

Adendum a la Revisión Meta-analítica del 2011 – el Examen Utah de Cuatro Preguntas (“Técnica Raskin”) / Equipo Editorial del ESS

APA (2011) publicó el reporte de la revisión meta-analítica de técnicas poligráficas validadas, en preparación para la evolución de los estándares de práctica que requieren el uso de técnicas validadas para la práctica de campo del polígrafo. Dos aspectos importantes en el diseño de ese proyecto fueron la especificación de los requisitos de inclusión en el estudio y una definición de lo que se entiende por una técnica poligráfica. Para los fines de esa revisión, se definió a la técnica poligráfica como una secuencia definida de preguntas con un método de análisis. Esta definición se basa en el supuesto de que la efectividad de una técnica poligráfica se encuentra en función de los datos de registro para el análisis e interpretación y también en el análisis de los datos registrados a través de un proceso válido y estructurado.

La inclusión en la revisión meta analítica requirió de estudios publicados y replicados que mostraran sensibilidad, especificidad, tasas de falsos positivos y falsos negativos de la prueba, además de la publicación de las medias y la varianza de las distribuciones muestrales. El requisito de publicación y réplica se basó en la premisa de que todas las muestras de investigación están sesgadas - son una representación imperfecta de la población - y del hecho de que las estadísticas de muestreo, si se seleccionan al azar, van a converger hacia los parámetros desconocidos de la población de acuerdo con el teorema del límite central. (Kwak y Kim, 2017). Los investigadores de todas las áreas de las ciencias sociales hacen uso de este teorema para desarrollar pruebas y mediciones de fenómenos amorfos como son los rasgos de personalidad, el funcionamiento intelectual, el rendimiento académico, la altura, peso o cualquier fenómeno de interés de la población de referencia.

Aunque en la revisión meta analítica se incluyó la prueba Utah de tres preguntas, la prueba Utah de cuatro preguntas - a veces referida como "técnica Raskin"- no se incluyó debido a la ausencia de información publicada específica sobre este formato. La prueba de cuatro preguntas Utah se menciona por APA (2011) en la nota al pie de página # 50 en la página 248, debido a su similitud estructural con el AFMGQT, junto con un aviso de que se puede generalizar la información para los dos formatos mencionados.

Una diferencia importante entre el formato de cuatro preguntas de Utah y el AFMGQT es que este último normalmente se interpreta usando la regla de puntaje por subtotal (SSR), mientras que el formato de cuatro preguntas de Utah normalmente se interpreta con la regla de gran total (GTR) o con la regla de dos etapas (TSR). [Ver Nelson (2018) para una revisión de las reglas de decisión del polígrafo.] Detrás de la selección de una regla poligráfica de decisión se encuentra una suposición acerca de si las preguntas relevantes son independientes o no independientes (dependientes).

En el contexto científico, la independencia requiere que las preguntas no tengan una fuente compartida de varianza en la cual los factores que influyen en las respuestas ante cualquiera de las preguntas pueden también influir en las respuestas ante otras preguntas. El AFMGQT se usa comúnmente en contextos de polígrafos exploratorios, en los que se formulan preguntas relevantes para investigar una serie de comportamientos de preocupación, en ausencia de cualquier incidente o alegato conocido, y comúnmente se interpretan asumiendo una variación de criterio independiente (a pesar de que la atención del examinado sigue siendo un factor de dependencia o de influencia potencial dentro de un examen de asuntos múltiples). Tanto por razones psicológicas como estadísticas (es decir, de multiplicidad), los exámenes de asuntos múltiples no pueden proporcionar el mismo nivel de exactitud o precisión que los exámenes de un solo asunto. Sin embargo, los exámenes de asuntos múltiples son útiles para los programas con polígrafos exploratorios.

En contraste, el formato de cuatro preguntas Utah se utiliza como un formato poligráfico de diagnóstico para evento específico, utilizado para investigar la veracidad de las declaraciones de un examinado con respecto a un incidente conocido. Debido a que todas las preguntas relevantes involucran un solo incidente o acusación conocida, la suposición de su independencia no tiene fundamento. Un aspecto conveniente y útil de la prueba de asunto único de cuatro preguntas Utah es que permite cualquier combinación de preguntas relevantes primarias y secundarias (es decir, pregunta relevante débil, de conexión de evidencia, de culpabilidad por conocimiento, de ayuda, planeación, participación y de descripción de roles). Es decir, los examinadores de campo tienen la libertad de utilizar cualquier tipo de pregunta relevante que se adapte mejor a las circunstancias y necesidades de la investigación. En la práctica, al menos una de las preguntas relevantes es normalmente una pregunta relevante primaria que describe la participación conductual del examinado en el tema de investigación. En muchos sentidos esto es similar a la formulación de preguntas de las versiones 2 del AFMGQT. Los aspectos centrales de la secuencia [CQ, RQ, RQ, CQ, RQ, RQ, CQ] son estructuralmente idénticos en la prueba de cuatro preguntas Utah y en el AFMGQT. Otra similitud en los dos formatos es que todas las CQ y RQ se rotan de forma aleatoria o pseudoaleatoria en cada iteración de la secuencia de preguntas.

Independientemente de la combinación de preguntas relevantes primarias y secundarias, la prueba de cuatro preguntas Utah se interpreta asumiendo que entre las RQ hay una variación de criterio compartida - el examinado estuvo involucrado o no en el alegato o incidente bajo investigación. Una suposición de varianza de criterio no independiente es la base para el uso de la GTR o TSR, y simplifica las suposiciones y requerimientos tanto psicológicos como estadísticos. El examinado no está sujeto a demandas de atención divididas, porque todas las preguntas relevantes pertenecen al mismo incidente o alegato. Los efectos de multiplicidad no son un problema cuando se utiliza el GTR y se reduce mediante el uso de una corrección estadística al utilizar el TSR.

Otro aspecto importante del formato de cuatro preguntas Utah es que, al utilizar cuatro preguntas relevantes en lugar de tres o dos, el resultado de la prueba se basará en mayor cantidad de información en comparación con los formatos de prueba de tres y dos preguntas. Como principio general, el uso de una mayor cantidad de datos conduce a una mayor precisión o exactitud en las conclusiones cuantitativas. Esto está relacionado con la *ley de los grandes números* (Dekking, 2005; Révész, 1968)

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

que sostiene que la frecuencia en la ocurrencia de un evento aleatorio converge hacia su probabilidad en la medida en que aumente el número de ensayos. Ésta es la razón por la que se prefieren las muestras más grandes en comparación con las más pequeñas - si se seleccionan al azar, pueden aproximarse más a un parámetro desconocido de la población.

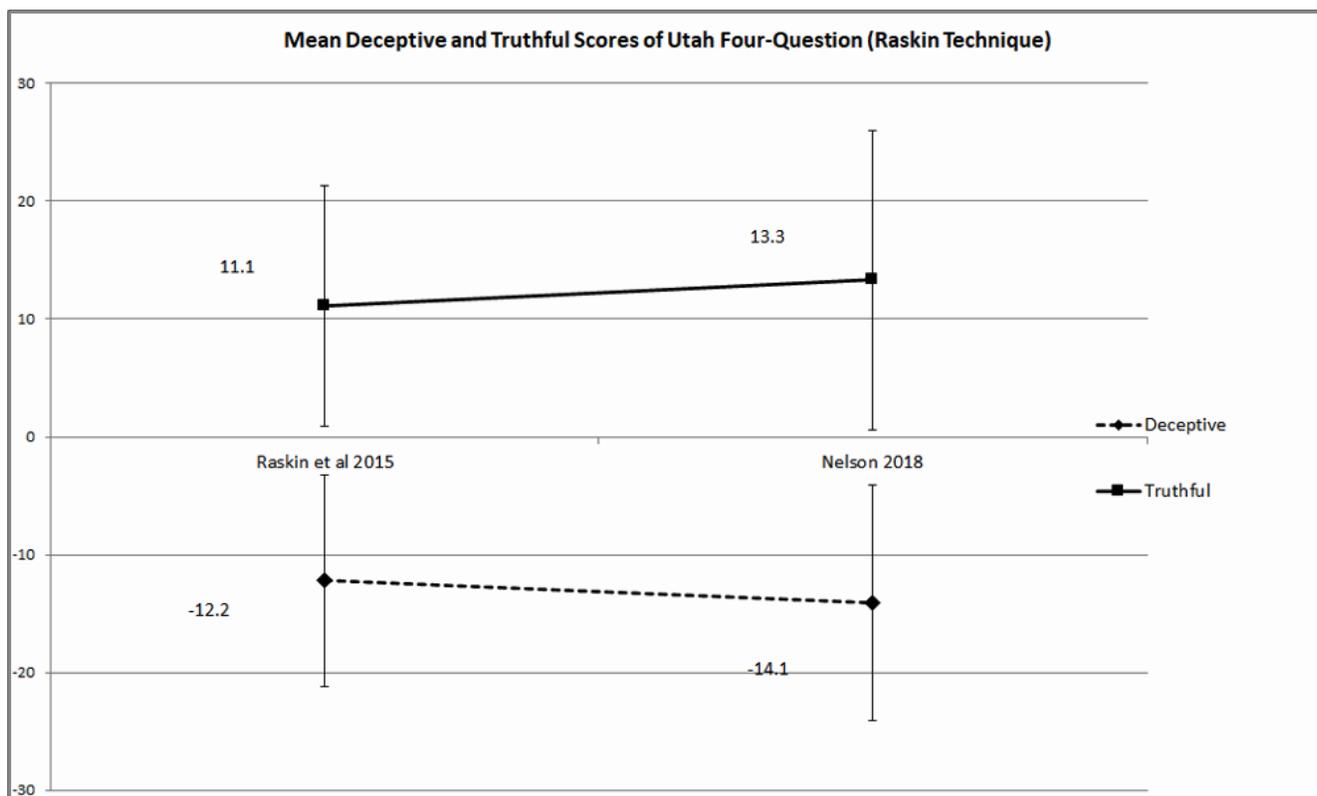
Con la prueba de cuatro preguntas Utah, la cantidad de datos de una prueba de polígrafo con cuatro RQ y hasta cinco iteraciones en la secuencia de preguntas, es de tres veces más que los de una prueba con tres iteraciones de dos RQ. Esto resulta en un aumento tanto en la sensibilidad como en la especificidad de la prueba, con su correspondiente reducción de resultados inconclusos y un aumento en la precisión general. Otro resultado es que la prueba de cuatro preguntas de Utah puede ser más sólida y menos vulnerable en un contexto con datos de prueba difíciles.

Raskin, Honts, Nelson y Handler (2015) publicaron los resultados de un estudio Monte Carlo de la prueba de cuatro preguntas Utah, que incluía puntajes tanto de ESS como de siete posiciones. Los datos iniciales Monte Carlo fueron N = 100 exámenes de la Universidad de Utah. Los resultados fueron estadísticamente indiferenciables para los dos métodos de puntuación. Para los puntajes ESS con el TSR usando un alfa = .05 / .05 para el engaño y la veracidad, la tasa de precisión no ponderada para cinco iteraciones de la secuencia de preguntas fue de .949 con una tasa no ponderada de inconclusos de .020.

Nelson (2018) publicó un segundo estudio sobre la técnica de Raskin utilizando datos de casos confirmados de archivo del DoD-PI. Los exámenes fueron una muestra de N = 30 casos de campo confirmados que se realizaron utilizando el formato AFMGQT. Este formato se describió anteriormente como estructuralmente similar a la prueba de cuatro preguntas Utah; todos los casos constaron de tres iteraciones de una secuencia de preguntas que incluía una combinación de preguntas relevantes primarias y secundarias. Los puntajes se obtuvieron usando una versión automatizada del ESS-M, y los resultados se clasificaron como de engaño o veracidad utilizando el TSR con un alfa = .05 / .05 para engaño y veracidad. La precisión no ponderada fue de .929 con una tasa de inconclusos de .033.

La Figura 1 muestra una gráfica de la media y desviación estándar de las puntuaciones de las distribuciones de muestreo de los estudios incluidos de la Técnica de Cuatro Preguntas (Raskin) de Utah. Un ANOVA de dos vías mostró que la interacción de la distribución muestral y del estado de criterio no fue estadísticamente significativa [$F(1,88) = 0.673, (p = .414)$], ni tampoco lo fue el efecto principal para la distribución muestral [$F(1, 88) = 0, (p = .993)$]. Los ANOVA unidireccionales no mostraron diferencias significativas en las puntuaciones de los dos estudios para las muestras de los engañosos [$F(1,44) = 0.018, (p = 0.895)$] o las de los veraces [$F(1,44) = 0.02, (p = 0,888)$].

Figura 1. Media y desviaciones estándar para los puntajes de muestras de veraces y engaño con el formato de cuatro preguntas Utah



La Tabla 1 muestra el resumen de las dos combinaciones en la Técnica Raskin.

Número de estudios utilizables	2
Total N	130
N Engaño	65
N Veracidad	65
Número de Examinadores/Calificadores	2
Puntajes Totales	130
Puntajes E	65
Puntajes V	65
Media E	-12.638
Des.St. E	10.154
Media V	11.608
Des.St. V	9.854

La Tabla 2 muestra el perfil y los intervalos de confianza estadísticos para las métricas de precisión de criterio.

Tabla 2. Criterio de precisión e intervalos de confianza para la técnica Raskin.	
Precisión promedio no ponderada	.944 (.021) {.897 a .984}
Inconclusos no ponderados	.031 (.026) {.010 a .092}
Sensibilidad	.923 (.033) {.852 to .984}
Especificidad	.908 (.036) {.831 to .971}
Errores FN	.046 (.026) {.010 to .104}
Errores FP	.062 (.030) {.014 to .125}
INC E	.031 (.021) {.010 to .078}
INC V	.031 (.021) {.010 to .078}
VPP	.938 (.031) {.871 to .986}
VPN	.952 (.027) {.892 to .990}
E Correctos	.952 (.027) {.893 to .990}
V Correctos	.936 (.031) {.871 to .986}
Coefficiente de Eficiencia de Detección	.875 (.041) {.788 to .949}

La tabla 3 muestra un resumen de los estudios individuales.

Tabla 3. Resumen de los estudios individuales de la Técnica Raskin.		
Estudio	Raskin et al., (2015)	Nelson (2018)
Muestra N	100	30
N Engaño	50	15
N Veracidad	50	15
Calificadores	1	1
Puntajes E	50	15
Puntajes V	50	15
Puntajes Totales	30	30
Media E	-12.2	-14.1
Des.St. E	10.2	10.0
Media V	11.1	13.3
Des.St. V	9.0	12.7
Precisión Promedio no Ponderada	.949	.929
Inconclusos no Ponderados	.020	.067
Sensibilidad	.940	.870
Especificidad	.920	.870
Errores FN	.040	.067
Errores FP	.060	.067
INC E	.020	.067
INC V	.020	.067
VPP	.940	.929
VPN	.958	.929
E Correctos	.959	.929
V Correctos	.939	.929

El nivel combinado de precisión en la decisión de los estudios de la prueba de cuatro preguntas Utah (“técnica de Raskin”), ponderado por el tamaño de la muestra y por el número de calificadores, fue de .944 con una tasa combinada de inconclusos de .031. El coeficiente de eficiencia en la detección, calculado como la correlación entre el resultado categórico codificado como [+1, 0, -1] y el estado de criterio para cada caso codificado como [+1, -1] fue de .875 con un intervalo de confianza del 95% desde .788 a .949.

Referencias

- American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. [Electronic version] Retrieved September 15, 2020, from <http://www.polygraph.org>.
- Dekking, Michel (2005). *A Modern Introduction to Probability and Statistics*. Springer.
- Nelson, R. (2018). Credibility assessment using Bayesian credible intervals: a replication study of criterion accuracy using the ESS-M and event-specific polygraphs with four relevant questions. *Polygraph and Forensic Credibility Assessment*, 47 (1), 85-90.
- Kwak, S. G. & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. Korean *Journal of Anesthesiology*. 70(2), 144–156.
- Nelson, R. (2018b). Practical polygraph: a survey and description of decision rules. *APA Magazine*, 51(2), 127-133.
- Raskin, D. C., Honts, C., Nelson, R. & Handler, M. (2015). Monte Carlo estimates of the validity of four-relevant-question polygraph examinations. *Polygraph* 44(1), 1-27.
- Révész, P. (1968). *The Laws of Large Numbers*. Academic Press

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor. rodolfo@poligrafia.com.mx
2020, 49 (2)

Polygraph & Forensic Credibility Assessment ,