

## **Uso del Sistema de Calificación Empírica**

**Raymond Nelson, Mark Handler, Pam Shaw,  
Michael Gougler, Benjamin Blalock,  
Chad Russell, Barry Cushman, y Marty Oelrich \***

### ABSTRACT

Los autores describen los procedimientos del Sistema de Calificación Empírica, un sistema normativo para el análisis de datos de prueba basado-en-evidencia, de datos de exámenes de la detección psicofisiológica del engaño, para formatos de prueba con preguntas de comparación. Se proporciona una descripción del procedimiento de todos los aspectos del modelo de calificación, incluidas las características fisiológicas, transformaciones matemáticas, reglas de decisión, y puntajes de corte basados en datos normativos. Se muestran los datos de validación para 13 dimensiones de precisión de criterio para 5,192 resultados de calificaciones de 732 exámenes confirmados ZCT Federales y Utah de tres-preguntas. Estos datos provienen de seis muestras, calificadas por 140 calificadores con y sin experiencia en 16 cohortes de participantes del estudio. En los apéndices se proporcionan los datos normativos.

---

\*Favor de contactar al autor principal en [raymond.nelson@gmail.com](mailto:raymond.nelson@gmail.com) por cualquier pregunta o comentario. Los autores permiten el uso ilimitado y derechos de duplicación a cualquier escuela acreditada por la American Polygraph Association (Asociación Americana de Poligrafía) o reconocida por la American Association of Police Polygraphists (Asociación Americana de Policías Poligrafistas) o la National Polygraph Association (Asociación Nacional de Poligrafía)

## Uso del Sistema de Calificación Empírica

Raymond Nelson, Mark Handler, Pam Shaw,

Michael Gougler, Benjamin Blalock,

Chad Russell, Barry Cushman, y Marty Oelrich

La escena es la sala de un tribunal y usted es el examinador que responde preguntas como testigo experto.

ABOGADO OPOSITOR: ¿Cuál es el nivel de significancia estadística o la probabilidad de error de su resultado de la examinación poligráfica calificada manualmente?

EXAMINADOR: Perdone, ¿puede repetir la pregunta? No la entiendo muy bien...

ABOGADO OPOSITOR: ¿Cuál es el valor-p o probabilidad de error asociado con su resultado de prueba calificado numéricamente?

EXAMINADOR: ¿valor-p? No lo sé ...

(Continuará ...)

## Introducción

El sistema de puntuación empírica (ESS) es un modelo de calificación numérica basado-en-evidencia para el análisis manual de datos de prueba (TDA), de datos de exámenes de detección psicofisiológica del engaño (PDD) en exámenes que se conducen utilizando formatos de examen con preguntas de comparación (CQT). El ESS está diseñado para proporcionar al examinador de campo un modelo de calificación validado y confiable basado en las soluciones más simples disponibles, para las que hay evidencia publicada de validez que soporta todos sus principios, supuestos y procedimientos. La simplicidad del ESS es un medio para que las habilidades de calificación se adquieran fácilmente por los calificadores inexpertos, se retengan fácilmente por calificadores con y sin experiencia,

y sean fáciles de entender cuando se explican a profesionales no-poligrafistas.

El ESS incluye una descripción de las características de los datos fisiológicos que están correlacionados con la veracidad y el engaño, de los métodos de transformación matemática para asignar y agregar las puntuaciones numéricas, de las reglas de decisión para la clasificación de las puntuaciones numéricas como indicativas de veracidad o engaño, y de los puntajes de corte numéricos que definen los umbrales a priori de significancia estadística<sup>1</sup>. Los puntos de corte del ESS se basan en datos normativos que permiten el cálculo de la probabilidad de un resultado de prueba erróneo. En consecuencia, el ESS permite la selección de puntos de corte estadísticamente óptimos basados en las necesidades operativas para la resolución y la precisión del resultado de prueba.

La ESS se puede utilizar con la variedad de los formatos CQT que se ajustan a los principios válidos para la construcción de exámenes de evento-específico de asunto-único y de facetas-múltiples, o de exámenes de asuntos-múltiples exploratorios. Las técnicas de polígrafo aplicables incluyen formatos de prueba diagnósticas que utilizan formatos de evento-específico de tres-preguntas y de dos-preguntas de la familia de las Técnicas de Zonas de Comparación (ZCT), formatos de prueba de investigación multi-facetas basados en la familia de las Técnicas de Preguntas Generales Modificada (MGQT), y formatos exploratorios de asunto-múltiple que se basan en la familia de las MGQTs o las Técnicas Exploratorias de Mentiras Dirigidas (DLST) (originalmente desarrolladas como Test de Espionaje y Sabotaje [TES]). El ESS se puede utilizar con preguntas comparativas de mentira probable o de mentira dirigida.

### Procedimientos ESS

El TDA comienza con un análisis global de los datos con el fin de garantizar una cantidad suficiente de datos que sean de calidad satisfactoriamente interpretable. El análisis global implica la inspección de artefactos o eventos en los datos que podrían afectar la calidad interpretable de los datos fisiológicos en cualquiera de los canales

---

<sup>1</sup> Para los puntajes ESS, la significancia se refiere al concepto de significancia estadística, o a la probabilidad de que un resultado numérico de prueba ocurrió debido al error o solamente a la probabilidad del azar.

grabados. Estos artefactos y eventos no interpretables podrían incluir falta de atención y falta de respuesta a los estímulos de prueba, indicadores de falta de cooperación en el comportamiento, falta de puntualidad en las reacciones, movimientos voluntarios o involuntarios, y respuestas fisiológicas a estímulos que son ajenos al examen. En el análisis global también se debe observar el segmento pre-estímulo de todas las preguntas de prueba. No deben ser calificadas las reacciones que ocurren en respuesta a las preguntas presentadas durante segmentos con artefacto o inestables.

### Características

El ESS no requiere mediciones mecánicas y sólo utiliza aquellas señales fisiológicas que históricamente han sido consideradas primarias. El ESS no incluye señales fisiológicas que se consideran secundarias. Cuando se utiliza el ESS, los datos se califican en pantalla a través del análisis visual, sin el uso de impresión o mediciones mecánicas. Con el ESS no se utiliza un período de medición fija. En su lugar, las respuestas se califican hasta el final de la respuesta, siempre que se pueda argumentar que el inicio de la respuesta se atribuye al estímulo de prueba, según se determine por la calidad global y la estabilidad de los datos y la oportunidad del inicio de la respuesta. Las respuestas no se califican cuando hay una causa alternativa aparente, diferente al estímulo de prueba. Las respuestas que comienzan antes de la aparición de los estímulos y las respuestas que comienzan varios segundos después de la respuesta no pueden atribuirse de manera confiable a los estímulos de prueba y por lo tanto no se califican.

La actividad electrodérmica (EDA) se observa por el incremento vertical de la amplitud desde el punto más bajo después de la aparición de estímulo hasta el final de la reacción. La actividad cardiovascular (CV) se observa por el aumento vertical de la amplitud en la línea base diastólica desde el punto más bajo después de la aparición de estímulo hasta el final de la reacción.

El ESS se basa en el reconocimiento de patrones visuales en lugar de mediciones físicas o lineales y emplea un conjunto simplificado de patrones de reacción respiratorios, que incluyen la reducción de la amplitud respiratoria, la desaceleración de la tasa respiratoria, y el aumento temporal de la línea base respiratoria durante tres o más ciclos.

La apnea, que está fuertemente correlacionada con el criterio de veracidad o de engaño, y que se ha descrito como la forma extrema de supresión respiratoria, es fácil de fingir y sólo debe ser calificada cuando se produce en las preguntas relevantes.

#### Transformaciones.

La ESS emplea una escala de tres posiciones (+, 0, -), y las calificaciones se asignan utilizando la regla de más-grande-es-mejor. Las puntuaciones se asignan cada vez que exista una diferencia visiblemente perceptible en la fuerza de reacción entre las preguntas relevantes y las de comparación. Este proceso se basa en el principio simple y robusto de que las reacciones fisiológicas de mayor magnitud son causadas por estímulos que son más salientes para el examinado debido a factores emocionales, cognitivos, o condicionados conductualmente. Se asignan puntuaciones positivas (+) cuando hay una mayor respuesta ante los estímulos comparativos. Puntuaciones negativas (-) cuando hay una mayor respuesta a los estímulos relevantes. Las preguntas relevantes normalmente deben compararse contra las preguntas de comparación que pida la técnica utilizada.

El ESS está diseñado para colocarle el 50% del peso de diagnóstico al EDA al duplicar todas las puntuaciones del EDA. A todas las reacciones del EDA se les asigna una puntuación de +2, 0 o -2, y a todas las reacciones pneumográficas y cardiográficas se les asignan valores de +1, 0 y -1, sin importar la diferencia del tamaño entre las reacciones. Se asigna un solo puntaje para los sensores pneumográficos superior e inferior combinados. Este esquema de ponderación genera una reducción de resultados inconclusos, un aumento en la sensibilidad de la prueba para el engaño, y ningún cambio en la especificidad de la prueba para la veracidad.

#### Reglas de Decisión.

Se utilizan conjuntamente las reglas de decisión y los puntajes de corte para interpretar las calificaciones numéricas dentro de resultados de prueba por categorías. Los puntos de corte se basan en datos normativos, y son seleccionados por su nivel deseado de significancia estadística y de tolerancia para el error ante resultados de engaño y veracidad. Las reglas de decisión son los pasos procesales para interpretar las puntuaciones numéricas transformadas y agregadas, dentro de resultados categóricos (es decir, veracidad o engaño) y están basadas en las tolerancias

establecidas para los diferentes tipos de error. Con el ESS se pueden utilizar las diferentes reglas de decisión y puntajes de corte para lograr diferentes objetivos operativos tales como sensibilidad de prueba, sensibilidad de prueba y especificidad de prueba balanceada, o la precisión global de la decisión.

Regla del Gran Total. La Regla del Gran Total es la más simple de todas las reglas de decisión y a menudo proporciona el más alto nivel de precisión en la decisión a costa de una tasa de inconclusos ligeramente alta. La Regla del Gran Total involucra la sumatoria de todas las puntuaciones en un resultado numérico único. No se utilizan los puntajes Subtotales con la Regla de Gran Total.

1. Si el gran total es menor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño<sup>2</sup>, entonces la interpretación correcta es de Engaño Indicado (Deception Indicated, DI).
2. Si el gran total es mayor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de veracidad, entonces la interpretación correcta es que No se Indica Engaño (No Deception Indicated, NDI).
3. Si el gran total no es menor que o igual al puntaje de corte requerido para un resultado de engaño, y no es mayor que o igual al puntaje de corte requerido para un resultado de veracidad, entonces la interpretación correcta es Inconcluso (INC) o No Opinión (NO).

Regla de Calificación-por-Punto. La Regla de Calificación-por-Punto se utiliza para interpretar los resultados de exámenes exploratorios de asuntos-múltiples y de exámenes de investigación de facetas-múltiples llevados a cabo con los formatos de la familia de los MGQTs o del DLST. Esta regla involucra el cálculo de las puntuaciones subtotales para cada pregunta relevante. La Regla de Calificación-por-Punto puede imponerse en el caso de las pruebas de asunto-único, cuando la Regla de Gran Total no conduce a un resultado definitivo, o como una regla única para las pruebas multi-faceta o de asunto-múltiple.

---

<sup>2</sup> Se han calculado las tasas de error para un rango de puntajes de corte, y se pueden encontrar en los apéndices al final de este artículo.

1. Si todos los puntajes subtotales son mayores que o iguales al puntaje de corte requerido para un resultado estadísticamente significativo de veracidad, entonces la interpretación correcta es que No hay Respuesta Significativa (No Significant Response, NSR) o NDI para la prueba como un todo. Las puntuaciones subtotales individuales que son mayores que o iguales al puntaje de corte requerido para un resultado estadísticamente significativo de veracidad podrían ser interpretadas y reportadas como NSR (NDI) para la puntuación subtotal y a la pregunta de prueba correspondiente, si es que se requiere por parte del agente o agencia referente.

2. Si cualquier puntuación subtotal es menor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño, entonces la interpretación correcta es Respuesta Significativa (Significant Response, SR) o DI para la prueba como un todo, así como para el subtotal.

3. Cuando la prueba en su conjunto no puede ser interpretada como SR (DI) o NSR (NDI), el resultado general de la prueba deberá ser interpretado como INC o NO. No se interpretan los resultados de cualquier subtotal como NSR (NDI) si es que cualquier subtotal ha sido interpretado como SR (DI). Todos los subtotales restantes que no son SR (DI) se deberán clasificar como INC o NO y la prueba como un todo deberá ser interpretada como SR (DI).

Reglas de Dos-Etapas. Las reglas de Dos-Etapas funcionan por el uso secuencial de la Regla de Gran Total y luego por las Reglas de Calificación-por-Punto. No se permite que las calificaciones subtotales reemplacen la calificación de gran total. Las clasificaciones basadas en calificaciones por gran total, siempre triunfarán sobre del resultado de calificaciones subtotales. Se ha demostrado que las Reglas de Dos-Etapas reducen la aparición de resultados inconclusos y que incrementan la sensibilidad al engaño. Los resultados de prueba deben ser interpretados y reportados a un nivel de la prueba-como-un-todo. Sin embargo, también podrían ser reportados los resultados calificados e interpretados mediante preguntas individuales si lo requiere el agente o agencia referente.

Primera Etapa: Regla de Gran Total. Se suman todas las puntuaciones de todas las preguntas para calcular una puntuación de gran total, representativa de la prueba en su conjunto. Durante la primera etapa no

se utilizan los puntajes subtotales. No proceda a la segunda etapa, si el gran total resulta en un DI o NDI.

1. Si el gran total es mayor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de veracidad, entonces la interpretación correcta es NDI.
2. Si el gran total es menor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño, entonces la interpretación correcta es DI.
3. Proceda a la segunda etapa sólo si el gran total resulta en INC o NO.

Segunda Etapa: Regla de Calificación-por-Punto. En la etapa dos, no se utilizan las calificaciones de gran total. Dentro de un mismo examen, no interprete los resultados de algunos subtotales como SR (DI) y otros subtotales como NSR (NDI).

1. Si la puntuación de cualquier subtotal es menor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño, entonces la interpretación correcta es DI para la prueba como un todo. Cualquier puntaje subtotal que sea menor o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño deberá interpretarse y reportarse como SR (DI) para la puntuación subtotal y para su pregunta de prueba correspondiente.
2. Si ninguna de las puntuaciones subtotales es menor que o igual al puntaje de corte requerido para un resultado estadísticamente significativo de engaño, entonces la interpretación correcta es INC o NO.
3. En la segunda fase, no hay solución para obtener un resultado NDI.

Puntajes de corte y datos normativos.

Los puntajes de corte del ESS se determinan mediante el uso de datos normativos, que fueron desarrollados para diferentes tipos de formatos de examinación CQT. Los puntajes de corte son seleccionados por su nivel de significancia estadística de acuerdo a los requerimientos de resolución o precisión de la prueba.

Los formatos ZCT de tres-preguntas y de dos-preguntas hacen uso de la Regla de Gran Total y también podrían emplear la Regla de Calificación-por-Punto en la segunda etapa. Se aplica la corrección de Bonferonni al

nivel deseado de significancia estadística cuando se utiliza la Regla de Calificación-por-Punto con los exámenes ZCT de evento-específico<sup>3</sup>.

Más del 50% de los exámenes ZCT veraces van a tener al menos un puntaje subtotal que será cero o inferior. El noventa por ciento de todos los exámenes va a producir una diferencia máxima entre los subtotales de siete puntos o menos, y el 10% de los exámenes tendrá una diferencia de ocho puntos o más. El efecto de estos aspectos sobre la precisión de la decisión es estadísticamente insignificante.

Los exámenes de investigación de faceta-múltiple, que utilizan formatos MGQT, también pueden hacer uso tanto de la Regla de Calificación-por-Punto como de la Regla de Gran Total para mejorar la precisión de criterio. Los exámenes exploratorios de asunto-múltiple solamente utilizan la Regla de Calificación-por-Punto, y no hacen uso de la Regla de Gran Total. La corrección de Bonferonni no se aplica con los puntajes de corte de los resultados de engaño de los exámenes exploratorios. Sin embargo, entre los casos veraces, se utiliza una inversa de la corrección Šidák<sup>4</sup> para

---

3 Bonferonni es el nombre de un estadista que mostró un procedimiento simple para corregir el incremento potencial de errores falso-positivo o error tipo-1 que ocurren como resultado de un fenómeno conocido como alfa inflado, que ocurre en contextos de prueba para los que la probabilidad de error estadístico se calcula repetidamente para un asunto único de preocupación, tal como cuando se va a decidir un resultado de prueba poligráfica al nivel de la prueba como un todo mediante la calificación de preguntas individuales. Por ejemplo, si nosotros establecemos nuestro nivel alfa a .05, entonces tenemos una probabilidad de error menor al 5% en cada decisión. Si tomamos varias decisiones con este nivel de tolerancia al error, la tasa de error acumulada estará compuesta por el número de decisiones (por ejemplo,  $.05 * 3$  preguntas relevantes = .15). Esto podría resultar en una tasa de error tan alta como del 15% cuando nosotros iniciamos con una declaración de tolerancia al error no mayor al 5%. La corrección Bonferonni involucra la selección de un puntaje de corte que corresponda al nivel alfa o el nivel de tolerancia al error, dividido por el número de decisiones que se tienen que realizar (por ejemplo,  $.05 / 3$  preguntas relevantes = alfa con corrección Bonferonni = .0167). La corrección Bonferonni normalmente no se usa en contextos exploratorios.

4 Šidák es el nombre de un estadista que desarrolló una corrección matemática para las distorsiones de un nivel alfa deseado; es similar a la corrección de Bonferonni a excepción de que en la corrección de Šidák se asume que los asuntos son independientes y se están evaluando simultáneamente dentro de un evento de evaluación único. La inversa de esta corrección se usa para corregir el alfa y los puntajes de corte para las clasificaciones de veracidad en exámenes de faceta-múltiple y de asunto-múltiple, cuando se asume que las preguntas son independientes y serán evaluadas utilizando las reglas de Calificación-por-Punto, que sin esta corrección, podrían resultar en una deflación del alfa y en un incremento de resultados inconclusos.

reducir los resultados inconclusos, al tiempo que se mantiene el nivel requerido de significancia estadística.

Los cuadros 1 y 2, muestran los puntos de corte para exámenes de tres-preguntas y de dos-preguntas para formatos de examen de la familia ZCT. La Tabla 3 muestra los puntajes de corte para exámenes se asunto-múltiple y faceta-múltiple de los formatos de examen de la familia MGQT y DLST.

Tabla 1. Puntajes de corte para exámenes ZCT de tres-preguntas.

	Puntaje de Corte	Alfa
NDI Gran Total	+2	< .10
NDI Gran Total (conservador)	+5	< .05
DI Gran Total	-4	< .05
DI Subtotal (cuando se usa)	-7 (cualquier pregunta)	< .017 (Corrección Bonferonni)
NDI Subtotal (cuando se usa)	ninguno	ninguno

Tabla 2. Puntajes de Corte para exámenes ZCT de dos-preguntas.

	Puntaje de Corte	Alfa
NDI Gran Total	+2	< .10
NDI Gran Total (conservador)	+4	< .05
DI Gran Total	-4	< .05
DI Subtotal (cuando se usa)	-6	< .025 (Corrección Bonferonni)
NDI Subtotal (cuando se usa)	ninguno	ninguno

Tabla 3. Puntajes de corte para resultados veraces en exámenes de asunto-múltiple y faceta-múltiple con dos, tres o cuatro preguntas relevantes.

	Cutscore	Alpha
NDI Gran Total	ninguno	ninguno
DI Gran Total	ninguno	ninguno
DI Subtotal	-3 (cualquier pregunta)	< .05
NDI Subtotal	+1 (todas las preguntas)	< .10 (1- Corrección Šidák)
NDI Subtotal (conservador)	+2 (todas las preguntas)	< .05 (1- Corrección Šidák)

Los apéndices A y B muestran los datos normativos y tablas de búsqueda para las puntuaciones ESS con exámenes ZCT de tres-preguntas y de dos-preguntas. El apéndice C muestra los datos normativos y los datos de búsqueda para exámenes de faceta-múltiple y asuntos-múltiples para dos, tres o cuatro objetivos de investigación. Los datos normativos están basados en un mínimo de tres presentaciones de cada pregunta estímulo de prueba para las que al menos dos de las tres puntuaciones de los componentes, de al menos dos de las tres presentaciones debieron ser interpretables, a menos que la técnica en particular incluya requisitos más estrictos. Las diferencias en las distribuciones normativas son estadísticamente insignificantes con presentaciones adicionales, por encima del mínimo de tres presentaciones de la secuencia de preguntas de

prueba. Por lo tanto, no es necesario cambiar los puntajes de corte normativos cuando se califica una cuarta o quinta presentación de los estímulos de prueba.

#### Validación de Datos

La Tabla 4 muestra el perfil de la validez de criterio para las puntuaciones ESS y cuatro reglas de decisión diferentes para exámenes ZCT de tres-preguntas. Los datos son los puntajes ( $N = 5192$ ) de 732 casos confirmados, que incluyeron formatos ZCT Federal y Utah ZCT recolectados de 16 cohortes de calificadoros con y sin experiencia, para un total de 140 participantes, que incluyeron examinadores de campo, aprendices, aprendices internacionales y psicólogos investigadores no-poligrafistas quienes calificaron 6 muestras diferentes de casos confirmados.

Tabla 4. Perfil de precisión de criterio para exámenes ZCT de tres-preguntas.

Reglas de Decisión para Media de exámenes ZCT, (Desviación Estándar) [95% CI]				
	Reglas Tradicionales ZCT .10/.05 (+2tot/-4tot/- 7sub)	Reglas Dos- Etapas .10/.05 (+2tot/-4tot/- 7sub)	Reglas Dos- Etapas .05/.05 (+5tot/-4tot/- 7sub)	Regla Gran Total .05/.05 (+5tot/-4tot)
Correcto	.930 (.032) [.867 to .993]	.922 (.028) [.867 to .977]	.935 (.027) [.882 to .987]	.947 (.025) [.899 to .996]
INC	.382 (.048) [.288 to .477]	.098 (.030) [.04 to .156]	.168 (.037) [.095 to .242]	.194 (.039) [.117 to .272]
D INC	.171 (.053) [.067 to .276]	.106 (.044) [.02 to .192]	.143 (.050) [.046 to .24]	.171 (.054) [.066 to .276]
T INC	.594 (.070) [.457 to .731]	.089 (.041) [.01 to .169]	.194 (.056) [.084 to .304]	.218 (.058) [.104 to .333]
Sensibilidad	.822 (.054) [.715 to .928]	.817 (.055) [.709 to .924]	.820 (.054) [.713 to .927]	.792 (.058) [.679 to .905]
Especificidad	.326 (.067) [.195 to .457]	.846 (.051) [.747 to .946]	.734 (.063) [.612 to .857]	.734 (.063) [.612 to .857]
FN	.007 (.012) [<.001 to .031]	.077 (.038) [.003 to .152]	.037 (.027) [<.001 to .089]	.037 (.027) [<.001 to .089]
FP	.08 (.038) [.005 to .154]	.064 (.034) [<.001 to .132]	.072 (.036) [.001 to .143]	.048 (.030) [<.001 to .107]
PPV	.912 (.042) [.830 to .993]	.927 (.039) [.852 to >.999]	.92 (.040) [.841 to .998]	.943 (.035) [.874 to 1.013]
NPV	.978 (.037) [.906 to >.999]	.916 (.041) [.835 to .997]	.952 (.035) [.884 to >.999]	.952 (.035) [.884 to >.999]
D Correcto	.991 (.015) [.962 to >.999]	.913 (.042) [.831 to .996]	.957 (.031) [.896 to >.999]	.955 (.032) [.892 to >.999]
T Correcto	.803 (.089) [.629 to .978]	.930 (.038) [.856 to >.999]	.911 (.045) [.823 to .999]	.939 (.038) [.864 to >.999]
Promedio Ponderado	.897 (.045) [.809 to .986]	.921 (.028) [.866 to .977]	.934 (.027) [.881 to .987]	.947 (.025) [.898 to .996]

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

El ESS logra un alto nivel de precisión de criterio, con tasas de precisión de criterio medias por encima del 90% e inconclusos promedio de aproximadamente 10%<sup>5</sup>. Se puede observar el nivel más alto de exactitud con la Regla de Gran Total, con una tasa de inconclusos que es alto, aunque todavía por debajo del 20%. Sin embargo, las reglas de Dos-Etapas proporcionarán una solución óptima para la mayoría de los propósitos diagnósticos y evidenciarios. Otras reglas de decisión van a producir perfiles dimensionales de precisión de criterio ligeramente diferentes. En comparación con las otras reglas de decisión, la regla tradicional ZCT mantiene un alto nivel de precisión en la decisión, un alto valor predictivo negativo, y error bajo en falsos negativos, aunque a un costo de inconclusos alto y débil especificidad a la veracidad.

## Conclusiones

El ESS es un modelo de TDA simple y altamente eficaz basado-en-evidencia. Está fundamentado en décadas de investigación acerca de los principios que se aplican en el análisis de datos de pruebas tanto manual como automatizado para los exámenes CQT. Las ventajas de un modelo de calificación simple basado-en-evidencia son muchas, y se puede esperar que incluyan: el incremento del acuerdo inter-calificadores en comparación con modelos más complejos, la adquisición acelerada de habilidades entre calificadores inexpertos, el incremento en la retención de habilidades entre calificadores con y sin experiencia, y una mayor generalización de los resultados experimentales en ambientes de campo. Entre los beneficios más importantes del ESS es que se basa en datos normativos. Los modelos de decisión PDD basados en datos normativos permitirán a los examinadores de campo hacer cálculos inferenciales de la probabilidad de un resultado erróneo de prueba. La habilidad para calcular y discutir un resultado de prueba PDD en términos de un nivel de significancia estadística o valor-p (es decir, valor de probabilidad o probabilidad de error), opuesto a las categorías de aprobado/reprobado, significará que los consumidores y los agentes referentes serán capaces de hacer un mejor

---

<sup>5</sup> Las reglas de decisión, junto con la elección de puntajes de corte, pueden afectar el balance entre sensibilidad de la prueba y especificidad de la prueba. Se desea una sensibilidad y especificidad balanceada en las aplicaciones evidenciarias donde los falsos positivos tienen un costo más elevado en comparación con aplicaciones de investigación. En entornos de investigación se podría requerir el evitar falsos-negativos, que podrían afectar el balance de los errores observados. Lo que es óptimo en un contexto no es necesariamente óptimo en ambos ambientes.

uso de los resultados del examen PDD como herramienta científica de soporte-de-decisión que puede agregar validez incremental a los juicios profesionales relacionados con asuntos como las prioridades de una investigación, los objetivos de una entrevista, la evaluación de riesgos y el manejo de riesgos. El ESS presenta una solución efectiva, expedita y comprensible basada-en-evidencia para los examinadores de campo, investigadores, evaluadores de riesgos, administradores de riesgos, investigadores judiciales de hechos y administradores de programas PDD.

La escena es una sala de tribunal y usted es el examinador respondiendo preguntas como testigo experto.

(Toma dos)

ABOGADO OPOSITOR: ¿Cuál es el nivel de significancia estadística o probabilidad de error para su resultado del examen poligráfico calificado manualmente?

EXAMINADOR: Basado en los datos normativos y en el Sistema de Puntuación empírica, un modelo de calificación numérica basado-en-evidencia para el análisis de datos de prueba de la detección psicofisiológica del engaño, el nivel de significancia estadística o probabilidad de error (p-valor) para este resultado de prueba se calcula en menos del 0.001. En otras palabras, la probabilidad de que este resultado de prueba fue producido por una persona engañosa (persona veraz) es menor a 1 en 1000.

ABOGADO OPOSITOR: No tengo más preguntas su señoría!

## References

- Backster, C. (1963). Do the charts speak for themselves? New standards in polygraph chart interpretation. *Law and Order*, 11, 67-68, 71.
- Barland, G. H. (1985). A method of estimating the accuracy of individual control question polygraph tests. In, *Anti-terrorism; forensic science; psychology in police investigations: Proceedings of IDENTA-'85* (pp. 142-147). The International Congress on Techniques for Criminal Identification.
- Blalock, B., Cushman, B. and Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph*, 38, 281-288.
- Department of Defense Polygraph Institute (2006). *Test Data Analysis: DoDPI numerical evaluation scoring system*. Retrieved from <http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf> on 3-31-2007.
- Handler, M., Nelson, R., Goodson, W. & Hicks, C. (2011). Empirical Scoring System: A Cross-cultural Replication and Extension Study of Manual Scoring and Decision Policies., *Polygraph*, 39, 200-215.
- Handler, M., Nelson, R., Krapohl, D., and Honts, C. (2010). An EDA Primer for Polygraph Examiners. *Polygraph*, 32, 68-108.
- Harris, J., Horner, A. and McQuarrie, D. (2000). An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations. Johns Hopkins University, Applied Physics Laboratory. SSD-PORPOR- 00-7272.
- Kircher, J. C., Kristjansson, S. D., Gardner, M. K. & Webb, A. (2005). Human and computer decision-making in the psychophysiological detection of deception. University of Utah. Final report.
- Kircher, J. C., and Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291-302.

- Krapohl, D. J. and McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.
- Nelson, R., Krapohl, D. J. and Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37, 185-215.
- Raskin, D. C., Kircher, J. C., Honts, C. R. and Horowitz, S. W. (1988). A study of the validity of polygraph examinations in criminal investigations. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542-547.
- Senter, S. M. (2003). Modified general question test decision rule exploration. *Polygraph*, 32, 251-263.
- Senter, S. M. and Dollins, A. B. (2008). Exploration of a two-stage approach. *Polygraph*, 37, 149-164.
- Van Herk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. *The Newsletter of the Canadian Association of Police Polygraphists*, 7, 28-47. Reprinted in *Polygraph*, 20, 70-79.

## Apéndice A

Tabla de búsqueda de Data Normativa para Exámenes ZCT de Tres-Preguntas

Puntuación Media de engaño = -9 (DS = 8)

Puntuación Media de veracidad = 8 (DS = 7)

Puntaje-de-Corte de Veracidad (NSR) en ZCT de Tres-Preguntas	
Puntaje-de-Corte Total NDI/NSR	Valor-p (alfa)
-1	0.159
0	0.130
1	0.106
2	0.085
3	0.067
4	0.052
5	0.040
6	0.030
7	0.023
8	0.017
9	0.012
10	0.008
11	0.006
12	0.004
13	0.003
14	0.002
15	0.001
Puntaje-de-Corte de Engaño (SR) en ZCT de Tres-Preguntas	
Puntaje-de-Corte Total DI/SR	Valor-p (alfa)
1	0.159
0	0.127

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

-1	0.099
-2	0.077
-3	0.058
-4	0.043
-5	0.032
-6	0.023
-7	0.016
-8	0.011
-9	0.008
-10	0.005
-11	0.003
-12	0.002
-13	0.001

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

## Apéndice B

Tabla de búsqueda de Data Normativa para Exámenes ZCT de Dos-Preguntas

Puntuación Media de engaño = -6 (DS = 6)

Puntuación Media de veracidad = 6 (DS = 6)

Puntaje-de-Corte de Veracidad (NSR) en ZCT de Dos-Preguntas	
Puntaje-de-Corte Total NDI/NSR	valor-p (alfa)
-1	0.202
0	0.159
1	0.122
2	0.091
3	0.067
4	0.048
5	0.033
6	0.023
7	0.015
8	0.010
9	0.006
10	0.004
11	0.002
12	0.001
Puntaje-de-Corte de Engaño (SR) en ZCT de Dos-Preguntas	

Puntaje-de-Corte Total DI/SR	valor-p (alfa)
1	0.202
0	0.159
-1	0.122
-2	0.091
-3	0.067
-4	0.048
-5	0.033
-6	0.023
-7	0.015
-8	0.010
-9	0.006
-10	0.004
-11	0.002
-12	0.001

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

## Apéndice C

### Tabla de búsqueda de Data Normativa para Exámenes de Asunto-Múltiple y Faceta-Múltiple

Puntuación Media de engaño = -2 (DS = 3)

Puntuación Media de veracidad = 2 (DS = 3)

Puntaje-de-Corte de Engaño (SR) en Asunto-Múltiple y Faceta -Múltiple		
Puntaje-de-Corte SR	valor-p (alfa)	
0	0.252	
-1	0.159	
-2	0.091	
-3	0.048	
-4	0.023	
-5	0.010	
-6	0.004	
-7	0.001	
Puntaje-de-Corte de Veracidad en Exploración de Asunto-Múltiple y Faceta -Múltiple		
Puntaje-de-Corte NSR (4RQs)	Alfa No Corregida	Alfa Observada
0	0.252	0.070
1	0.159	0.042
2	0.091	0.024
3	0.048	0.012
4	0.023	0.006
5	0.010	0.002
6	0.004	0.001
Puntaje-de-Corte de Veracidad en Exploración de Asunto-Múltiple y Faceta -Múltiple		
Puntaje-de-Corte	Alfa No Corregida	Alfa Observada

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

NSR (3RQs)		
0	0.252	0.092
1	0.159	0.056
2	0.091	0.031
3	0.048	0.016
4	0.023	0.008
5	0.010	0.003
6	0.004	0.001
Puntaje-de-Corte de Veracidad en Exploración de Asunto-Múltiple y Faceta -Múltiple		
Puntaje-de-Corte NSR (2RQs)	Alfa No Corregida	Alfa Observada
0	0.252	0.135
1	0.159	0.083
2	0.091	0.047
3	0.048	0.024
4	0.023	0.011
5	0.010	0.005
6	0.004	0.002
7	0.001	0.001

This article is copyrighted by the American Polygraph Association (APA), and appears here with the permission of the APA. La American Polygraph Association (APA) tiene los Derechos de Autor de este artículo, y aparece aquí con el permiso de la APA

*Para cualquier corrección o consulta, el traductor puede ser contactado en [rodolfo@poligrafia.com.mx](mailto:rodolfo@poligrafia.com.mx)*