



John C. Kircher*

David C. Raskin

Universidad de Utah,
Salt Lake City, Utah, EE.UU.

Investigación en laboratorio y en el campo sobre la prueba del engaño por observación del movimiento ocular

Лабораторные и практические исследования применения «ocular-motor» для выявления лжи

Palabras clave: “motor-ocular”, “prueba de detección motor-ocular”, “detección instrumental del engaño”

Divulgación: los autores tienen intereses financieros en Converus, Inc. (www.converus.com), una compañía que ha comercializado la tecnología descrita en este informe. Hemos divulgado esos intereses a la Universidad de Utah y hemos establecido un plan aprobado para manejar cualquier posible conflicto que surja debido a nuestra participación en Converus.

El presente artículo examina la lógica y suposiciones teóricas que respaldan la prueba del engaño por observación del movimiento ocular (ODT, por sus siglas en inglés) así como la evidencia empírica de su validez con relación a criterios. Las investigaciones sugieren que la ODT puede contribuir a programas de evaluación preempleo y periódica, particularmente en agencias gubernamentales encargadas de hacer cumplir la ley y de la seguridad nacional.

* John.Kircher@utah.edu

** davidc.raskin@me.com

Lógica subyacente a la prueba del engaño por observación del movimiento ocular (ODT)

Cook et al. (2012) introdujeron un nuevo método para la detección del engaño que se llama prueba del engaño por observación del movimiento ocular (ODT, por sus siglas en inglés). En contraste con el polígrafo, la ODT se automatiza y puede completarse en aproximadamente 40 minutos. Una computadora presenta instrucciones sintetizadas y escritas seguidas por declaraciones escritas de prueba tipo verdadero / falso con respecto a la posible participación del examinado en actividades ilícitas. Las instrucciones informan al examinado que, si no responde de manera rápida y precisa, no aprobará la prueba. Seguidamente el examinado lee las declaraciones presentadas en serie por la computadora mientras un rastreador ocular remoto registra los movimientos oculares y los cambios en el tamaño de la pupila. El examinado presiona una tecla en el teclado para responder verdadero o falso. La computadora procesa los datos del movimiento ocular, combina las mediciones en una ecuación de regresión logística y clasifica al individuo como veraz o engañoso en la prueba.

La ODT utiliza un formato de prueba conocido como prueba de comparación relevante (RCT, por sus siglas en inglés). Originalmente, desarrollamos la RCT como una nueva técnica de polígrafo para usar en los puertos de entrada para detectar en los viajeros el posible tráfico de drogas y / o el transporte de explosivos (Kircher et al., 2012). La RCT incluye preguntas sobre dos temas relevantes (R1 y R2) que se mezclan con preguntas neutrales, y utiliza la diferencia entre las reacciones a los dos conjuntos de preguntas relevantes para determinar si el examinado fue sincero o engañoso en cualquiera de los temas relevantes. Cada tema relevante sirve como un control para el otro. Los examinados que reaccionan con más fuerza a las preguntas relacionadas con uno de los temas son hallados engañosos en sus respuestas a las preguntas sobre ese tema relevante. Los examinados que muestran poca o ninguna diferencia en las reacciones a los dos conjuntos de preguntas relevantes se consideran veraces sobre ambos temas.

La ODT se basa en el supuesto de que mentir es cognitivamente más exigente que decir la verdad. Un tema recurrente en la literatura sobre técnicas de detección del engaño (Johnson, Barnhardt y Zhu, 2005; Kircher, 1981; Raskin, 1979; Steller, 1989; Vrij, Fisher, Mann y Leal, 2006). En contraste con las personas veraces, un individuo engañoso debe identificar las preguntas que contesta de manera veraz y las que contesta de manera engañosa. Cuando reconoce una pregunta como inculpatoria, debe inhibir la respuesta veraz prepotente y hacerlo de manera consistente, rápida y precisa. Mientras realizan esta tarea, los individuos engañosos también pueden autocontrolar su desempeño para detectar señales que revelen su engaño, ya sea respondiendo muy lentamente o cometiendo demasiados errores. El reclutamiento de recursos mentales para realizar estas actividades cognitivas y metacognitivas adicionales podría explicar el impacto observado en la dilatación de la pupila, los movimientos oculares, el tiempo de respuesta y las tasas de error (Hacker et al., 2012; Kahneman, 1973; Loewenfeld, 1999; Rayner, 1998).

La pupila reacciona no solo a la carga cognitiva sino también a los estímulos

emocionales. Varios investigadores han informado que los estímulos emocionales provocan respuestas de la pupila cuya magnitud depende de la intensidad, no de la valencia del estímulo emocional (Bradley, Micolli, Escrig y Lang, 2008; Hess y Polt, 1960; Hess & Polt, 1964; Steinhauer, Boller, Zubin & Pearlman, 1983).

Las pruebas de polígrafo se basan en el concepto de que los individuos engañosos mostrarán respuestas emocionales más fuertes a las preguntas de la prueba que fueron contestadas de manera engañosa que a las que respondieron con la verdad. En la medida en que las reacciones emocionales a las preguntas de prueba distinguen entre individuos engañosos y verdaderos, las respuestas de las pupilas deben reflejar esas diferencias y ser diagnósticas de engaño. De acuerdo con esta predicción, varios investigadores han respondido que, durante exámenes poligráficos con información oculta y mentiras probables, la pupila se dilata más cuando las personas son engañosas que cuando son veraces (por ejemplo, Bradley y Janisse, 1979; Dionisio et al., 2001; Janisse y Bradley, 1980; Webb et al., 2009).

Un lector que tiene dificultades para leer o comprender texto muestra más fijaciones de los ojos, aumento de la pupila y tiempos de lectura más largos (Rayner, 1998; Rayner, Chace, Slattery y Ashby, 2006). Si los individuos engañosos experimentan una mayor carga cognitiva y dificultad para procesar los ítems de prueba que los individuos veraces, deberíamos ver diferencias entre los grupos en estas medidas.

Investigación de laboratorio de delitos simulados con la ODT

Hemos realizado una serie de estudios de laboratorio y de campo para determinar si las medidas del movimiento ocular discriminan entre individuos veraces y engañosos. Los estudios de laboratorio utilizan un procedimiento de delitos simulados que modelamos según investigaciones de laboratorio sobre técnicas de polígrafo. Los experimentos realistas de simulacros de delitos producen efectos de diagnóstico en reacciones electrodérmicas, cardiovasculares y respiratorias que son similares a las que se obtienen de sospechosos reales (Kircher, Horowitz y Raskin, 1988; Raskin y Kircher, 2014). En nuestros experimentos, reclutamos participantes del campus universitario o de la comunidad en general y los asignamos al azar a condiciones de tratamiento de culpables e inocentes. Les damos instrucciones a los participantes culpables para que cometan un delito simulado, como robar un examen de la oficina de un profesor, o sacar \$20 del bolso de una secretaria, y luego mentir al respecto en la ODT. Para simplificar el diseño de la investigación, hemos comenzado a realizar experimentos con uno en lugar de dos delitos simulados. A los participantes en estos experimentos se les hace creer que algunos participantes culpables roban un examen de la oficina de un profesor, mientras que otros sacan \$20 del bolso de una secretaria. En realidad, todos los participantes culpables sacan \$20 del bolso. Dado que los examinados en estudios de campo usualmente se sienten altamente motivados para pasar las pruebas, ofrecemos a nuestros participantes un bono monetario sustancial por parecer inocentes de los delitos. La ODT consta de 16 declaraciones tipo Verdadero / Falso relacionadas con el robo de

los \$20 por ejemplo, "Tomé los \$20 del bolso de la secretaria"), 16 declaraciones relacionadas con el robo del examen (por ejemplo, "No tomé el examen de la

oficina del profesor") y 16 declaraciones neutrales (por ejemplo, "Estoy sentado frente a una computadora, tomando una prueba"). En el formato estándar, organizamos las 48 declaraciones para que no aparezcan dos declaraciones del mismo tipo en sucesión. La computadora presenta el conjunto de 48 ítems 5 veces con una breve pausa entre las repeticiones durante las cuales el examinado responde a problemas aritméticos simples. Excluyendo los problemas aritméticos, el examinado responde a un total de 80 declaraciones V / F de cada tipo. La respuesta correcta (no incriminatoria) a la mitad de los ítems de cada tipo es Verdadero y la respuesta correcta a la otra mitad es Falso.

Medidas del movimiento ocular y del comportamiento. Los rastreadores oculares en nuestros experimentos midieron la posición de la mirada horizontal, la posición de la mirada vertical y uno o dos canales de diámetro de la pupila, dependiendo de si el rastreador registró datos de los ojos izquierdo y derecho o solo del ojo derecho. Los rastreadores almacenaron muestras de cada canal a 30Hz o 60Hz. Cuando el rastreador midió los ojos izquierdo y derecho, calculamos la media de las medidas de ambos ojos, porque los dos canales estaban altamente correlacionados.

La media del cambio en el tamaño de la pupila para un experimento de simulacro de delito reportado por Cook et al. (2012) se muestra en la Ilustración 1. Muestra la respuesta de la pupila evocada durante cuatro segundos después del inicio de la declaración para declaraciones neutrales, declaraciones sobre el robo del dinero en efectivo y declaraciones sobre el examen. Como se predijo, los participantes culpables reaccionaron con más fuerza a las declaraciones sobre el efectivo que a las del examen, mientras que los participantes inocentes mostraron poca diferencia entre las declaraciones sobre el efectivo y las del examen. Es de destacar que la media del cambio en el diámetro de la pupila (DP) fue inferior a 0,1 mm.

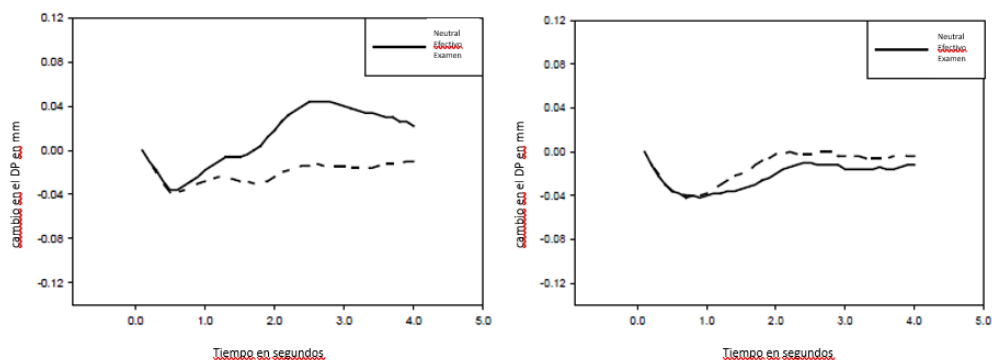


Ilustración 1. Cambio en la media del diámetro de la pupila (PD) desde el inicio de la declaración para los participantes culpables (izquierda) e inocentes (derecha)

Para ajustar las diferencias individuales en la reactividad y las diferencias en las condiciones de luz ambiental, transformamos los datos de la pupila para cada una de las cinco repeticiones de ítems de prueba en calificaciones estándar. La estandarización también estableció una métrica común a través de repeticiones dentro de los individuos. De las curvas de reacciones estandarizadas de la pupila,

extrajimos dos características: el área bajo la curva de reacción y el nivel de reacción en la respuesta del participante.

Derivamos medidas de lectura del análisis de las fijaciones oculares en el texto. Una computadora identificó las fijaciones oculares en el texto donde había poca variación tanto en la posición de la mirada horizontal como en la vertical, con una duración mínima de 100ms y una duración máxima de 1000ms (Cook et al., 2012). Se definió un área de interés para cada elemento de prueba antes del cálculo de las medidas de lectura. El área de interés comenzó con el primer carácter del ítem y finalizó después del período al final del mismo.

Además de medir la validez diagnóstica de varias medidas del movimiento ocular, también usamos el coeficiente alfa para evaluar la confiabilidad de esas medidas (Chronbach, 1951). La confiabilidad indica la medida en que las mediciones obtenidas de las cinco repeticiones de los ítems de prueba son consistentes. Por ejemplo, si los datos de la primera repetición indicaban que el individuo era engañoso, ¿parecía la persona también engañoso en la segunda, tercera, cuarta y quinta repeticiones?

Exactitud de la ODT en estudios de laboratorio

La Tabla 1 presenta los coeficientes de validez y consistencia interna (confiabilidad) para el conjunto de características en experimentos realizados por dos de nuestros estudiantes (Patnaik, 2015; Webb, 2008). También muestra estadísticas de confiabilidad de consistencia interna para las funciones generadas por computadora.

Tabla 1. Coeficientes de validez y confiabilidad para dos estudios de delitos simulados con la ODT

	Media		Webb (2008)		Patnaik (2015)		Media Mean	
Área bajo la curva	R1 - R2	0.409	0.640	0.586	0.615	0.505	0.628	
Área bajo la curva	R1 - (NT+R2)/2	0.396	0.759	0.554	0.639	0.482	0.699	
Nivel en la respuesta	R1 - R2	0.557	0.465	0.585	0.510	0.571	0.488	
Nivel en la respuesta	R1 - (NT+R2)/2	0.548	0.527	0.634	0.575	0.593	0.551	
Lectura								
Número de fijaciones	R1 - R2	-0.509	0.572	-0.406	0.627	-0.460	0.600	
Número de fijaciones	R1 - (NT+R2)/2	-0.329	0.807	-0.293	0.720	-0.312	0.764	
Duración de la primera pasada	R1 - R2	-0.549	0.582	-0.253	0.540	-0.427	0.561	
Duración de la primera pasada	R1 - (NT+R2)/2	-0.293	0.622	-0.166	0.585	-0.238	0.604	
Duración de relectura	R1 - R2	-0.488	0.516	-0.342	0.397	-0.421	0.457	
Duración de relectura	R1 - (NT+R2)/2	-0.224	0.683	-0.115	0.407	-0.178	0.545	
Comportamiento								
Tiempo de respuesta	R1 - R2	-0.529	0.434	-0.497	0.329	-0.513	0.382	
Tiempo de respuesta	R1 - (NT+R2)/2	-0.312	0.788	-0.348	0.671	-0.330	0.730	
Índice de error	R1 - R2	0.082	0.052	0.093	0.209	0.088	0.131	
Índice de error	R1 - (NT+R2)/2	0.242	0.741	-0.002	0.690	0.171	0.716	
Frecuencia de parpadeo								
Frecuencia de parpadeo por ítem	R1 - R2	-0.014	0.247	-0.388	0.182	-0.275	0.215	
Frecuencia de parpadeo por ítem	R1 - (NT+R2)/2	-0.015	0.572	-0.191	0.101	-0.135	0.337	
Frecuencia de parpadeo del siguiente ítem	R1 - R2	0.169	0.104	-0.088	0.351	0.135	0.228	
Frecuencia de parpadeo del siguiente ítem	R1 - (NT+R2)/2	0.010	0.315	-0.105	0.381	0.075	0.348	

Los coeficientes de validez en negrita son estadísticamente significativos a $p < 0.05$.

Las fuertes correlaciones positivas para las características de las pupilas indicaron que los participantes culpables mostraron mayores aumentos en el tamaño de las pupilas en respuesta a las declaraciones de R1 (efectivo) que a las de R2 (examen). Estos hallazgos son consistentes con los datos presentados en la Figura 1. Las correlaciones negativas para las mediciones de tiempo de lectura y respuesta indicaron que los participantes culpables hicieron menos fijaciones y pasaron menos tiempo leyendo las declaraciones R1 que las declaraciones R2. Los efectos en el tiempo de respuesta son sustanciales y consistentes en múltiples experimentos y grupos culturales. Basados en la psicología de la lectura de literatura, inicialmente pensamos que los participantes culpables experimentarían más dificultades y pasarían más tiempo en las declaraciones R1 que en las R2. Sin embargo, los datos sugieren que los participantes culpables invierten más esfuerzo mental para responder esas declaraciones de manera rápida y precisa en un intento por evitar la detección (Cook et al., 2012). Los participantes culpables logran su objetivo, pero revelan su engaño. Esta hipótesis explica los efectos sobre el tiempo de respuesta y las medidas de lectura, así como los aumentos observados en el tamaño de la pupila asociados con las declaraciones R1.

El examen de los coeficientes de validez promedio indica que las medidas de la pupila fueron más diagnósticas que las de lectura, de comportamiento y de parpadeo. El contraste (R1 - R2) para el tiempo de respuesta fue casi tan diagnóstico como el

tamaño de la pupila. Las medidas de la frecuencia de parpadeo y el índice de error fueron las menos predictivas del estado de culpabilidad.

Modelo de decisión. Para clasificar a los individuos como veraces o engañosos, combinamos las puntuaciones en un subconjunto de medidas de diagnóstico de una manera matemáticamente óptima para calcular la probabilidad de engaño. Si la probabilidad de engaño es superior a 0.5, clasificamos a la persona como engañosa; si la probabilidad es inferior a 0.5, clasificamos a la persona como veraz. Se han desarrollado varios procedimientos estadísticos para identificar un subconjunto de medidas de diagnóstico que representarán la mayor parte de la varianza diagnóstica en el conjunto completo de medidas y funcionarán bien cuando se evalúen en una muestra independiente de casos. Estos procedimientos tienden a seleccionar medidas que están más altamente correlacionadas con el estado de culpabilidad y menos altamente correlacionadas entre sí. Luego utilizamos el análisis de regresión logística para derivar un peso único para cada medida del movimiento ocular que maximiza la separación entre grupos veraces y engañosos.

Gran parte de nuestra investigación ha sido diseñada para evaluar los efectos de los factores que podrían afectar la precisión de la ODT y para explorar métodos alternativos para presentar los elementos de prueba. Con respecto a este último objetivo, no hemos mejorado el formato de presentación y los procedimientos de delitos simulados evaluados en nuestro primer experimento de simulacros de delitos (Osher, 2005), que llamamos el protocolo estándar. Debido a que nuestros intentos de mejorar el protocolo estándar han dado resultados inferiores, la Tabla 2 presenta los resultados obtenidos con el protocolo estándar, y la Tabla 3 presenta los resultados de los protocolos no estándar.

Tabla 2. Porcentaje de decisiones correctas en condiciones estándar en experimentos de delitos simulados

Experimento	Factores	N	n _G	n _I	Culpable	Inocente	Media	Validación _G	Validación _I	Media
Osher (2005) ^a	temas; formato de serie	40	20	20	85.0	85.0	85.0	85.0	70.0	77.5
Webb (2008) ^b	sexo; motivación; dificultad	112	56	56	82.1	89.2	85.7	89.3	80.4	84.9
Patnaik (2013) ^a	interrogación directa	48	24	24	83.3	95.8	89.6	83.3	83.3	83.3
Patnaik (2015) ^a	distribuidos; retroalimentación previa a la prueba; intervalo postrespuesta	80	40	40	82.5	90.0	86.3	80.0	90.0	85.0
Patnaik et al. (2016) ^c	lenguaje; cultura	145	82	63	84.1	87.3	85.7	81.9	87.5	84.7
Medio Oriente (2016a)	lenguaje; cultura	112	51	61	80.4	88.5	84.5			
Medio Oriente (2016b) ^d	lenguaje; cultura	101	52	49				75.0	85.7	80.4
Protocolo estándar		638	325	313	82.8	89.0	85.9	82.1	84.1	83.1

^a Los resultados de la validación se obtuvieron con el procedimiento de dejar uno fuera

^b El modelo de decisión basado en Patnaik et al. (2016) se utilizó para clasificar a los participantes en Webb (2008)

^c El modelo de decisión basado en Webb (2008) se utilizó para clasificar a los participantes en Patnaik et al. (2016)

^d El modelo de decisión basado en Medio Oriente (2016a) se utilizó para clasificar a los participantes en Medio Oriente (2016b)

Los resultados presentados en la Tabla 2 indican que el protocolo estándar en los experimentos de delitos simulados produjo aproximadamente el 86% de clasificaciones correctas en la muestra original de estandarización y aproximadamente el 83% de clasificaciones correctas cuando se probó en muestras independientes (validación cruzada). En la validación cruzada, la precisión fue ligeramente mayor para los participantes inocentes (84.1%) que para los culpables (82.1%).

La Tabla 3 resume los resultados de condiciones no estándar. Los resultados de Osher (2005) sugieren que obtenemos más información de diagnóstico de presentaciones en serie de declaraciones de prueba individuales que con la visualización simultánea de varias declaraciones de prueba. Webb (2008) encontró que el sexo de la persona no modera los efectos del engaño en las medidas del movimiento ocular, mientras que una mayor motivación para aprobar la prueba y la simplicidad semántica en la redacción de las declaraciones de prueba mejoran la validez diagnóstica de algunas medidas de movimiento ocular.

Tabla 3. Porcentaje de decisiones correctas en condiciones no estándar en experimentos de delitos simulados

Experimento	Factores	N	n _c	n _i	Culpable	Inocente	Media
Osher (2005)	temas; formato paralelo	40	20	20	70.0	95.0	82.5
USTAR (2010)	cuestionario de prueba previa; temas	71	47	27	59.6	77.8	68.7
NSA (2012)	estandarización	94	51	43	72.5	88.4	80.5
NSA (2013)	validación	60	34	26	50.0	80.8	65.4
Patnaik (2013)	interrogación indirecta	48	24	24	58.3	79.2	68.8
Patnaik (2015)	bloqueado	80	40	40	77.5	85.0	81.3
Protocolos no estándar		393	216	180	65.3	84.5	74.9

En conjunto, los estudios USTAR (2010) y Patnaik (2013) indicaron que las declaraciones de prueba que se refieren directamente al asunto en cuestión ("No tomé los \$20") producen precisiones más altas que las declaraciones que preguntan indirectamente si la persona falsificó información para encubrir su culpa ("No falsifiqué mis respuestas a las preguntas sobre el robo de los \$20"). En los estudios de la NSA, reclutamos empleados y los probamos sobre violaciones menores de seguridad. Los estudios utilizaron un protocolo no estándar porque no se nos permitía proporcionar incentivos a los empleados del gobierno para aprobar la ODT, muchos participantes eran examinadores de polígrafo federales que sabían que esencialmente no había consecuencias por reprobado la ODT y teníamos que confiar en el autoreporte como un representante de la verdad fundamental.

Patnaik (2015) encontró que la secuenciación pseudoaleatoria estándar de las declaraciones NT, R1 y R2 mejora la validez diagnóstica de las medidas del movimiento ocular, mientras que la retroalimentación sobre el rendimiento en una prueba de práctica pre-ODT y el alargamiento del intervalo entre la respuesta y la presentación del siguiente ítem no lo hace. Patnaik et al. (2016) encontraron que los efectos en las medidas del movimiento ocular eran similares para las pruebas administradas en inglés o español a hablantes nativos inscritos como estudiantes universitarios. Los experimentos en el Medio Oriente requirieron la modificación del software de visualización para presentar texto en árabe de derecha a izquierda. Las tasas de precisión en la validación cruzada en el Medio Oriente fueron ligeramente más bajas que las obtenidas en los Estados Unidos y México, particularmente para los participantes culpables. Si bien las diferencias en la precisión de las decisiones entre los participantes de Oriente Medio y Occidente no fueron estadísticamente significativas, encontramos que es necesario reducir el número de repeticiones de las pruebas para medir la respuesta de las pupilas a fin de lograr niveles comparables de discriminación entre participantes verdaderos y engañosos de habla árabe, como habíamos encontrado para participantes que hablan inglés y español. No estamos seguros de por qué fue necesario hacer esos cambios.

Estudio de campo de la ODT

Recientemente completamos un estudio de validez de campo de la ODT que evaluó a los solicitantes para puestos en las oficinas mexicanas del fiscal general, inmigración y policía federal. Probamos a algunos solicitantes en torno al uso reciente de drogas ilegales (R1) en comparación con la corrupción, el tráfico de armas o la afiliación a una organización terrorista religiosa (R2). La verdad fundamental para los casos engañosos fue que los solicitantes admitieron durante una prueba de polígrafo posterior que habían consumido drogas ilegales en el período cubierto por las declaraciones de la ODT ($n = 71$). La falta de admisión es insuficiente para establecer de manera concluyente que una persona fue veraz en la ODT. Para obtener datos de personas veraces, creamos una nueva prueba para los solicitantes de puestos de inmigración que les preguntó si habían cometido espionaje (R1) o sabotaje (R2). Asumimos que todas las personas evaluadas fueron veraces en sus respuestas a los dos temas relevantes porque la tasa base de engaño en esos temas en México es muy baja, especialmente para las personas que no tenían empleo gubernamental anterior ($n = 83$).

Para desarrollar y validar un modelo de decisión con los datos de campo, extrajimos medidas del movimiento ocular de los datos del rastreador ocular. Calculamos los coeficientes de validez de las medidas y utilizamos la regresión lineal para seleccionar un subconjunto de cuatro medidas para distinguir entre los grupos veraces y los grupos engañosos confirmados. Luego utilizamos las variables seleccionadas en una validación de 5 veces de un modelo de regresión logística binaria para clasificar los casos como veraces o engañosos. Para llevar a cabo la validación de 5 veces, dividimos la muestra de 154 casos de campo en cinco submuestras aleatorias de modo que cada submuestra constaba de aproximadamente el 20% de los casos engañosos ($n = 14$ o 15) y el 20% de los casos veraces ($n = 16$ o 17). La primera submuestra de 14 casos veraces y 16 casos engañosos ($N = 30$) se eliminó, y se creó un modelo de decisión con las cuatro submuestras restantes ($N = 124$). Utilizamos ese modelo de decisión para clasificar la muestra reservada de 30 casos y registramos el porcentaje correcto para cada grupo. La segunda submuestra se dejó de lado ($N = 30$), se desarrolló un nuevo modelo de decisión con los 124 casos restantes y se calculó la precisión de las clasificaciones para la segunda muestra reservada. Repetimos este proceso para las tres submuestras restantes. Los resultados se recogen en la Tabla 4.

Tabla 4. Porcentaje de decisiones correctas en cinco modelos de decisión en submuestras independientes

	Submuestra 1	Submuestra 2	Submuestra 3	Submuestra 4	Submuestra 5	Media	Media
veraz	N=30	N=30	N=31	N=31	N=32		N=154
engañoso	75.0	87.5	88.2	88.2	100.0	87.8	86.1
	100.0	71.4	85.7	78.6	86.7	84.5	

El porcentaje de decisiones correctas varió entre las cinco submuestras de 75% a 100% para los solicitantes veraces y de 71.4% a 100% para los solicitantes engañosos. Nuestras mejores estimaciones del rendimiento del modelo basado en los 154 casos cuando se analizaron en una nueva muestra de casos de campo son las medias para solicitantes veraces (87.8%) y engañosos (84.5%). Si bien los resultados obtenidos en

la validación quintuple utilizando solicitantes reales para cargos en el gobierno mexicano sugieren que la ODT puede agregar valor a un programa de selección previa al empleo, hasta cierto punto nuestros métodos pueden haber aprovechado el azar y producido tasas de precisión que son demasiado optimistas. Específicamente, utilizamos toda la muestra para seleccionar el subconjunto de medidas del movimiento ocular para el modelo de decisión que se validó posteriormente. Este subconjunto particular de medidas funcionó bien en la validación de 5 veces, pero podría ser subóptimo para un nuevo conjunto de casos de campo confirmados. Por este motivo, recomendamos que el modelo de decisión actual se reevalúe con nuevos casos desde la configuración de campo representativa.

Conclusiones

Los resultados de las investigaciones de laboratorio y de campo indican que la ODT arroja una precisión superior al 80% tanto con los examinados veraces como con los engañosos, aunque las tasas de precisión tienden a ser entre un 3% y un 6% más altas para los individuos veraces que para los engañosos. Las tasas de precisión se generalizan en los idiomas inglés y español, en las culturas estadounidenses y mexicanas y, en menor medida, en el árabe en el Medio Oriente. Creemos que la ODT es una nueva tecnología prometedora que es la más adecuada para las aplicaciones de detección. También creemos que se adapta mejor a las aplicaciones de detección que a la investigación criminal de incidentes específicos, ya que sería difícil elaborar un ECA que contenga dos problemas relevantes no superpuestos con validez nominal. Por la misma razón, los examinadores de polígrafo rara vez utilizan la pregunta del Complejo de Culpa para los exámenes de polígrafo de incidentes específicos (Office of Technology Assessment, 1983).

La investigación piloto no publicada con lectores deficientes indicó, como se esperaba, que la ODT es ineficaz. Parece que las dificultades cognitivas experimentadas por los examinados que luchan por comprender las afirmaciones de las pruebas eclipsan los efectos del engaño en las medidas del movimiento ocular. Estamos explorando alternativas audiovisuales y solo de audio a la ODT actual basada en texto. Sin embargo, hasta ahora no hemos desarrollado una versión de audio funcional de la prueba que pueda ser efectiva para una población más amplia de personas, incluidas aquellas con habilidades de lectura deficientes.

Con entrenamiento y práctica especializados, las pruebas de polígrafo pueden ser superadas (Honts, 2012). Hasta la fecha, no ha habido intentos de investigar los efectos de las contramedidas en la ODT. Sin embargo, debido a que los examinados están bajo presión para responder de manera rápida y precisa, los intentos de implementar contramedidas pueden ser ineficaces o fácilmente detectados con medidas de comportamiento, tales como tiempo de respuesta e índices de error.

Se necesita investigación adicional para explorar estas posibilidades.

Finalmente, la investigación sobre la ODT se ha llevado a cabo principalmente en un solo laboratorio por un equipo de investigadores. A medida que el equipo de rastreo ocular y el software de análisis de datos de alta calidad estén disponibles a costos más bajos, esperamos que otros investigadores contribuyan con nuevos conocimientos a esta área de la ciencia aplicada.

Referencias

- Bradley, M.M., Micolli, L., Escrig, M.A., Lang, P.J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45, pp. 602-607.
- Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology*, 18, pp. 307-315.
- Cook, A.E., Hacker, D.J., Webb, A.K., Osher, D., Kristjansson, S., Woltz, D.J., & Kircher, J.C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. *Journal of Experimental Psychology: Applied*, 18(3), pp. 301-313.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, pp. 297-334.
- Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38, pp. 205-211.
- Hacker, D.J., Kuhlman, B., & Kircher, J.C., Cook, A.E., & Woltz, D.J. (2014). De-tecting deception using ocular metrics during reading. In D.C. Raskin, C.R. Honts, & J.C. Kircher (Eds.), *Credibility assessment: Scientific research and applications*. Elsevier, pp 159-216.
- Hess, E.H., Polt, J.M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132, pp. 349-350.
- Hess, E.H., Pok, J.M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, 143, -1190-1192.
- Honts, C.R. (2012). Countermeasures and credibility assessment. In DC Raskin, CR Honts, & JC Kircher (Eds.), *Credibility assessment: Scientific research and applications*. Elsevier, pp. 131-156.
- Johnson, R., Jr., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research*, 24, pp. 386-404.
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall: New Jersey.
- Kircher, J. C. (1981). *Psychophysiological processes in the detection of deception*. Department of Psychology, University of Utah, Salt Lake City, UT. Unpublished manuscript.
- 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, pp. 79-90.
- Kircher, J.C., Raskin, D.C., Honts, C.R., & Horowitz, S.W. (1994). Generalizability of statistical classifiers for the detection of deception. *Psychophysiology*, 31, S73. (Abstract)
- Loewenfeld, I. E. (1999). *The pupil: Anatomy, physiology, and clinical application* (Vol 1). Boston: Butterworth-Heinemann.
- Middle East (2016a). *Ocular-motor detection of deception with Middle Eastern college students*, Phase 1. Unpublished research. University of Utah, Department of Educational Psychology.
- Middle East (2016b). *Ocular-motor detection of deception with Middle Eastern college students*, Phase 2. Unpublished research. University of Utah, Department of Educational Psychology.
- NSA (2012). *Ocular-motor detection of deception with employees at the National Security Agency Phase 1*. Unpublished research. University of Utah, Department of Educational Psychology.
- NSA (2013). *Ocular-motor detection of deception with employees at the National Security Agency Phase 2*. Unpublished research. University of Utah, Department of Educational Psychology.
- Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation*. OTA-TM-H-15. Washington, DC: US Government Printing Office.

INVESTIGACIÓN EN LABORATORIO Y EN EL CAMPO SOBRE
LA PRUEBA DEL ENGAÑO POR OBSERVACIÓN DEL MOVIMIENTO OCULAR

- Osher, D. (2005). *Multi-method assessment of deception: Oculomotor movement, pupil size, and response time measures*. Unpublished dissertation, University of Utah, Department of Educational Psychology.
- Patnaik, P. (2013). *Ocular-motor methods for detecting deception: Direct versus indirect interrogation*. Unpublished master's thesis, University of Utah, Department of Educational Psychology.
- Patnaik, P. (2015). *Oculomotor methods for detecting deception: Effects of practice feedback and blocking*. Unpublished dissertation, University of Utah, Department of Educational Psychology.
- Patnaik, P., Woltt, D.J., Hacker, D.J., Cook, A.E., Ramm, M.L., Webb, A.K., & Kircher, J.C. (2016). Generalizability of an ocular-motor test for deception to a Mexican population. *International Journal of Applied Psychology*, 6(1), pp. 1-9.
- Raskin, D.C. & Kircher, J.C. (2014). Validity of polygraph techniques and decision methods. In D.C. Raskin, C.R. Honts, & J.C. Kircher (Eds.), *Credibility assessment: Scientific research and applications*. Elsevier. pp. 63-129.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, pp. 372-422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10, pp. 241-255.
- Steinhauer S. R., Boller F., Zubin J., Pearlman S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology*, 20, p. 472.
- Steller, M. (1987). *Psychophysiologische Aussagebeurteilung* [Psychological assessment]. Hogrefe: Gottingen.
- Stern J.A., Walrath L.C., Goldstein, R. (1984). The endogenous eyeblink. *Psycho-physiology* 21, pp. 22-33.
- USTAR (2010). Oculomotor deception detection. Unpublished research funded by the Utah Science Technology and Research Initiative.
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10, pp. 141-142.
- Webb, A.K. (2008). *Effects of Motivation, and Item Difficulty on Oculomotor and Behavioral Measures of Deception*. Unpublished dissertation, University of Utah, Department of Educational Psychology.
- Webb, A. K, Honts, C.R., Kircher, J. C., Bernhardt, P.C., & Cook, A. E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminal Psychology*, 14(2), pp. 279-292.