles los requisitos y detalles experimentales del estudio. Los participantes fueron informados de su derecho a poner fin a su participación en el estudio en cualquier momento, sin penalidad, y de que se les pagaría la cantidad anunciada a pesar de que se retiraran del estudio. Se informó también a los participantes de que no sufrirían daño alguno como resultado de su participación. Participaron en el estudio un total de 114 voluntarios.

A los participantes se les asignó al azar un estatus de culpable o de inocente. De los 114 participantes originales, 57 fueron designados como "inocentes" y 57 como "culpables". Se les pidió a los participantes culpables participar en una actividad que programaría su estatus de culpable para el estudio. Se dio instrucción a los participantes inocentes para involucrarse en una actividad que aseguraría su estatus de inocente con respecto a las RQs.

La carta para el estatus de "culpable" daba instrucciones que se resumen de la siguiente manera:

Vaya hacia el comedor. En la mesa del café se encuentra una mochila roja, justo debajo de la mesa. Sírvese una taza de café y "robe" \$ 100 (pesos mexicanos / MXN) que va a encontrar dentro de la mochila. Después de que robe el dinero, tendrá que pasar por la prueba de polígrafo (detector de mentiras), y el asunto bajo investigación será el robo de ese dinero.

A los participantes de este grupo se les dijo que durante la prueba tendrían que negar el haber robado ese dinero, y que si durante la prueba no era descubierto su involucramiento, serían recompensados con un bono de \$ 50 pesos, más los \$ 200 pesos originales, además de la oportunidad de poder mantener los \$ 100 pesos que tomó de la mochila. Se les dijo también que si era descubierto su involucramiento, sólo recibirían el pago original por ser voluntario.

Con el fin de confirmar que las instrucciones eran entendidas, el monitor pedía a los voluntarios que revisaran las instrucciones. Todos los sujetos fueron capaces de describirle adecuadamente la tarea al monitor, quien no tuvo contacto con los examinadores durante el proyecto.

La carta para el estatus de "inocente" daba las siguientes instrucciones:

Vaya hacia el comedor. Sírvese una taza de café y espere a que alguien lo llame. A continuación, tendrá que pasar por una prueba de polígrafo (detector de mentiras). El asunto bajo investigación es el robo de cierta cantidad de dinero. Durante esa prueba solamente tendrá que negar honestamente su involucramiento en el crimen, porque usted no tomó ningún dinero.

A los sujetos con estatus de "inocentes" también se les dijo que si pasaban la prueba y demostraban su inocencia, serían recompensados con un bono de \$ 50 pesos adicionales. También se les dijo que en caso de que el resultado de la prueba, de alguna manera los mostrara como culpables, solamente recibirían los \$200 pesos originales de recompensa por aceptar participar en el proyecto.

Los datos fueron recolectados entre el 04 y 14 de diciembre del 2012 utilizando la muestra de conveniencia descrita de participantes de la comunidad. Se utilizaron Instrumentos poligráficos modelo LX4000 Lafayette para registrar actividad electrodérmica (EDA), movimientos respiratorios, actividad cardiovascular y actividad voluntaria. La EDA se midió utilizando la resistencia de la piel, medida por los electrodos estándar de metal Lafayette colocados en las falanges mediales en los dedos primero y segundo. El movimiento respiratorio torácico y abdominal se midió mediante el juego de tubos neumáticos estándar Lafayette. Las respuestas cardiovasculares se registraron mediante el set de presión sanguínea Lafayette a una presión de 80 a 90 mmHg y se colocó en la pantorrilla del participante. Se colocó un sensor de actividad en el asiento del sujeto.

El estudio se realizó en dos instalaciones diferentes: Cuando se habló con los participantes, la primera ubicación - Instalación A - fue denominada como "Casa N ° 3", y consistió en una oficina utilizada por el monitor para recibir los participantes voluntarios y para asignar y programar los estatus de culpable o inocente. Fue programado un participante a la vez, y los participantes no interactuaron entre sí. Después de la programación, el coordinador del estudio (autor principal) escoltaba a cada participante a la segunda instalación, denominado "Instalación B", donde el examinador llevaba a cabo la prueba. El coordinador se mantuvo

ciego con respecto al estatus de culpable o inocente de los participantes, hasta después de que se recolectaron y analizaron todos los datos.

Se contó con 15 examinadores, trabajando en cinco salas de evaluación poligráfica. El coordinador asignaba cada participante a un examinador y a una sala de examinación en el orden de aparición. Al final de la prueba el coordinador acompañaba al sujeto a la sala de espera en la Instalación A. Cada examen fue evaluado por el examinador y después por un revisor de control de calidad. Se determinó que algunos exámenes eran inutilizables cuando el revisor de control de calidad notaba alguna de las siguientes violaciones de protocolo:

- Enfermedad física o achaque en el sujeto.
- Sujetos culpables que no negaron apropiadamente el robo, es decir, confesaron el crimen.
- Examinadores que no utilizaron correctamente el formato experimental.
- Pruebas interrumpidas o incompletas.
- Datos de prueba no interpretables.

Una vez terminada la prueba y la revisión de control de calidad, se le proporcionaban los resultados de prueba al monitor, quien mantuvo la posesión de la información relacionada con el criterio de estado real de cada participante. Después de comparar el resultado de la prueba con el criterio de estado, el monitor recompensó a cada participante, quien después abandonaba las instalaciones. Debido a que los participantes se retiraban después de la finalización de los exámenes, ninguno de los participantes fue sometido a reexaminación cuando el examen resultó en una violación del protocolo del estudio.

Formato de prueba Experimental

El formato poligráfico experimental se compuso de dos RQs repetidas cuatro veces en una sola secuencia. Los datos deberían calificarse utilizando el EES, utilizando sólo las primeras tres presentaciones útiles de cada pregunta estímulo. La cuarta presentación

se utilizaría solamente en caso de que alguna de las presentaciones de estímulo previas fuera inutilizables debido a un artefacto o por tener reacción fuera de tiempo, y cuando los resultados en las calificaciones de las primeras tres presentaciones no fueran concluyentes. La información con respecto a las preguntas, su orden y su tipo se resume, en la Tabla 1.

El formato de la prueba incluyó:

- 2 preguntas neutrales en las posiciones 2 y 8 y que se repitieron en las posiciones 13 y 18
- 1 pregunta relevante de sacrificio en la posición 3
- 2 preguntas relevantes en las posiciones 5 y 7 (primera presentación), repetidas en las posiciones 10 y 12 (segunda presentación), 15 y 17 (tercera presentación), y finalmente en las posiciones 20 y 22 (cuarta presentación).
- 3 preguntas de comparación de mentira dirigida en las posiciones 4, 6 y 9, repetidas en 11, 14 y 16 (segunda presentación), y por último en las posiciones 19, 21 y 23 (tercera presentación).

Tabla 1. Preguntas presentadas durante el formato.

#	ID	Tipo	Texto	Respuesta
1	X		La prueba está por comenzar, por favor no se mueva y responda con un sí o un no a cada pregunta.	
2	1N1	N	¿Estamos en el año 2012?	Sí
3	SR	SR	Con respecto al robo de ese dinero, ¿intenta contestar con la verdad a cada una de las preguntas?	Sí
4	1C1	C	¿Alguna vez le mintió a alguien que confiaba en usted?	No
5	1R1	R	El día de hoy, ¿tomó usted cualquier cantidad de dinero de una mochila en la casa 3?	No
6	1C2	C	¿Alguna vez le robó algo a alguien que confiaba en usted?	No
7	1R2	R	El día de hoy, ¿tomó cualquier cantidad del dinero reportado robado de una mochila roja?	No
8	1N2	N	¿Estamos en la Ciudad de México?	Sí
9	1C3	C	¿Alguna vez hizo algo que lo haría sentirse avergonzado en frente de su familia?	NO

10	2R1	R	El día de hoy, ¿tomó usted cualquier cantidad de dinero de una mochila en la casa 3?	No
11	2C1	C	¿Alguna vez le mintió a alguien que confiaba en usted?	No
12	2R2	R	El día de hoy, ¿tomó cualquier cantidad del dinero reportado robado de una mochila roja?	No
13	2N1	N	¿Estamos en el año 2012?	Sí
14	2C2	C	¿Alguna vez le robó algo a alguien que confiaba en usted?	No
15	3R1	R	El día de hoy, ¿tomó usted cualquier cantidad de dinero de una mochila en la casa 3?	No
16	2C3	C	¿Alguna vez hizo algo que lo haría sentirse avergonzado en frente de su familia?	No
17	3R2	R	El día de hoy, ¿tomó cualquier cantidad del dinero reportado robado de una mochila roja?	No
18	2N2	N	¿Estamos en la Ciudad de México?	Sí
19	3C1	C	¿Alguna vez le mintió a alguien que confiaba en usted?	No
20	4R1	R	El día de hoy, ¿tomó usted cualquier cantidad de dinero de una mochila en la casa 3?	No
21	3C2	C	¿Alguna vez le robó algo a alguien que confiaba en usted?	No
22	4R2	R	El día de hoy, ¿tomó cualquier cantidad del dinero reportado robado de una mochila roja?	No
23	3C3	C	¿Alguna vez hizo algo que lo haría sentirse avergonzado en frente de su familia?	No
24	XX		La prueba está por terminar, por favor no se mueva hasta que libere el aire de la manga.	

Análisis de los datos de prueba

Debido a que el formato de preguntas experimental fue previsto para exámenes de evento-específico, las categorías de decisión deberían obtenerse a nivel de la prueba como un todo. Todos los exámenes fueron evaluados utilizando el Sistema de Puntuación Empírica, Empirical Scoring System en inglés (ESS, Nelson, et al., 2011). Información detallada

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

del ESS, sus distribuciones de referencia estadística y sus reglas de decisión pueden encontrarse en Nelson et al. Las RQs fueron comparadas en contra de la CQ previa. Si esta CQ estaba distorsionada por un artefacto, la RQ se comparaba en contra de la CQ libre de artefactos más cercana.

Se determinaba entonces la puntuación subtotal para cada pregunta relevante de las primeras tres presentaciones, y se sumaron las puntuaciones subtotales para tener una puntuación de gran total. Las clasificaciones iniciales se realizaron utilizando la Regla de Gran Total, cuyo puntaje de corte de gran total fue de ± 4 . Esto corresponde a una puntuación de corte con probabilidad a .05, tanto para clasificaciones veraces como de engaño. En otras palabras, un gran total de 4 o mayor, daba como resultado una clasificación de veracidad, mientras que un gran total de -4 o menos daba como resultado una clasificación de engaño. Las puntuaciones entre -3 y 3 daban como resultado una clasificación de inconcluso. Para un resultado inconcluso, el examinador podría calificar la cuarta presentación y sumar nuevamente los puntajes de subtotal y gran total. Se mantienen iguales los puntajes de corte de ± 4 cuando se califican las cuatro presentaciones del estímulo.

Si el resultado se mantiene inconcluso, o si la diferencia entre 2 de los subtotales es superior a 7 puntos, entonces el examinador utiliza una Regla de Puntuación Subtotal, en inglés Subtotal-Score-Rule (SSR). El SSR requiere que cualquiera de los subtotales obtenga un total de -6 o menos para que resulte en una clasificación de engaño. Un puntaje subtotal de -6 corresponde a una puntuación de corte con probabilidad estadísticamente corregida de .05 para el engaño.

Análisis

De los 114 sujetos incluidos en el estudio, 22 casos (19,3%) no se incluyeron en el análisis estadístico debido a alguna forma de violación de protocolo. Las violaciones de protocolo excluidas redujeron el tamaño de la muestra a 92 casos, incluyendo 29 casos programados como culpables y 43 casos programados como inocentes.

El análisis se realizó utilizando una combinación de software matemático que incluyó:

- R: Un lenguaje y entorno para el cálculo estadístico (R Development Core Team, 2008).
- STATA v. 11.0 (Statacorp, 2009)

Resultados

Se calculó la precisión diagnóstica excluyendo todos los casos inconclusos. Esto produjo un tamaño de muestra de 75, 40 de los cuales fueron programados como "Inocentes" y 35 fueron programados como "Culpables." Las tasas de sensibilidad y especificidad se calcularon incluyendo los casos inconclusos. Los resultados se presentan en la Tabla 1. La Precisión de una prueba es la proporción de individuos clasificados correctamente y la Tasa de Error es lo contrario, la proporción de sujetos clasificados incorrectamente. La Sensibilidad se refiere a la proporción de individuos "Culpables" correctamente clasificados por engaño. La Especificidad se refiere a lo opuesto, la proporción de individuos "Inocentes" correctamente clasificados como no-engaño. Los Falsos Positivos y Falsos Negativos se refieren a la proporción de individuos incorrectamente clasificados por Engaño o No-Engaño, respectivamente.

La razón de verosimilitud (RV) utiliza tanto la sensibilidad como la especificidad de la prueba para proporcionar una estimación de qué tanto un resultado de prueba cambiará la probabilidad o posibilidad de tener una condición. La razón de verosimilitud para un resultado positivo (RV+) indica el factor por el que la probabilidad o posibilidad de la condición se incrementa cuando una prueba es positiva (SR). La razón de verosimilitud para un resultado negativo (RV-) indica el factor por el cual la probabilidad o posibilidad de la condición se decrementa cuando una prueba es negativa (NSR). La razón de verosimilitud de un resultado de engaño (RV+) es la proporción de la probabilidad de un resultado de Engaño en un "Culpable" individual en comparación con la probabilidad asociada de un sujeto "Inocente". En otras palabras, este número se describe como la proporción: Sensibilidad / (1- Especificidad). La razón de verosimilitud de una prueba negativa (RV-) funciona de la misma manera; sin embargo, en este caso describe la proporción de la probabilidad de un resultado de No-Engaño en un sujeto "Culpable" en comparación con la probabilidad asociada de un sujeto "Inocente". Este número se describe como la proporción: (1- Sensibilidad) / Especificidad.

Tabla 1. Precisión del formato experimental.

PRECISIÓN DIAGNÓSTICA	
Precisión (Intervalo de Confianza de Wilson)	86.667 % (77.2 %, 92.6 %)
Sensibilidad (Intervalo de Confianza de Wilson)	80.0 % (64.1 %, 90.0 %)
Especificidad (Intervalo de Confianza de Wilson)	92.5 % (80.2 %, 97.4 %)
Tasa de Error (Intervalo de Confianza de Wilson)	13.3 % (7.4 %, 22.8 %)
Falsos Positivos (Intervalo de Confianza de Wilson)	4 % (1.4 %, 11.1 %)
Falsos Negativos (Intervalo de Confianza de Wilson)	9.3 % (4.6 %, 18.0 %)
Razón de Verosimilitud (+)	10.67
CRazón de Verosimilitud (-)	0.22

Los resultados inconclusos se muestran en la Tabla 2, junto con los intervalos de confianza al 95%. Los intervalos de confianza se obtuvieron a través de un método computacionalmente sencillo propuesto por Wilson (1927), utilizando un refinamiento del método asintótico simple. Diecisiete de los 92 casos produjeron resultados inconclusos. Más del 18% de los participantes de la muestra tuvieron un resultado de examinación inconcluso con el formato experimental.

Tabla 2. Resultados inconclusos e intervalos de confianza

TASA DE INCONCLUSOS	
Número de Resultados Inconclusos	17
Número de Resultados Inconclusos (Dentro de los sujetos "Inocentes")	9
Número de Resultados Inconclusos (Dentro de los sujetos "Culpables")	8
Tasa Total de Inconclusos	18.5 %
(Intervalo de Confianza de Wilson)	(11.9 %, 27.6 %)
Tasa de Inconclusos (Dentro de los sujetos "Inocentes")	18.4 %
(Intervalo de Confianza de Wilson)	(9.9 %, 31.357%)
Tasa de Inconclusos (Dentro de los sujetos "Culpables")	18.6 %
(Intervalo de Confianza de Wilson)	(9.7 %, 32.6 %)

Las estimaciones de confiabilidad de los resultados de prueba con el formato experimental se muestran en la Tabla 3, basadas en el Estadístico Kappa de Cohen (Cohen, 1960). El uso de esta estadística requiere de un supuesto subyacente de clasificaciones homogéneas proporcionadas por los diferentes evaluadores. Esta hipótesis fue aceptada debido a que todos los calificadores y revisores utilizaron un procedimiento estandarizado. Este estudio consistió en múltiples evaluadores, incluyendo tanto a los examinadores originales como a los revisores de control de calidad. Esto se redujo a dos categorías de calificadores para este análisis, examinador y revisor.

Tabla 3. Confiabilidad diagnóstica del formato

CONFIABILIDAD DIAGNÓSTICA	
Estadística Kappa de Cohen (Método analítico de Intervalo de Confianza)	0.730 (0.576 , 0.885)
Área Bajo la Curva ROC (Método analítico de Intervalo de Confianza)	0.862 (0.783 , 0.941)
Acuerdo	86.67%
Correlación	0.734

La tabla 4 muestra las frecuencias de los resultados correctos e incorrectos con los casos de culpables e inocentes. La tabla 5 muestra las medias y las desviaciones estándar de las puntuaciones de los grupos de inocentes y culpables, y en la Figura 1 se muestra un diagrama de datos de las puntuaciones numéricas.

Tabla 4. Tabulación cruzada del rendimiento de clasificación, excluyendo casos inconclusos.

		Resultados de Prueba		
		DI	NDI	TOTAL
Estatus asignado	Culpable	28	7	35
	Inocente	3	37	40
TOTAL		31	44	TOTAL= 75 CASOS

Tabla 5. Estadística Descriptiva de las Puntuaciones Calculadas

Estadística Descriptiva de las Puntuaciones	
Media aritmética de los puntajes (Dentro de los inocentes)	5.449
Desviación estándar de los puntajes (Dentro de los inocentes)	5.545
Media aritmética de los puntajes (Dentro de los culpables)	-3.256
Desviación estándar de los puntajes (Dentro de los culpables)	5.350

Discusión

Este proyecto involucró el estudio de la precisión de decisión de una técnica poligráfica experimental en una sola secuencia, con una cohorte de participantes de la comunidad que fueron asignados al azar a los estatus de culpable e inocente, relacionados con un delito de robo simulado. La precisión de la técnica experimental fue del 87%. La revisión de los intervalos de confianza reportados en el presente estudio y por la APA (2011) indican que la precisión observada es consistente con el rango normal de exactitud reportado previamente para una técnica diagnóstica. Aunque que la tasa observada de inconclusos del 18% fue mayor que la reportada por otros formatos de polígrafos diagnósticos, se mantuvo dentro del límite máximo del 20% que fue previamente descrito por APA. La precisión de la decisión observada para el formato experimental no cumple el requisito APA de una precisión superior a .90 para los exámenes evidenciarios. Serán necesarios resultados de otros estudios de este formato experimental, combinados con los presentes resultados, para lograr cualquier conclusión acerca la precisión media de esta técnica.

Se pueden describir un número de limitaciones en lo que respecta a este estudio, comenzando con el uso de la actuación de un crimen simulado, y también el hecho de que algunos de los examinadores había completado muy recientemente su formación académica en poligrafía y prácticamente no tenían experiencia de campo real. La validez ecológica – aunque las condiciones de evaluación fueron muy aproximadas a las condiciones de prueba de campo reales– de las investigaciones de

laboratorio y de la actuación de crímenes falsos es una reflexión importante, aunque en algunas ocasiones puede ser malinterpretada o confundida fácilmente con la validez externa. La validez externa es un problema más importante, porque se refiere directamente a si los resultados observados en el estudio, probablemente podrían ser observados en la práctica de campo. La preocupación práctica es el grado en el que los resultados de laboratorio podrían sobrestimar la efectividad de ambientes de campo. Como una cuestión de conocimiento generalizado, en la mayoría de los ámbitos de la investigación científica, Hay una tendencia clara de que una investigación de laboratorio realizada adecuadamente, es útil e informativa cuando se intenta entender y estimar el rango de efectividad que se podría esperar en ambientes de campo.

En general, la investigación en poligrafía en ambientes de laboratorio ha tenido como resultado estimaciones más bajas o más conservadoras, con respecto a la precisión del polígrafo comparado con investigaciones en ambientes de campo. Ya sea por la pericia del examinador o por los aspectos motivacionales de los sujetos evaluados en los exámenes de campo reales, es posible que esto diera como resultado estos incrementos previamente observados en la efectividad de los resultados de estudios de campo sobre aquellos en ambientes de laboratorio. También es posible que las investigaciones llevadas a cabo en ambientes de campo fueron conformadas por metodologías de selección de muestras no aleatorias, lo que incrementa la precisión en los resultados de los estudios de campo por encima de lo que se observa mediante el muestreo aleatorio. Aunque a menudo los estudios de campo carecen de completo control experimental y de selección al azar, ofrecen la ventaja de tener validez ecológica percibida. La investigación de laboratorio ofrece la ventaja de un mayor control experimental y del potencial de la asignación aleatoria del estatus de culpable o inocente. Independientemente de las ventajas y desventajas, las diferencias en los resultados entre estudios de laboratorio y de campo bien diseñados, no han demostrado tener hasta ahora diferencias estadísticamente significativas. (Anderson, Lindsay, y Bushman, 1999), y no hay indicadores obvios de que los resultados observados en el presente estudio son una sobreestimación de la precisión de la prueba.

Merece discusión el gran número de violaciones protocolo que generaron datos de examinación inutilizables. Casi el 20% de los exámenes realizados no pudieron ser utilizados debido a datos con artefactos muy fuertes que no podían ser interpretados y debido a violaciones de protocolo

por parte de los examinadores. La mayoría de estos exámenes se debieron a lo último, y atribuimos esto a la falta general de experiencia por parte de muchos de los examinadores y también a la falta de familiaridad de parte de los examinadores hacia un protocolo de prueba experimental, para el cual los examinadores no habían recibido instrucción o práctica previa hasta el inicio de este proyecto.

Existen algunas ventajas hipotéticas de un formato diagnóstico en secuencia individual, comenzando con el potencial de una prueba con mayor efectividad; resultado de la reducción de una fuente de varianza de respuestas no controlada por iniciar y detener la grabación al utilizar los formatos CQT diagnósticos tradicionales. Debido a que el formato experimental no supera el desempeño de los formatos de diagnóstico poligráfico existentes de ninguna forma, estas ventajas hipotéticas no se observaron en los resultados de este estudio.

A pesar de estas limitaciones obvias, en este formato diagnóstico en una sola secuencia de dos RQ, se observaron tasas de precisión y de inconclusos remarcablemente similares a las reportadas previamente por la APA (2011) para otros formatos poligráficos diagnósticos con dos RQs. La exactitud observada durante este estudio fue esencialmente idéntica a la de otros formatos: no mejor, pero tampoco peor. Es posible que futuras modificaciones y perfeccionamiento del formato de una sola secuencia podrían conducir a incrementos graduales en la efectividad de la prueba, sobre de los formatos existentes.

Otro aspecto interesante e importante de este estudio es que todos los participantes y los examinadores - y el primer y segundo autor - son personas de habla hispana nativa. Todos los exámenes fueron realizados en español, en la Ciudad de México. La importancia de esto, aunque no fue el objetivo de este proyecto, son las tasas trans-culturales de precisión de decisión y de inconclusos que fueron observadas y que son consistentes con resultados de investigaciones previamente reportados en estudios en inglés para otros formatos de diagnóstico con dos RQs. Aunque este estudio no se enfocó en la eficacia de las preguntas del PLC, hay que destacar que este estudio añade soporte para el supuesto de que la precisión y la eficiencia de las técnicas poligráficas CQT que utilizan preguntas DLC se pueden mantener estables a través de las diferencias lingüísticas y culturales. Previamente, las preguntas DLC también se reportaron como eficaces en estudios sobre polígrafo exploratorio,

realizados bajo circunstancias de lenguaje nativo en Irak (Nelson, Hander y Morgan, 2012). Junto con otras ventajas de las preguntas DLC, que incluyen la transparencia, la no dependencia de la ingenuidad de los sujetos, y la administración no manipulada – las preguntas DLC pueden seguir desempeñando un papel cada vez más importante tanto en poligrafía diagnóstica como exploratoria.

Finalmente, recomendamos cautela en el uso inmediato de este formato experimental en ambientes de campo. Es difícil encontrar un argumento ético sensato para la selección de un protocolo experimental que no supera las prácticas existentes basadas en evidencia. Hasta que futuras investigaciones y la experiencia puedan confirmar que el formato experimental proporcionará un nivel de eficiencia que sea igual o superior a otros métodos establecidos, los profesionales deben permanecer informados de que el uso de procedimientos experimentales con el público se puede realizar de forma ética, solamente cuando se informa debidamente al examinando y se le proporciona la oportunidad de ejercer su consentimiento informado con respecto a si se somete a una prueba con un método experimental. Se recomienda mantener interés en los formatos poligráficos de en una sola secuencia, aunque es necesario una mayor investigación y desarrollo antes de que este formato experimental pueda ser utilizado en ambientes de campo.

Referencias

- American Polygraph Association (2009a). Model Policy for Post-conviction Sex Offender Testing. [Electronic version] Retrieved January 25, 2012, from <http://www.polygraph.org>
- American Polygraph Association. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40, 194-305.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychology laboratory: Truth or Triviality? *Current Directions in Psychological Science*, 8, 3-9.
- Cohen, Jacob (1960). "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement* 20 (1): 37-46.
doi:10.1177/001316446002000104
- Department of Defense (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retrieved from <http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf> on 3-31-2007. Reprinted in *Polygraph*, 40(1), 2-66.
- Handler, M., Nelson, R. & Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph*, 37(2), 100-111.
- 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.
- Kircher, J. C., Packard, T., Bell, B. G., & Bernhardt, P. C. (2010). Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests. *Polygraph* 39, 22-67.
- 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. Final report.
- 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.
- National Research Council (2003). *The Polygraph and Lie Detection*. National Academy of Sciences Press.

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. Handler, M. Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System, *Polygraph*, 40 (2).
- Nelson, R., 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).
- R Development Core Team, (2008). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic (1-49).
- Research Division Staff (1995). A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats. Report number DoDPI94-R-0008. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 26(2), 79-106.
- Research Division Staff (1995). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. DoDPI94-R-0009. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 27, (3), 171-180.
- StataCorp. (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP
- Wilson, E. B. "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, 22, 209-212 (1927).

*Para cualquier corrección o consulta, el traductor puede ser
contactado en **rodolfo@poligrafia.com.mx***