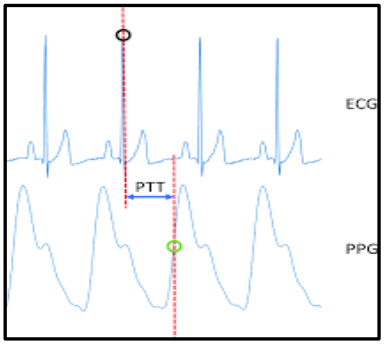
Prueba de comparación de múltiples temas híbrida EyeDetect (HMCT)
Resumen de desarrollo y validación

28 de mayo de 2021

Recientemente Converus lanzó la prueba de comparación de múltiples temas híbrida (HMCT) de EyeDetect. El protocolo HMCT es similar al MCT de base óculo motora y al MCT (AMCT) de audio. El HMCT cubre hasta cuatro temas relevantes, calcula la probabilidad de que un sujeto de prueba haya sido engañoso acerca de cada uno de los temas relevantes y utiliza probabilidades para clasificar al sujeto como veraz para todos los temas relevantes o engañoso para uno o más de ellos. El HMCT se diferencia de todas las tecnologías de detección de engaño anteriores en que combina los métodos tradicionales de prueba del polígrafo con los métodos óculo motores.

Las pruebas que utiliza el protocolo HMCT se realizan en dos fases. La primera fase es muy parecida a una prueba de polígrafo tradicional, pero está automatizada. La computadora presenta preguntas de prueba al sujeto con una voz de texto a voz (digital) por auriculares mientras graba la actividad electro dérmica, respiratoria, cardiovascular y óculo motora. La segunda fase consta de dos sesiones del protocolo MCT de base óculo motora. Mientras que la fase de polígrafo de la prueba asume que el engaño se asocia principalmente con la excitación emocional, la fase MCT asume que el engaño es cognitivamente más exigente que decir la verdad. En la fase MCT de la prueba, un guión automatizado anima al sujeto a leer y responder de la manera más rápida y precisa posible a enunciados de tipo Verdadero / Falso presentados en la pantalla. Esta prueba utiliza la presión del tiempo para aumentar las demandas cognitivas de la tarea al sujeto.

Al igual que con otros polígrafos digitales, los registros electro dérmicos se obtienen a partir de electrodos desechables adheridos a la superficie palmar de los dedos de la mano no dominante, y la respiración se registra con un medidor de tensión alrededor del pecho o el abdomen. La estación EyeDetect+ 2.0 también registra el electrocardiograma (ECG) de los electrodos conectados a cada brazo y la actividad vasomotora periférica con un fotopletismógrafo (PPG) de un dedo de la mano no dominante. A partir de las señales de ECG y PPG, la computadora obtiene una medida conocida como tiempo de tránsito del pulso (PTT). PTT es la diferencia en el tiempo desde que late el corazón hasta que aparece el pulso en el dedo (ver la imagen acontinuación).



Se ha demostrado que el PTT varía inversamente con los cambios en la presión arterial (Geddes et al., 1981; Obrist et al., 1978; Obrist et al., 1979). A medida que aumenta la presión arterial, disminuye el tiempo que tarda el pulso en viajar desde el corazón a la periferia. Si bien los cambios en el PTT pueden ser del orden de solo 5 ms, las investigaciones han demostrado que el PTT es al menos tan diagnóstico como el cardiógrafo tradicional (Webb y Kircher, 2005). EyeDetect+ 2.0 reemplaza el electrocardiógrafo con PTT. El PTT es menos invasivo que el electrocardiógrafo, alivia la incomodidad del sujeto y ofrece más flexibilidad en la construcción de la prueba. La

duración de los gráficos de polígrafo han estado limitado por el electrocardiógrafo, ya que debe desinflarse aproximadamente cada cinco minutos para restaurar la circulación en la parte inferior del brazo del sujeto durante la prueba. Reemplazo del electrocardiógrafo con PTT, la prueba puede incluir más preguntas en cada sesión que una prueba de polígrafo tradicional. Más preguntas de prueba producen más mediciones de reacciones fisiológicas, y las mediciones adicionales pueden mejorar la confiabilidad y precisión de la prueba.

El protocolo HMCT introduce un nuevo formato de prueba para la fase optimizada del polígrafo de la prueba. Las preguntas de la prueba se organizan en todos los pares posibles de preguntas relevantes. En la primera mitad de la primera sesión, las preguntas sobre el primer tema relevante (R1) se emparejan con preguntas sobre el segundo tema relevante (R2). En la segunda mitad de la primera sesión, las preguntas sobre el tercer tema relevante (R3) se emparejan con preguntas sobre el cuarto tema relevante (R4). En sesiones posteriores, R1 se empareja con R3, R1 se empareja con R4, R2 se empareja con R3 y R2 se empareja con R4. Durante cualquier subsesión, el sujeto se enfoca en solo dos de los cuatro temas relevantes.

Fue óptima la presentación auditiva de las preguntas de la prueba de tipo Sí / No cada 22 segundos para registrar los cambios de diagnóstico en las medidas tradicionales del polígrafo, como la amplitud máxima de la respuesta de conductancia de la piel, los aumentos de la presión arterial y la supresión de la respiración. En las últimas sesiones, las demandas cognitivas de leer y responder con rapidez y precisión fueron óptimas para las medidas óculo motoras, como el número de fijaciones y los aumentos del diámetro de la pupila. Las características extraídas de las sesiones optimizadas de polígrafo y las sesiones optimizadas para las medidas óculo motoras se combinaron mediante una ecuación de regresión logística binaria para clasificar a los sujetos como creíbles o no creíbles en la prueba y en cada uno de los temas relevantes individualmente.

En la validación cruzada, las decisiones del protocolo HMCT fueron 86,3% correctas cuando clasificó al sujeto como veraz en todas las preguntas relevantes o engañoso en una o más de las preguntas relevantes. Las decisiones de HMCT fueron 90% correctas en temas relevantes individuales.

El resto del presente artículo describe los procedimientos utilizados para desarrollar el MCT híbrido y estimar su precisión.

Experimento

Realizamos un experimento de crimen simulado siguiendo el modelo de Cook et al. (2012) para desarrollar y validar de forma cruzada un modelo estadístico de medidas de polígrafo y óculo motoras que produjeron un puntaje de credibilidad para cada tema relevante. En el presente experimento, hubo cuatro temas: (1) robo de efectivo, (2) robo de un anillo, (3) robo de un teléfono celular y (4) robo de un conjunto de audífonos AirPods®.

Reclutamos a 182 sujetos de la comunidad local. Dos personas fueron eliminadas de la muestra debido a registros inadecuados de ECG o PPG. Se les dijo a los sujetos que algunos participantes cometerían uno o más de los robos, mientras que otros serían inocentes y no cometerían ninguno de los delitos. Los

sujetos fueron asignados aleatoriamente a uno de tres grupos. Un grupo de sujetos culpables robó $20 en efectivo de la oficina de una secretaria (n=60). Mintieron acerca de tomar los $20 en efectivo, pero dijeron la verdad sobre el robo del anillo de diamantes, el teléfono celular y los AirPods. Otro grupo de sujetos culpables tomó $20 en efectivo de una secretaria y también robó un anillo de un cajón del escritorio (n=60). Mintieron sobre dos de los cuatro temas relevantes. El tercer grupo de sujetos era inocente de los cuatro delitos simulados (n=60); estos fueron veraces acerca de los cuatro temas relevantes cubiertos en la prueba.

Una vez que los sujetos llevaron a cabo sus tareas asignadas, se les dio la prueba del protocolo HMCT. Se sentaron en un escritorio frente a una pantalla de computadora. El supervisor de pruebas adjuntó sensores de conductancia cutánea, ECG, PPG y respiración al sujeto y luego calibró el rastreador de ojos. Un sensor de fisiología multicanal registró la conductancia de la piel y la respiración a 340 Hz y ECG y PPG a 1000 Hz. Un rastreador de ojos remoto Tobii 4C conectado a la parte inferior del monitor de la computadora registró la posición de la mirada bilateral y el diámetro de la pupila a 60 Hz.

La fase optimizada del polígrafo del protocolo HMCT contenía un conjunto de 22 preguntas de tipo Sí / No, repetidas tres veces. Una voz digital de texto a voz presentó instrucciones de prueba y preguntas por auriculares. Se indicó a los sujetos que evitaran moverse, que mantuvieran los pies apoyados en el suelo y que miraran la pantalla de la computadora durante la prueba.

La Tabla 1 muestra las preguntas de la prueba para la primera de las tres sesiones de polígrafo. Tenga en cuenta que la primera subsesión preguntó sobre los temas de R1 y R2, mientras que la segunda subsesión preguntó sobre los temas de R3 y R4. Las sesiones dos y tres emparejaron los temas R1 con R3, R2 con R4, R1 con R4 y R2 con R3. Antes de cada subsesión, se informó a los sujetos sobre los dos temas relevantes que se tratarían en esa subsesión, por ejemplo, ahora se le preguntará sobre el robo del anillo de diamantes y los $20 en efectivo.

Tabla 1. Preguntas de prueba para la primera sesión de polígrafo

No. de pregunta	Sub sesión	Tipo de pregunta	Pregunta
1	1	Neutral	¿Es prudente mirar a ambos lados antes de cruzar la calle?
2	1	R2	¿Es culpable del robo del anillo de diamantes?
3	1	R1	¿Es culpable del robo de los $20 en efecitvo?
4	1	R2	¿Robó el anillo de diamantes?

5	1	Neutral	¿Se necesitan alrededor de 180 días para que la tierra gire alrededor del sol una vez?
6	1	R1	¿Robó los $20 en efectivo?
7	1	R2	¿Es inocente de robar el anillo de diamantes?
8	1	R1	¿Es inocente de robar los $20 en efectivo?
9	1	Neutral	¿Obtienen los camiones grandes un mejor rendimiento de combustible que los autos compactos más nuevos?
10	1	R2	¿Se llevó el anillo de diamantes?
11	1	R1	¿Sacó los veinte dólares de la billetera?
12	2	Neutral	¿Ha leído al menos un libro, revista o artículo de periódico en su vida?
13	2	R3	¿Es la persona que robó el teléfono celular?
14	2	R4	¿Tomó los AirPods del carro Toyota Camry?
15	2	R3	¿Es culpable de robar el celular perdido?
16	2	Neutral	¿Facebook le permite mantenerse en contacto con amigos y familiares?
17	2	R4	¿Robó los AirPods del carro Toyota Camry rojo?
18	2	R3	¿Robó el teléfono celular de la oficina?
19	2	R4	¿Es inocente del robo de los AirPods?
20	2	Neutral	¿Sirven hamburguesas y papas fritas en McDonald's, Wendy's y Burger King?
21	2	R4	¿Es la persona que robó los AirPods?
22	2	R3	¿Es inocente del robo del celular?

La fase optimizada para las medida óculo motoras del protocolo HMCT siguió a las tres sesiones de polígrafo. La computadora informó a los sujetos que debían leer y responder a los enunciados de manera rápida y precisa o podrían reprobar la prueba. Se presentaron 32 declaraciones de tipo Verdadero / Falso dos veces en diferentes órdenes. Ocho enunciados abordaron cada uno de los cuatro

temas relevantes. Además de las medidas fisiológicasy óculo motoras, la computadora también grabó los tiempos de respuesta y el número de preguntas respondidas incorrectamente. La computadora informó a los sujetos que debían leer y responder las declaraciones de manera rápida y precisa o podrían reprobar la prueba. La Tabla 2 contiene los enunciados de prueba en la primera repetición de la porción óculo motora del HMCT.

Tabla 2. Enunciados para la primera sesión óculo motora

No. de enunciado	Tipo de enunciado	Enunciado V / F
1	R3	Admito haber robado el teléfono celular perdido.
2	R1	No robé los $20 en efectivo de la secretaria.
3	R2	El anillo de diamantes no está en el cajón del escritorio por mi culpa.
4	R4	Los AirPods desaparecieron porque los tomé.
5	R3	No robé el teléfono celular de la oficina.
6	R1	Yo soy la persona que robó los $20.
7	R4	No robé los AirPods de un carro en el estacionamiento.
8	R2	El anillo de diamantes del escritorio fue robado por mí.
9	R4	Los AirPods desaparecidos fueron robados por mí.
10	R1	La razón por la que los $20 se han ido es porque los robé.
11	R3	Soy la persona que robó el celular.
12	R2	Robé el anillo de diamantes del cajón del escritorio.
13	R3	Soy inocente del robo del teléfono celular.
14	R2	No tomé el anillo de diamantes del escritorio.

15	R4	Soy inocente del robo de los AirPods.
16	R1	Admito que me llevé los $20 robados.
17	R4	Soy culpable del robo de los AirPods.
18	R2	No robé el anillo de diamantes del escritorio.
19	R3	Faltaba el teléfono celular porque lo robé.
20	R1	No soy la persona que robó los $20 en efectivo.
21	R4	No robé los AirPods del coche.
22	R1	Soy inocente de robar los $20 en efectivo.
23	R3	No robé ese teléfono celular perdido.
24	R2	Soy inocente del robo del anillo de diamantes.
25	R3	El celular perdido no fue robado por mí.
26	R2	Admito haber robado el anillo de diamantes del escritorio.
27	R4	Robé los AirPods del Toyota Camry rojo.
28	R1	No robé el dinero en efectivo de la billetera.
29	R4	No tomé los AirPods del carro Toyota Camry.
30	R1	Los $20 de la secretaria fueron robados por mí.
31	R3	El celular perdido fue robado por mí.
32	R2	No me llevé el anillo de diamantes del cajón del escritorio.

A los sujetos se les pagó $40 en efectivo por su participación y se les ofreció un bono adicional de $30 si aprobaron la prueba.

Análisis

Se analizaron datos óculo motores para identificar características que discriminaran entre preguntas y enuciados respondidos de manera veraz y engañosa. Se identificó un subconjunto de características del polígrafo y óculo motoras que lograron más del 89% de precisión en el conjunto completo de preguntas relevantes. Para cada tema, esas características fueron ponderadas y combinadas mediante una ecuación de regresión logística que generó un puntaje de credibilidad para cada pregunta relevante. El índice de credibilidad fue la probabilidad de que la persona fuera sincera sobre ese tema. Si el índice de credibilidad era 0.5 o mayor, el tema se clasificaba como veraz a las preguntas sobre ese tema. Si el índice de credibilidad era inferior a 0.5, el sujeto se clasificaba como engañoso sobre ese tema.

Validación de N-veces

Un modelo estadístico que es óptimo para clasificar los casos en un experimento en particular rara vez es óptimo para la población de la que se tomaron muestras de los sujetos. El modelo no es óptimo porque la muestra no representa perfectamente la población más general de la que se extrajo. En consecuencia, el modelo produce estimaciones de precisión sesgadas si se prueba en los casos utilizados para crear el modelo.

Se pueden obtener mejores estimaciones de precisión con la validación de “N” veces. Una validación de N veces divide el conjunto de datos en N subconjuntos. El primer subconjunto comprende una submuestra de reserva y se elimina del conjunto de datos. Los subconjuntos restantes se combinan para crear un conjunto de entrenamiento. Se desarrolla un modelo de regresión logística utilizando los casos del conjunto de entrenamiento. Luego, ese modelo de regresión se utiliza para clasificar los casos en la submuestra reservada. La precisión observada en la muestra reservada proporciona una estimación de precisión menos sesgada porque los casos reservados no se utilizaron para optimizar los coeficientes de características en la ecuación de regresión. Se registra la precisión alcanzada en la muestra retenida.

Este proceso continúa para cada partición del conjunto de datos. El primer subconjunto se devuelve al conjunto de entrenamiento y el segundo subconjunto se elimina para que sirva como una nueva muestra reservada. Se crea una nueva ecuación de regresión logística con todos menos el segundo subconjunto de casos. Ese modelo se utiliza para clasificar casos en la muestra reservada y se registra su precisión. Este proceso se repite para cada uno de los subconjuntos restantes. La mejor estimación de la precisión del modelo es la precisión media en las k muestras reservadas.

Validación del HMCT

Cada uno de los 180 sujetos fue veraz o engañoso en cada una de las cuatro preguntas relevantes. Esto proporcionó un total de 720 (180 X 4) preguntas relevantes donde la persona respondió de manera veraz o engañosa. La validación de “N” veces se completó con 240 de esas 720 preguntas relevantes. Para lograr un equilibrio adecuado entre casos veraces y engañosos y respuestas engañosas por parte de sujetos culpables que cometieron uno o dos delitos, el número de preguntas para el proceso de validación estuvo limitado por el número de sujetos culpables que cometieron un delito (n=60). Las 60 preguntas respondidas engañosamente por sujetos que robaron solo efectivo se incluyeron en la

muestra de validación. Luego, se obtuvo un número igual de preguntas engañosas de una muestra aleatoria de los 60 sujetos culpables que cometieron dos delitos. Dado que la muestra de validación ahora contenía 120 (2 x 60) preguntas relevantes respondidas de manera engañosa, se seleccionó al azar un número igual de preguntas respondidas de manera veraz de la muestra completa de 180 sujetos. Esa submuestra de 120 casos veraces contenía una mezcla aleatoria de preguntas respondidas con veracidad por sujetos que eran veraces en todas las preguntas relevantes (inocentes) o eran veraces en algunas preguntas pero no en otras (culpables).

Se realizó una validación de 4 veces. La muestra de 240 preguntas se dividió en 4 submuestras de 60 preguntas. La mitad de las preguntas en cada submuestra fueron preguntas respondidas con veracidad y la otra mitad fueron preguntas respondidas de manera engañosa. La mitad de las preguntas respondidas de manera engañosa fueron de sujetos que cometieron un delito y la otra mitad fueron de sujetos que cometieron dos delitos. La Tabla 3 muestra el porcentaje correcto de preguntas veraces y engañosas para cada vez, así como la precisión media en los cuatro validaciones.

Tabla 3. Porcentaje de decisiones correctas para preguntas respondidas de manera veraz o engañosa en una validación cuádruple

	1ra vez	2da vez	3ra vez	4ta vez	Media
n	180	180	180	180	
Veraz	93.3	88.9	90.4	94.1	91.7
Engañoso	91.1	86.7	77.8	97.8	88.3
				Precisión media	**90.0**

En promedio, la precisión fue ligeramente mayor para las preguntas respondidas verazmente (91,7%) que para las preguntas respondidas de forma engañosa (88,3%). En el nivel de las preguntas relevantes individuales, la precisión media en la validación cruzada fue del 90,0%. Con base en estos resultados, *esperaríamos que el HMCT produzca un 90% de decisiones correctas sobre cuestiones individuales relevantes cuando el modelo se utilice con una nueva muestra.*

La computadora usó los puntajes de credibilidad no sesgados para los cuatro temas relevantes de la validación cuádruple para decidir si un tema era engañoso para una o más de las preguntas relevantes. Clasificó a los sujetos que aprobaron los cuatro temas relevantes como inocentes y clasificó a los sujetos que fallaron en uno o más de los temas relevantes como culpables. La precisión media fue del 86,3%. El modelo de decisión fue un poco más preciso para los culpables (87,5%) que para los inocentes (85,0%). El modelo de decisión fue significativamente más preciso para los sujetos culpables que cometieron un delito (96,7%) que para los sujetos culpables que cometieron dos delitos (78,3%); $z = 3,05$, $p < 0,01$.

Referencias

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)*, 301-313.

Geddes L.A., Voelz, M.H., Babbs C.F., Bourland J.D., & Tacker W.A. (1981). Pulse transit time as an indicator of arterial blood pressure. *Psychophysiology, 18,* 71–74.

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1978). Pulse transit time: Relationship to blood pressure. *Behavior Research Methods & Instrumentation, 10,* 623-626.

Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1979). Pulse transit time: Relationship to blood pressure and myocardial performance. *Psychophysiology, 16,* 292-301.

Webb, A. K. & Kircher, J. C. (2005). *Use of Pulse Transit Time for the Psychophysiological Detection of Deception.* Final report to the U.S. Department of Defense. Salt Lake City: University of Utah, Department of Educational Psychology.