

# Generalizabilidad de un examen oculomotor para detectar el engaño a una población mexicana

Pooja Patnaik<sup>1\*</sup>, Dan J. Woltz<sup>2</sup>, Douglas J. Hacker<sup>2</sup>, Anne E. Cooke<sup>2</sup>,  
Maria de Lourdes Francke Ramm<sup>3</sup>, Andrea K. Webb<sup>1</sup>, John C. Kircher<sup>2</sup>

<sup>1</sup>Draper, Cambridge, MA, EE.UU.

<sup>2</sup>Psicología Educacional, Universidad de Utah, Salt Lake City, EE.UU.

<sup>3</sup>Departamento de Psicología, Tec de Monterrey, Monterrey, México

**Resumen** Hemos desarrollado un examen oculomotor para detectar el engaño (EODE) que clasifica a la gente como veraz o engañosa según los movimientos de los ojos y las respuestas de las pupilas mientras los examinados leen y contestan preguntas tipo verdadero o falso con respecto a su posible participación en actividades ilícitas. El propósito del presente estudio era investigar si los efectos del engaño en las mediciones oculomotoras se generalizan a los hispanohablantes nativos en México. Ciento cuarenta y siete estudiantes de una universidad grande de México participaron en un experimento con simulacro de crimen; 83 eran culpables y 64 eran inocentes de tomar \$200 pesos de la cartera de una secretaria. Al efectuar la validación cruzada, la precisión de las clasificaciones según las mediciones de conducta y oculomotoras excedió el 80% tanto para los participantes veraces como para los engañosos en las poblaciones mexicana y estadounidense.

**Términos clave** Detección del engaño, oculomotor, mediciones de lectura, psicología transcultural

## 1. Introducción

Desde 2007, el gobierno mexicano, en combinación con el Departamento de Estado de Estados Unidos, ha fomentado las investigaciones anticorrupción de empleados gubernamentales. El gobierno de Estados Unidos ha destinado más de US\$1,000,000,000 para ayudar a eliminar la corrupción dentro de los rangos de oficiales de policía y seguridad en México. De esos fondos, los funcionarios mexicanos han invertido decenas de millones de dólares en equipo poligráfico, capacitación y exámenes a solicitantes de empleo en la policía y a empleados de dichos organismos.

Lamentablemente, los exámenes poligráficos toman mucho tiempo y son laboriosos, requiriendo 1.5 a 3 horas con un examinador capacitado. La administración de exámenes poligráficos a fiscales federales y funcionarios de policía es permitida por la ley mexicana, pero los resultados por sí solos no son suficientes para despedir a un empleado [1, 2]. Las exigencias de tiempo y trabajo que requieren los exámenes poligráficos junto con la necesidad de las investigaciones adicionales de los empleados limitan su viabilidad para disuadir la corrupción en México.

### 1.1. Detección del engaño

Cook et al. desarrollaron un nuevo método para detectar el engaño llamado examen oculomotor para detectar el engaño (EODE) [3]. En contraste con el polígrafo, el EODE es automatizado y puede administrarse en aproximadamente 40 minutos. Una computadora da instrucciones escritas y hace preguntas

tipo verdadero o falso con respecto a la posible participación del examinado en actividades ilícitas. El examinado lee las preguntas que le hace la computadora individualmente y usa un teclado para contestar mientras un rastreador de ojos remoto registra los movimientos de sus ojos y los cambios en el tamaño de sus pupilas. La computadora procesa los datos oculomotores, combina sus mediciones en una regresión logística, y clasifica al individuo como veraz o engañoso en el examen. Con suficiente evidencia de validez en un entorno de campo con lectores hispanos, el gobierno mexicano podría usar el examen por sí sólo o decidir que conviene administrar también un examen poligráfico.

La mayoría de teorías de detección del engaño sostienen que mentir es cognitivamente más exigente que decir la verdad [4-6]. Se necesitan recursos cognitivos para inhibir la verdad, fabricar la mentira y mantener su constancia, coherencia y credibilidad al paso del tiempo. La inhibición de respuestas veraces prepotentes, la retención de la credibilidad al paso del tiempo y la automonitorización para evitar discurso involuntario son procesos cognitivos que requieren esfuerzo mental.

Durante las últimas cuatro décadas, estudios psicofisiológicos han demostrado que las pupilas proporcionan un índice sensible del esfuerzo cognitivo (e.g., [7, 8]). Generalmente, entre mayor es la carga cognitiva, mayor es el aumento en el tamaño de las pupilas. De conformidad con la idea de que el engaño requiere más esfuerzo mental que decir la verdad, los psicofisiólogos también han comprobado que las reacciones pupilares muestran la diferencia entre individuos veraces y engañosos durante exámenes poligráficos (e.g., [9, 10]).

Las pupilas reaccionan no sólo a la carga cognitiva sino

\*Autor corresponsal:

pooja.patnaik@gmail.com (Pooja Patnaik)

Publicado en línea en <http://journal.sapub.org/ijap>

Copyright © 2016 Scientific & Academic Publishing. Reservados todos los derechos

también a los estímulos emocionales. Varios investigadores han reportado que los estímulos emocionales evocan reacciones en los examinados, la magnitud de las cuales depende de la intensidad, pero no del valor del estímulo emocional [11-14]. Los exámenes poligráficos se basan en la idea de que los individuos engañosos muestran reacciones emocionales más fuertes a ciertas preguntas del examen que a otras. Al grado que las reacciones emocionales a las preguntas del examen ayuden a distinguir entre individuos engañosos y veraces, las reacciones pupilares deberían reflejar esas diferencias y ser diagnósticos de engaño.

Un lector que tiene dificultades para leer o comprender el texto muestra más fijaciones de la mirada, agrandamiento de las pupilas y tiempos de lectura más largos [15, 16]. El número de fijaciones es un conteo del número de fijaciones que caen en un área definida de interés alrededor de un texto. El tiempo de lectura es la suma de las duraciones de las fijaciones que caen dentro del área de interés. Los investigadores a veces dividen el tiempo de lectura en lectura inicial del texto (duración de la primera pasada) y relectura subsiguiente del texto.

Cook et al. midieron los efectos del engaño en las reacciones de las pupilas, las conductas de lectura y las mediciones de conducta. Los participantes fueron asignados a condiciones de tratamiento como culpables o inocentes. Los participantes culpables cometieron un robo simulado de \$20 de la cartera de una secretaria o descargaron la información de la tarjeta de crédito de la computadora de un estudiante graduado. A todos los participantes se les pidió que contestaran rápida y precisamente preguntas tipo verdadero o falso presentadas en serie en el monitor de una computadora mientras que los movimientos oculares y los cambios en el tamaño de las pupilas eran registrados por un rastreador de ojos. Las preguntas tenían que ver con el robo de los \$20, el robo de información o eran de contenido neutral (por ejemplo: Los osos polares son originarios de México).

Tal y como se esperaba, el engaño se vio asociado con los mayores aumentos en el tamaño de las pupilas, y los participantes culpables también tomaron más tiempo para leer y contestar las preguntas del examen. Aunque los efectos principales de la culpa en las mediciones de tiempos de lectura y respuesta a lo largo de todos los tipos de preguntas fueron congruentes con la literatura sobre la lectura, el estatus de culpa interactuó con el tipo de pregunta, y esas interacciones no fueron congruentes con las expectativas iniciales basadas en la literatura sobre la lectura. En ambos experimentos, los participantes culpables tomaron considerablemente menos tiempo para leer y contestar engañosamente las preguntas del examen relacionadas con el crimen que habían cometido que cuando contestaron verazmente las otras preguntas. Para los participantes culpables, las respuestas engañosas se caracterizaron por menos fijaciones, y tiempos de lectura y relectura más breves si se comparan con sus respuestas veraces. Cook et al. concluyeron que cuando a los participantes se les informó que no pasarían el examen si no contestaban rápida y precisamente las preguntas, los participantes culpables hicieron un esfuerzo concertado para pasar el menor tiempo posible en las preguntas incriminatorias para evitar la detección.

## 1.2. El presente estudio

A la fecha, todos los experimentos de laboratorio con el EODE se han restringido a participantes anglohablantes en Estados Unidos. Las investigaciones han demostrado que miembros de todas las culturas creen que los mentirosos experimentan temor, vergüenza o dificultades cognitivas [17, 18],

pero las culturas difieren en sus normas para la comunicación interpersonal [19], y mentir es una forma de comunicación interpersonal [20]. Además de las diferencias culturales entre Estados Unidos y México, sus idiomas (inglés y español) difieren en las dimensiones semántica, pragmática, sintáctica, ortográfica y morfológica.

Una simple teoría que depende de la carga cognitiva y la emoción no predice que el idioma puede afectar las mediciones de conducta o fisiológicas observadas durante el EODE. Sin embargo, en contraste con los resultados de los experimentos con simulacro de crimen en Estados Unidos, no descubrimos casi ningún efecto del engaño en las mediciones oculomotoras en un estudio de campo no publicado que se realizó en Colombia. Aunque los errores en la medición de criterios (verdad fundamental) probablemente contribuyeron a la baja precisión del EODE en ese estudio, también creemos que los participantes no poseían aptitud de lectura para mostrar los efectos del engaño. Los individuos engañosos que tienen problemas para entender el significado de las preguntas de un examen no exhiben los cambios de patrón esperados en la lectura, las pupilas o las mediciones de conducta, y no son buenos candidatos para un EODE.

Aunque la aptitud de lectura podría explicar la menor precisión en el campo que en el laboratorio, las diferencias culturales o lingüísticas entre las poblaciones de Estados Unidos y Colombia también podrían explicar la inhabilidad del EODE para generalizar a lo largo de entornos. La meta del presente estudio era examinar si las diferencias culturales o lingüísticas moderan los efectos del engaño en las mediciones oculomotoras. Utilizamos los procedimientos detallados en Cook et al. para realizar un experimento con simulacro de crimen en una universidad mexicana grande. Siguiendo las teorías de Cook et al., evaluamos los efectos del engaño en las mediciones oculomotoras individuales. Luego combinamos los conjuntos de datos de México y Estados Unidos para examinar si los efectos del engaño en las mediciones oculomotoras varían entre especificaciones. El fracaso del EODE en una población de lectores mexicanos competentes indicaría que la tecnología no se generaliza bien a otras culturas o idiomas.

## 2. Método

### 2.1. Participantes

Ciento setenta y dos participantes aceptaron la invitación en el campus de una universidad urbana en el noreste de México. Se postearon volantes de invitación en el campus anunciando la oportunidad de ganar \$200 pesos (aproximadamente \$15 dólares estadounidenses en ese momento) y un posible bono de \$300 pesos (aproximadamente \$23 dólares estadounidenses en ese momento) por participar en un experimento psicológico. Se programó una sesión con los interesados que hablaban español fluido, que tenían más de 18 años de edad y podían leer la pantalla de una computadora sin anteojos. De estos 172 individuos, 27 participantes fueron omitidos debido a problemas con la calibración, por no seguir las instrucciones y/o por registros inadecuados. Dos participantes culpables fueron excluidos debido a que sus índices de error al responder excedían el 50% (azar). Los otros 145 participantes tenían edades de 18-52 años. Eran estudiantes de la universidad, predominantemente mexicanos

hispanos / latinos, y el español era su idioma principal. Sesenta y dos de los 145 participantes eran inocentes de los crímenes simulados (24 mujeres) y 83 eran culpables (29 mujeres).

## 2.2. Reseña del diseño y del procedimiento

El diseño fue del tipo 2 x (3 x 5) mixto. El factor entre-sujetos fue la culpa con dos niveles (culpable o inocente). Los factores intrasujeto fueron los tipos de pregunta (neutral, efectivo y examen) y la repetición (5 repeticiones de las preguntas del examen EODE). El tiempo con 40 niveles (10 muestras Hz x 4 segundos) también se incluyó como factor intrasujeto para el análisis del diámetro pupilar (DP).

## 2.3. Aparato

Un rastreador de ojos SensoMotoric Instruments (SMI) RED-m remoto incorporado al monitor de una computadora de escritorio registró los movimientos oculares y el diámetro de las pupilas a 60 Hz. La visualización fue binocular y se utilizó un descansabarrillas para mantener inmóvil la cabeza del participante. Los estímulos se le presentaron al participante en un monitor LCD Lenovo con pantalla plana de 19 pulgadas con proporción de aspecto de 5:4. El monitor se colocó aproximadamente a 65 cm de los ojos del participante.

## 2.4. Examen oculomotor para detectar el engaño

Las instrucciones y preguntas de práctica se presentaron en español al participante en letra negra con fondo gris pálido. Los participantes contestaron 15 preguntas de práctica seguidas por 48 preguntas de examen, y estas mismas 48 preguntas fueron presentadas cinco veces en órdenes diferentes. 16 preguntas tenían que ver con el robo de los \$200 pesos (por ejemplo: Yo no tuve nada que ver con el robo de los \$200 pesos), 16 tenían que ver con el robo del examen (por ejemplo: Yo no tomé nada de la oficina del profesor) y 16 eran preguntas neutras (por ejemplo: Yo tengo menos de 75 años de edad). Las preguntas se arreglaron de tal manera que no aparecieran dos de la misma categoría en sucesión. Las afirmaciones se presentaron una a la vez y los caracteres V/F(verdadero/falso) aparecieron a la extrema derecha de cada afirmación para recordar a los participantes sus opciones de respuesta. La respuesta correcta (es decir, no incriminatoria) era verdadera para 8 de 16 preguntas en una categoría y falsa para las restantes 8 preguntas en cada categoría.

Entre las repeticiones de las 48 preguntas del examen, los participantes completaron una tarea intermedia consistente en 18 preguntas tipo verdadero y falso sobre conocimiento mundial general las cuales fueron diseñadas para proporcionar un descanso y despejar la memoria de trabajo de las preguntas y respuestas del EODE.

## 2.5. Mediciones dependientes

Mediciones de resultados de conducta. Las mediciones de conducta consistieron en porcentaje de error y tiempo de respuesta.

El porcentaje de error para un tipo de pregunta en particular (neutral, efectivo, examen) fue el número de respuestas incorrectas dividido entre el número de preguntas ( $16 \times 5 = 80$ ) multiplicado por 100.

El tiempo de respuesta (TR) fue el tiempo en minutos desde la aparición de la pregunta en la pantalla hasta el momento en que el participante presionó un botón para contestar la pregunta.

**Mediciones de resultados oculomotores.** Se definió un área de interés (ADI) para cada pregunta tipo V/F del examen. El ADI comenzaba con el primer carácter de la pregunta y terminaba con el punto final de la misma. Se computaron mediciones oculomotoras de lectura para las fijaciones en cada ADI. Las fijaciones se determinaron usando los archivos de datos producidos por el rastreador de ojos SMI mediante la identificación de una secuencia de muestras en las cuales los ojos mostraban poco movimiento durante por lo menos 100 ms. Las fijaciones mayores de 1,000 ms se consideraron como artefactos y se descartaron [15]. El tiempo de lectura fue la suma de duraciones de las fijaciones que cayeron dentro del área de interés.

El número de fijaciones fue el número de fijaciones detectadas en el ADI.

La duración de la primera pasada fue la suma de todas las duraciones de fijaciones dentro del ADI antes de que los ojos se fijaran fuera de la misma.

La duración de relectura fue la suma de las duraciones de fijaciones asociadas con los movimientos oculares hacia la izquierda en el ADI.

Las muestras de datos del DP del comienzo de un bloque de 48 preguntas de examen hasta el final de ese bloque fueron estandarizadas (convertidas a calificaciones-z) dentro de los participantes. Para cada pregunta del examen, el área de DP debajo de la curva (DP ADC) se obtuvo de la curva estandarizada de reacciones pupilares que comenzó al momento en que la pregunta del examen apareció en la pantalla de la computadora y terminó 4 s más tarde. La computadora identificó los puntos alto y bajo en esta curva de reacciones de cuatro segundos y computó la diferencia entre cada punto bajo y cada punto alto subsiguiente. La amplitud pico fue la mayor diferencia observada, y la reacción inicial se definió como el punto bajo desde el cual se midió la amplitud pico. En el paso final, la medición del área se obtuvo mediante la integración del área debajo de la curva de reacción pupilar desde el inicio de la reacción al punto en el cual la reacción volvió al nivel inicial o al final del intervalo de muestra de cuatro segundos, el que ocurriera primero.

El nivel de DP al momento de la respuesta V/F fue la media de calificaciones estándar para un período que comenzó 1 s antes de la respuesta V/F del participante y terminó 1 s después de la respuesta V/F.

## 2.6. Procedimientos

Los participantes se presentaron solos a un cuarto de un edificio en el campus. Se les dejaron instrucciones en un sobre pegado a la puerta con cinta adhesiva donde se les indicaba que entraran en el salón, y que leyeran y firmaran el formulario de consentimiento. El participante entonces escuchaba una grabación que le daba las instrucciones del estudio. A todos los participantes se les informó que algunos de ellos robarían un examen de la computadora de un profesor, otros tomarían \$200 pesos de la cartera de una secretaria, y un tercer grupo de participantes sería inocente de ambos crímenes simulados. En realidad, sólo se formaron dos grupos; los participantes eran culpables de tomar los \$200 de la cartera de la secretaria o eran inocentes del crimen. A todos los participantes se les prometió un bono de \$300 pesos si pasaban el EODE. A los participantes se les dio un número telefónico para llamar si decidían no participar.

Cada participante culpable fue a la oficina de una secretaria y le preguntó dónde estaba ubicada la oficina del profesor. La secretaria (una mujer aliada) dijo al participante que el profesor

no trabajaba en ese edificio. El participante dio las gracias a la secretaria y salió de la oficina. El participante entonces espero inadvertidamente a que la secretaria dejara sola la oficina, entró en la misma, encontró la bolsa, sacó \$200 pesos de la billetera que había en la bolsa, y escondió el dinero en su persona. A los participantes culpables se les pidió que prepararan una coartada en caso de que los descubrieran y que no dejaran huellas digitales. Se les informó que no tenían más de 20 minutos para cometer el crimen y presentarse ante el examinador.

A los participantes inocentes se les dijo que se les harían preguntas en cuanto a los robos pero que no tomaran nada, y que debían esperar aproximadamente 20 minutos antes de presentarse ante el examinador.

Cuando los participantes se presentaron para el EODE, se sentaron frente a una computadora con la barbilla en el descansa-barbillas y efectuaron una breve calibración del rastreador de ojos. Después de calibrar el rastreador de ojos, la computadora administró el EODE. Después del examen, el examinador interrogó y pagó al participante. Además del pago básico de \$200 pesos, los participantes recibieron una bonificación de \$300 pesos si la computadora indicó que eran inocentes.

### 3. Resultados

La meta principal del presente estudio era determinar si los efectos del engaño en las mediciones de conducta y oculomotoras eran similares para estudiantes en una universidad mexicana y estudiantes en una universidad estadounidense. El análisis de varianza con mediciones repetidas (Repeated Measures Analysis of Variance, RMANOVA) se utilizó para analizar cada variable dependiente. Debido a que el EODE fue diseñado como un examen de comparación relevante, la evaluación de la detección del engaño dentro de cada medición se basó en la diferencia entre las medias de las personas para los dos

asuntos relevantes (es decir, preguntas de examen y preguntas sobre el efectivo). Se comprobó la detección del engaño mediante la comparación de este contraste para individuos culpables versus inocentes, y esto correspondió al efecto de interacción de la culpa y el contraste del contenido relevante. El nivel de significancia se estableció a  $p < .05$  en todo el análisis. También se han reportado análisis complementarios con la inclusión de 112 participantes (56 inocentes y 56 culpables) del Experimento 2 que se realizó bajo condiciones experimentales similares en Estados Unidos [3]. Analizamos la interacción de 3 vías entre la condición de culpable, el contraste de preguntas relevantes y el entorno (México versus Estados Unidos) para cada medición a fin de examinar si los efectos del engaño diferían en los dos entornos.

#### 3.1. Mediciones de conducta y lectura

La Ilustración 1 presenta las medias de errores al responder (Panel a) y tiempo de respuesta (Panel b). Las barras de error representan los intervalos del 95% de confianza calculados para comparaciones intrasujeto del tipo de pregunta. Como puede verse en la Ilustración 1a, los participantes culpables cometieron más errores al responder a lo largo de todos los tres tipos de preguntas en comparación con los inocentes,  $F(1,143) = 10.04$ ,  $\text{parcial } \eta^2 = .07$ . Además, hubo un efecto principal considerable para el contraste entre las preguntas relevantes  $F(1,143) = 17.47$ ,  $\text{parcial } \eta^2 = .11$ . En general, los participantes en ambas condiciones cometieron más errores al contestar las preguntas del examen en comparación con las preguntas sobre el efectivo. La condición de culpa no interactuó con el contraste del contenido de preguntas relevantes,  $F(1,143) < 1$ ,  $p > .05$ . Cuando se compararon los datos de Cook et al. con los datos actuales, la interacción de 3 vías de Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) < 1$ ,  $p > .05$ .

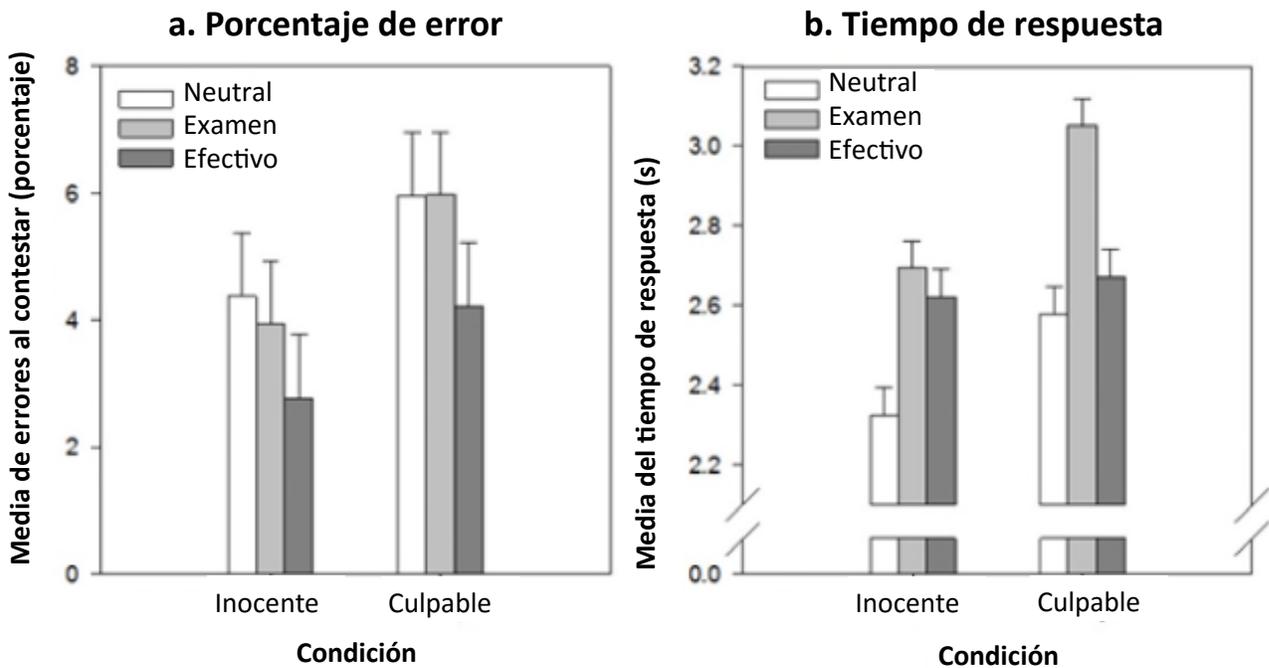


Ilustración 1. Media del porcentaje de error y del tiempo de respuesta por contenido de preguntas y condición de culpable

Para el tiempo de respuesta, el efecto principal de la culpa a lo largo de todas las preguntas no fue considerable,  $F(1,143) = 3.52$ ,  $p = .06$ . El efecto principal del contraste de preguntas relevantes fue considerable,  $F(1,143) = 63.20$ ,  $\text{parcial } \eta^2 = .31$ . En general, los participantes contestaron más rápidamente a las preguntas sobre el efectivo en comparación con las preguntas del examen. Un dato de gran importancia es que la culpa interactuó con el contraste de preguntas relevantes,  $F(1,143) = 29.37$ ,  $\text{parcial } \eta^2 = .17$ . Como se puede ver en la ilustración 1b, los participantes inocentes contestaron similarmente las preguntas del examen y las preguntas sobre el efectivo, mientras que los participantes culpables contestaron más rápidamente las preguntas sobre el efectivo que las del examen. Este patrón fue similar al reportado por Cook et al. Cuando los datos del Experimento 2 de Cook et al. fueron incluidos con los datos actuales, la interacción de 3 vías de Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) < 1$ ,  $p > .05$ .

La Ilustración 2 presenta los resultados del tiempo de primera lectura por pregunta (Panel a) y el tiempo de relectura por pregunta (Panel b). Al igual que con el tiempo de respuesta, no hubo efecto principal por culpa en la duración de la primera lectura,  $F(1, 143) = 2.52$ ,  $p = .11$ . Hubo un efecto principal para el contraste de preguntas relevantes,  $F(1,143) = 13.51$ ,  $\text{parcial } \eta^2 = .09$ , con una duración de primera lectura más breve para las preguntas sobre el efectivo que para las del examen. Sin embargo, como es evidente en la Ilustración 2a, el efecto general del contenido de las preguntas se debió enteramente a la rápida primera lectura de los participantes culpables en las preguntas sobre el efectivo. Esto se reflejó en la interacción del contenido de preguntas relevantes con la culpa,  $F(1,143) = 37.91$ ,  $\text{parcial } \eta^2 = .21$ . El patrón descubierto por Cook et al. (2012) fue similar en que sólo los participantes culpables tuvieron duraciones de primera lectura más breves para las preguntas sobre el efectivo que para las del examen. La interacción de 3 vías de Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) < 1$ ,  $p > .05$ .

Para el tiempo de relectura, no hubo efecto principal de culpa,  $F(1, 143) = 1.75$ ,  $p = .19$ . No hubo efecto principal para el contraste de preguntas relevantes,  $F(1,143) = 36.74$ ,  $\text{parcial } \eta^2 = .20$ . Al igual que con la primera lectura, el promedio de

duración de relectura fue más breve para las preguntas sobre el efectivo que para las del examen. Algo importante es que la condición de culpa interactuó con el contraste de preguntas relevantes,  $F(1,143) = 19.78$ ,  $\text{parcial } \eta^2 = .12$ . Como puede verse en la Ilustración 2b, con respecto a los participantes inocentes, los individuos culpables pasaron menos tiempo releendo las preguntas sobre el efectivo que las del examen. Este patrón fue similar al reportado por Cook et al. (2012). La interacción de 3 vías de Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) = 1.89$ ,  $p = .28$ .

La Ilustración 3 presenta los resultados para el número de fijaciones por pregunta. No hubo efecto principal de culpa en el número de fijaciones,  $F(1, 143) < 1$ ,  $p > .05$ . Los participantes tuvieron menos fijaciones en las preguntas sobre el efectivo que en las del examen,  $F(1,143) = 13.30$ ,  $\text{parcial } \eta^2 = .09$ , pero no hubo interacción del contenido de preguntas relevantes con la culpa,  $F(1,143) = 2.80$ ,  $p = .10$ . Cuando se combinaron las poblaciones mexicana y estadounidense, la interacción de 3 vías entre Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) = 1.66$ ,  $p = .20$ . La interacción de 3 vías entre Entorno x Culpa x Contenido Relevante no fue considerable,  $F(1,253) < 1$ ,  $p > .05$ .

### 3.2. Mediciones pupilares

La Ilustración 4 muestra los cambios en la media a partir del valor inicial en el diámetro pupilar. Una diferencia positiva indicó un aumento en el DP con respecto al valor inicial, y una diferencia negativa indicó una reducción en el DP con respecto al valor inicial.

De uno a tres segundos después del inicio de la pregunta, la reacción pupilar comenzó a distinguir entre los tipos de pregunta. Los participantes inocentes mostraron una reacción pupilar ligeramente más fuerte a las preguntas del examen que a las preguntas sobre el efectivo, y las preguntas neutrales se asociaron con las respuestas más débiles. En contraste, los participantes culpables mostraron una reacción más fuerte a las preguntas sobre el efectivo que a las del examen. La reducción en la curva de las preguntas neutrales del segundo al 6 indica constricción pupilar y probablemente refleja recuperación de la reacción previa a una pregunta relevante.

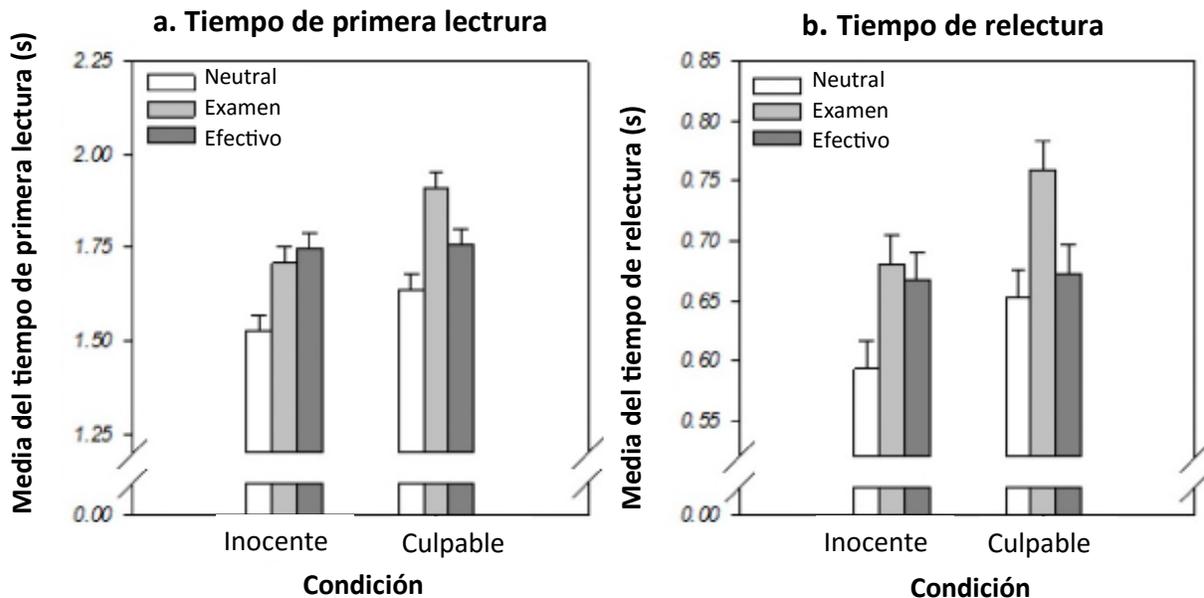
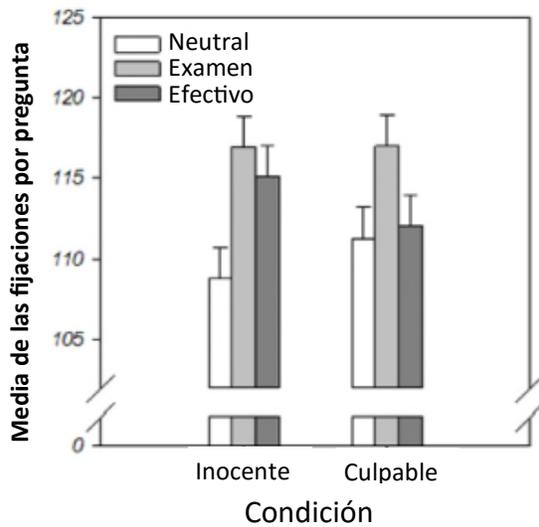


Ilustración 2. Media de los tiempos de primera lectura y relectura por contenido de preguntas y condición de culpable



**Ilustración 3.** Media del número de fijaciones por pregunta por contenido de preguntas y condición de culpable

La magnitud de la reacción pupilar según lo midió el DP ADC se muestra en la Ilustración 5a para cada grupo y tipo de pregunta. La Ilustración 5b muestra el nivel de DP al momento en que el participante contestó la pregunta para cada grupo y tipo de pregunta.

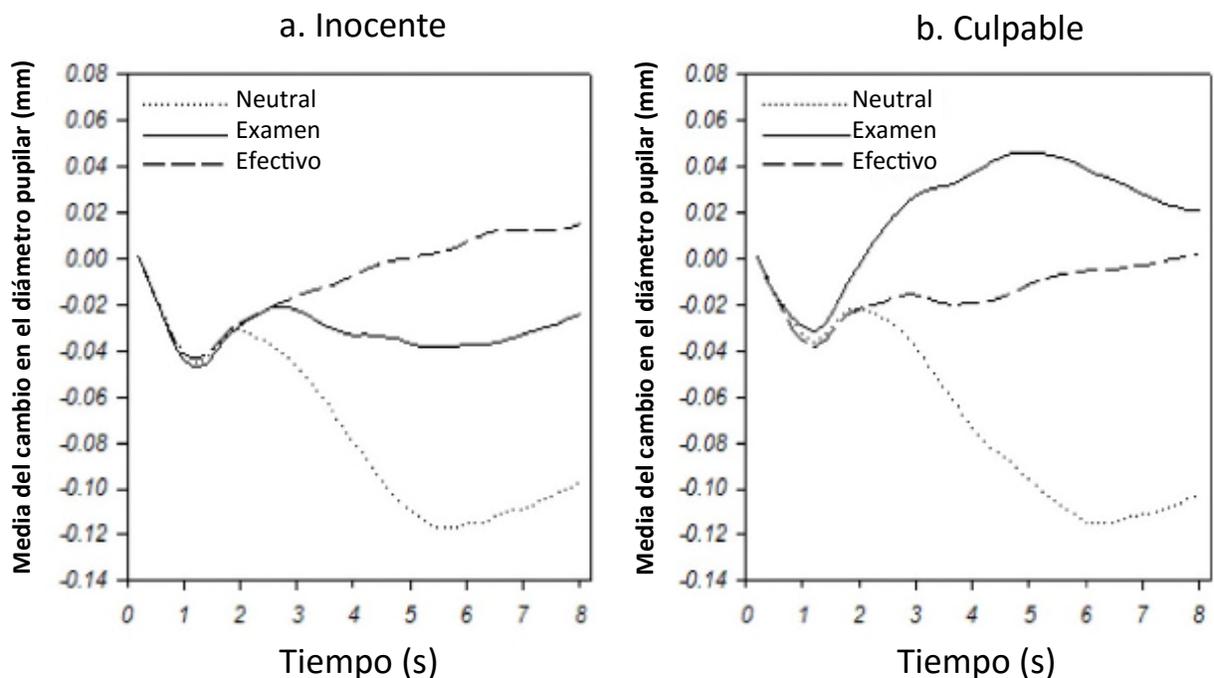
No hubo efecto principal de culpa a lo largo de todos los tipos de preguntas ni en el DP ADC,  $F(1, 143) < 1, p > .05$ , o nivel de DP,  $F(1, 143) = 1.53, p = .22$ . Hubo un pequeño efecto principal para el contraste de preguntas relevantes en DP ADC con valores menores para las preguntas de examen en contraste con las preguntas sobre el efectivo,  $F(1, 143) = 5.02$ ,  $\text{parcial } \eta^2 = .03$ . No hubo efecto principal para el contenido de preguntas relevantes en el nivel de DP. Un punto de importancia primordial es que hubo interacciones sustanciales entre

culpa y contenido de preguntas relevantes tanto para DP ADC,  $F(1, 143) = 60.03$ ,  $\text{parcial } \eta^2 = .30$ , como para nivel de DP,  $F(1, 143) = 73.22$ ,  $\text{parcial } \eta^2 = .34$ . Como puede verse en las Ilustraciones 5a y 5b, comparados con los participantes inocentes, los individuos culpables mostraron reacciones pupilares más fuertes durante las preguntas sobre el efectivo y reacciones más débiles durante las preguntas del examen. Estos patrones fueron similares a los reportados por Cook et al. La interacción de 3 vías, incluyendo el entorno, no fue considerable para DP ADC,  $F(1, 253) < 1, p > .05$ , ni para nivel de DP,  $F(1, 253) < 1, p > .05$ .

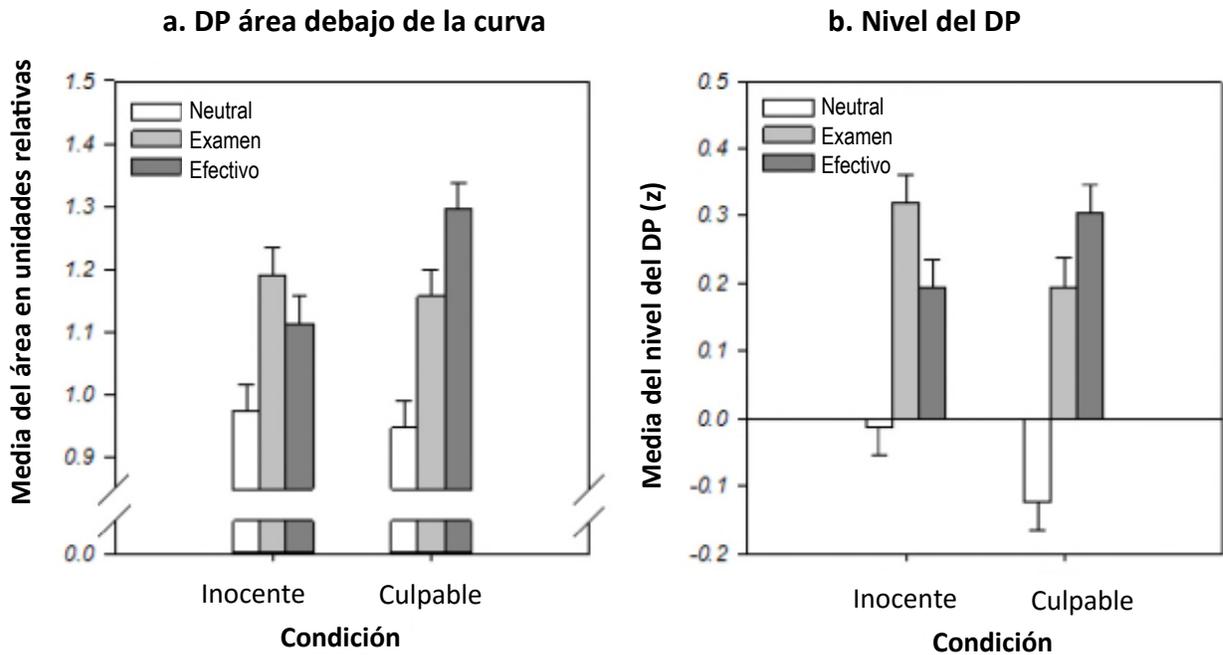
### 3.3. Validez predecible de las mediciones oculomotoras

Para evaluar la precisión de las clasificaciones efectuadas por el EODE con hispanohablantes, se realizó un análisis de regresión logística escalonada usando mediciones contrastantes que tuvieron efectos interactivos considerables con la culpa. El modelo de regresión resultante tenía tres mediciones que representaban las reacciones de conducta, lectura y pupilas a las preguntas de examen. Esta función clasificó correctamente 55 de los 63 participantes inocentes (87.3%) y 69 de los 82 participantes culpables (84.1%), para una precisión general del 85.5%. Estos resultados fueron similares a los reportados por Cook et al., donde las precisiones generales de clasificación fueron del 85% en el Experimento 1 y del 86% en el Experimento 2.

Para hacer más comparaciones de los resultados mexicanos actuales con los de la muestra poblacional de Estados Unidos, realizamos un análisis de validación cruzada doble con datos del Experimento 2 de [3]. El modelo mexicano clasificó correctamente el 80.4% de los inocentes y el 89.3% de los participantes culpables de Estados Unidos, para una media de precisión del 84.9%. Usando las variables seleccionadas en el modelo mexicano, el modelo estadounidense clasificó correctamente el 87.5% de los inocentes y el 81.9% de los participantes mexicanos culpables, para una media de precisión del 84.7%.



**Ilustración 4.** Media del cambio en el diámetro pupilar por 8 segundos después del inicio de la pregunta por contenido de preguntas y condición de culpable



**Ilustración 5.** Media del área debajo de la curva y del nivel de diámetro pupilar al momento de respuesta por contenido de preguntas y condición de culpable

#### 4. Discusión

El presente estudio de simulacro de crimen fue realizado en una universidad mexicana con participantes nativos hispanohablantes. El estudio reveló varios efectos grandes en las mediciones de conducta y oculomotoras usadas por el EODE para distinguir entre individuos hispanohablantes veraces y engañosos. Lo más importante es que el presente estudio no tuvo éxito en revelar ninguna evidencia de diferencias considerables entre los participantes mexicanos y estadounidenses en lo concerniente a ninguna medición de conducta u oculomotoras de engaño. A pesar de las diferencias en cultura e idioma, las mediciones de conducta de errores al contestar y tiempos de respuesta no distinguieron entre los entornos mexicano y estadounidense, como tampoco lo hicieron los patrones de lectura o las reacciones pupilares. Al efectuarse la validación cruzada, el EODE correctamente clasificó aproximadamente el 85% de participantes veraces y engañosos. Los altos y uniformes niveles de decisión precisa logrados por los modelos de regresión logística en las poblaciones mexicana y estadounidense son índices de la solidez de los efectos del engaño en las mediciones oculomotoras, así como de la estabilidad de las estructuras con covarianzas entre las mediciones y entre entornos.

Aunque es difícil sacar conclusiones firmes de resultados nulos, varios aspectos del diseño apoyados por una teoría rudimentaria alivian las preocupaciones en cuanto a la imposibilidad de detectar los efectos de los entornos. Para cada participante, hubo 80 preguntas por media para cada uno de los tres tipos de preguntas, y hubo 257 participantes en las muestras poblacionales mexicana y estadounidense. El agrupamiento de mediciones de las diferentes preguntas de examen del mismo tipo mejoró la confiabilidad de las mediciones de los participantes individuales, y los grandes números de casos proporcionaron exámenes estadísticos sensibles para los efectos de los entornos. El poder estadístico también se mejoró mediante la conducción de contrastes enfocados [21]. El análisis potenciado reveló que nuestros exámenes estadísticos tenían más del 80% de poder para detectar un efecto que sería responsable de menos del 3% de la varianza en una medición de resultados, y sin embargo ninguno de estos exámenes reveló efecto alguno del entorno en ninguna medición.

El EODE se basa en la idea de que es más difícil mentir que decir la verdad y que el engaño puede asociarse con una fuerte reacción emocional. Si estas ideas son correctas, los efectos del idioma en la validez de los diagnósticos de nuestras mediciones de lectura podrían ser demasiado pequeños para tener importancia alguna. Por otra parte, el marco teórico propuesto predeciría efectos de la cultura sobre el contexto engañoso creado en un experimento con simulacro de crimen si hubiese diferencias entre las culturas en desarrollo moral. Si el robo de dinero fuera algo más reprehensible en una cultura que en otra, la comisión del crimen podría evocar una reacción emocional más fuerte que podría afectar una o más de las mediciones resultantes. Sin embargo, no hubo evidencia de tal efecto en el presente estudio.

#### 4.1. Direcciones futuras

En el presente estudio, todos los participantes veraces tuvieron una experiencia común de pre examen y todos los individuos engañosos tuvieron otra. Las experiencias pre examen de individuos que se someten a un EODE en entornos de campo serían más diversas. Los examinados en el campo variarían más en edad, educación, inteligencia y habilidad de lectura en comparación con los participantes del presente experimento. La varianza en cualquier factor que modere los efectos del engaño sobre las mediciones oculomotoras tendría que controlarse para mantener altos niveles de precisión en las decisiones. Por ejemplo, esfuerzos no publicados para evaluar la eficacia del EODE con lectores deficientes en Colombia produjo apenas efectos débiles en las mediciones oculomotoras. Los lectores deficientes batallan para comprender el contenido de las preguntas del examen, y esos esfuerzos parecen ser de mayor peso que los efectos del engaño en las mediciones oculomotoras. Un examen de lectura podría administrarse antes del EODE para determinar si la persona es un candidato apto para el examen. De manera alternativa, los índices de error y los tiempos de respuesta en el EODE deberían proporcionar evidencia de que el examinado es capaz de producir datos de diagnóstico.

Otra pregunta que podría surgir es si un EODE diseñado para un incidente específico funcionaría con la misma eficacia si el asunto relevante fuera más general. En el presente estudio preguntamos a los participantes si robaron \$200 pesos de una cartera, mientras que en un examen para evaluar solicitantes de empleo, el examen podría preguntar si el solicitante alguna vez le robó a un empleador previo. La generalidad de las preguntas relevantes en un contexto de evaluación podría amenazar la validez de un EODE si los solicitantes no están seguros de ser culpables [22]. ¿Qué significa exactamente robarle a un empleador previo? Un individuo podría sentirse culpable por tomar papel del trabajo, mientras que otro individuo podría pensar que tomar papel del trabajo es irrelevante. A menos que los asuntos relevantes se definan claramente antes del examen, los diferentes individuos podrían interpretar el significado de las preguntas del examen diferentemente, y la precisión podría sufrir.

Aunque existen razones para esperar que las condiciones del campo [4] reducirán la precisión del EODE, también hay razones para pensar que las condiciones del campo aumentarán su precisión. Los examinados del campo posiblemente estarían más motivados para pasar el examen que los participantes de un experimento con simulacro de crimen. Cook et al. (2012) manipuló incentivos monetarios para pasar el examen y [5] descubrió que mayores incentivos mejoraron la validez de diagnóstico de algunas mediciones de lectura.

Existe evidencia adicional de que el EODE será eficaz en el campo, a pesar de la generalidad de los asuntos que se cubren en el examen. En 2012, la Sección de Narcóticos de la Embajada Estadounidense en México nos pidió que realizáramos una evaluación del EODE para investigar abogados que estaban solicitando empleo con la Procuraduría General de la República, el equivalente mexicano de la U.S. Office of the Attorney General. El asunto cubierto en el EODE era si el solicitante había usado alguna vez estupefacientes ilícitos. Un día después del EODE, al solicitante se le administró un examen poligráfico pre empleo. Durante una entrevista con el examinador, 26 de los 332 examinados confesaron haber mentido en cuanto a su uso de estupefacientes el día anterior en el EODE. El modelo de regresión logística para los estudiantes mexicanos en el presente estudio correctamente clasificó 23 de los 26 abogados como engañosos (88.5%), precisión que es ligeramente mayor que la del modelo con estudiantes mexicanos engañosos (84.1%). Por supuesto, no pudimos verificar el estatus de engaño de los solicitantes que no confesaron, pero los resultados de los casos de quienes lo hicieron sugieren que la precisión podría ser mayor en entornos de campo donde las consecuencias de no pasar el EODE son más serias que no recibir un bono en efectivo. Un importante paso siguiente sería realizar un estudio de campo donde exista una verdad fundamental para todos los participantes, y existan consecuencias de la vida real por no pasar el examen. Hasta que no conozcamos mejor los límites de generalizabilidad, el uso del EODE debe limitarse a individuos que compartan las características comunes de las muestras poblacionales de nuestro estudio y muestren resultados similares en las mediciones de conducta.

## OBRAS CITADAS

- [1] Byrne, E.V. (2013, February 15). Mexico's supreme court approves polygraph tests for federal prosecutors, but with some limitations. *The Mexico Gulf Reporter*. Retrieved from <http://www.mexicogulfreporter.com/2013/02/mexicos-supreme-court-approves.html>.
- [2] Byrne, E.V. (2014, March 12). Mexico's supreme court upholds police vetting process. *The Mexico Gulf Reporter*. Retrieved from <http://www.mexicogulfreporter.com/2014/03/mexican-supreme-court-upholds-police.html>.
- [3] Cook, A. E., Hacker, D. J., Webb, A.K., Osher, D., Kristjansson, S., Woltz, D. J., & Kircher, J. C. (2012). Lying eyes: Ocular-motor measures of reading reveal deception. *Journal of Experimental Psychology Applied*, 18(3), 301-313. doi: 10.1037/a0028307.
- [4] 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, 386-404. DOI: 10.1016/j.cogbrainres.2005.02.011.
- [5] Kircher, J. C. (1981, March). *Psychophysiological processes underlying the detection of deception*. Unpublished manuscript, University of Utah.
- [6] Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10, 141-142.
- [7] Kahneman, D., Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583-1585.
- [8] Klingner, Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance.
- [9] Bradley, M. T., & Janisse, M. P. (1979). Pupil size and lie detection: The effect of certainty on deception. *Psychology: A Quarterly Journal of Human Behavior*, 16, 33-39.
- [10] 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), 279-292. doi:10.1348/135532508X398602.
- [11] Bradley, M.T., Micolli, L., Escrig, M.A., Lang, P.J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45, 602-607.
- [12] Hess, E.H., Polt, J.M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132, 349-350.
- [13] Hess, E.H., Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, 143, 1190-1192.
- [14] Steinhauer S. R., Boller F., Zubin J., Pearlman S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology*, 20, 472.
- [15] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422. doi: 10.1037/0033-2909.124.3.372
- ess, E.H., Polt, J.M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132, 349-350.

- [16] 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*, 241-255. doi:10.1207/s1532799xssr1003\_3
- [17] Bond C.F., Jr, Robinson M.A. (1988) The evolution of deception. *Journal of Nonverbal Behavior, 12*, 295-308.
- Poulos, H. G., 1971, Behavior of laterally loaded piles-I: Single piles., *J. Soil Mech. and Found. Div.*, 97(5), 711-731.
- [18] Ekman P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.
- [19] Anderson, P. A., Hecht, M. L., Hoobler, G. D., & Smallwood, M. (2002). Nonverbal communication across cultures. In W. G. Gudykunst & B. Mody (Eds.), *Handbook of international and intercultural communication* (pp. 89-106). Thousand Oaks, CA: Sage.
- [20] Buller, D.B. and J.K. Burgoon (1996). Interpersonal Deception Theory. *Communication Theory, 6*(3), 203-242.
- [21] Keppel, G. & Wicken, T. D. (2004). *Design & analysis: A researcher's handbook* (4rd Ed.). New Jersey: Prentice Hall.
- [22] Meijer, E.H., Verschuere, B., Merckelbach, H.L.G.J., & Crombez, G. (2008). Sex offender management using the polygraph: A critical review. *International Journal of Law and Psychiatry, 31*(5), 423-429.