Guía de Referencia del Sistema de Puntuación Empírico - Multinomial (ESS-M) Raymond Nelson

Abstract

El Sistema de Puntuación Empírico - Multinomial (ESS-M*) es un protocolo estandarizado y basado en evidencia, para analizar los datos de las pruebas de detección psicofisiológica del engaño (PDD*). El ESS-M convierte en puntajes numéricos las respuestas fisiológicas de los sensores poligráficos, que luego se analizan mediante el uso de un modelo de referencia multinomial y un clasificador Bayesiano. Sus componentes clave incluyen la selección de las parejas de análisis entre las preguntas relevantes y comparativas(RQ*/CQ*), la asignación de puntajes multinomiales de tres posiciones y de puntajes ponderados para la actividad electrodérmica (EDA*). Se muestran los puntajes de corte y las tablas de referencia del ESS-M para una probabilidad previa equivalente (con un alfa = .05) y que se basan en el límite inferior del intervalo de credibilidad Bayesiano. Esta guía de referencia proporciona información para la extracción de características, transformación numérica y reducción de datos, cálculo de probabilidades a posteriori e intervalos creíbles, y la ejecución de reglas de decisión estructuradas ESS-M para analizar los resultados de la prueba a partir de los datos numéricos y probabilísticos. La guía también proporciona información para reportar los resultados del ESS-M y para seleccionar los puntajes de corte para los diferentes valores alfa y condiciones previas. Esta guía, que hace énfasis en el soporte empírico y el rigor científico, es meramente informativa y no pretende ser una política o norma de cumplimiento obligatoria. Para obtener información sobre la validez y el tamaño de los efectos de las pruebas PDD, se remite a los lectores a los estudios publicados.

This paper was translated with permission from the Editor and is hereby distributed among AIPP Members for knowlege-sharing purposes. Any other use is forbidden. Asociación Internacional de Profesionales en Poligrafía is an affiliate association of the American Polygraph Association.

Este artículo fue traducido con permiso del Editor de la revista, y se distribuye entre los miembros de la AIPP con propósitos de compartir conocimiento. Cualquier otro uso queda prohibido. La Asociación Internacional de Profesionales en Poligrafía es una asociación afiliada a la Asociación Americana de Poligrafía.

En este documento, el traductor conservará las siglas de todos los conceptos en inglés, ya que son los que utilizan los expertos en poligrafía de habla hispana.

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

Traductores jpadilla@deltasis.co, rodolfo@poligrafia.com.mx, eduardo@poligrafia.com.pa

Contenidos

1. Introducción y propósito
2. Ámbito de autoridad
3. Definiciones
4. Análisis global
5. Extracción de características
   - 5.1 Actividad electrodérmica
   - 5.2 Actividad cardiovascular
   - 5.3 Actividad respiratoria
   - 5.4 Actividad vasomotora
   - 5.5 Actividad fisica
   - 5.6 Rechazo de artefactos
   - 5.7 Valores de extracción de características ausentes

6. Transformación numérica y reducción de datos
   - 6.1 Selección de las parejas de análisis RQ/CQ
   - 6.2 La regla de Más Grande es Mejor
   - 6.3 Puntajes de tres posiciones
   - 6.4 Puntajes ponderados del EDA
   - 6.5 Puntajes ausentes
   - 6.6 Hoja de puntuación ESS-M
   - 6.7 Una puntuación respiratoria para ambos sensores respiratorios
   - 6.8 Reducción de datos mediante suma

7. Función de verosimilitud (distribuciones de referencia)
   Tabla 1. Distribución de referencia para puntajes de gran total (exámenes de asunto único)
   Tabla 2. Distribución de referencia para puntajes subtotales (exámenes de asunto único)
   Tabla 3. Distribución de referencia para puntajes subtotales (exámenes de asuntos múltiples)
   - 7.1 Obtención de posibilidades posteriores de engaño o veracidad en exámenes de asunto único
   - 7.2 Obtención de posibilidades posteriores de veracidad en exámenes de asuntos múltiples
   - 7.3 Puntajes de corte numéricos
     Tabla 4. Puntajes de corte ESS-M
     - 7.3.1 Cálculo de puntajes de corte para exámenes de asunto único
     - 7.3.2 Cálculo de puntajes de corte para exámenes de asuntos múltiples

8. Interpretación de los resultados de prueba
   - 8.1 Reglas de decisión
     - 8.1.1 Regla de gran total (GTR)
     - 8.1.2 Regla de subtotal (SSR)
     - 8.1.3 Regla de dos etapas (TSR)
   - 8.2 Análisis de los resultados y de las preguntas de la prueba
   - 8.3 Documentación y reporte

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

Traductores jpadilla@deltasis.co, rodolfo@poligrafia.com.mx, eduardo@poligrafia.com.pa

9. Trabajo con posibilidades previa y valores alfa diferentes
9.1 Cálculo de posibilidades posteriores mediante el Teorema de Bayes
Tabla 5. Posibilidades Posteriores de Culpabilidad con probabilidades previas de culpabilidad altas y bajas
9.2 Cálculo del intervalo de credibilidad mediante el método de Clopper-Pearson
9.3 Cálculo de los puntajes numéricos a partir del alfa y la posibilidad previa

Referencias
Apéndice A: Guía de referencia rápida del ESS-M
Apéndice B: Hoja de puntuación manual del ESS-M
Apéndice C-1: Tabla de referencia ESS-M para los puntajes de gran total
Apéndice C-2: Tabla de referencia ESS-M para los puntajes subtotales de exámenes de asunto único
Apéndice C-3: Tabla de referencia ESS-M para puntajes subtotales de exámenes de asunto múltiple
Apéndice D. Hoja de trabajo para el análisis del ESS-M
Apéndice E-1: Puntajes de corte ESS-M para exámenes de asunto único
Apéndice E-2: Puntajes de corte ESS- M para exámenes de asunto múltiple

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
Traductores jpadilla@deltasis.co, rodolfo@poligrafia.com.mx, eduardo@poligrafia.com.pa

1. Introducción y Propósito

El Sistema de Puntuación Empírico - Multinomial (ESS-M*) es un protocolo para el análisis de los datos de prueba (TDA*) de la detección psicofisiológica del engaño (PDD*), estandarizado y basado en evidencia. El ESS-M utiliza un clasificador Bayesiano para evaluar la *significancia estadística* de las posibilidades posteriores de engaño o veracidad. Las distribuciones de referencia del ESS-M se calculan con base en la teoría de la técnica de preguntas de comparación (CQT*) (Nelson, 2016a). Estas describen las probabilidades multinomiales asociadas a los puntajes numéricos de los formatos de prueba, incluyendo tanto los métodos de mentira probable (PL*) y mentira dirigida (DL*), que constan de tres a cinco presentaciones de dos a cuatro preguntas relevantes y que utilizan el conjunto tradicional de sensores de registro poligráfico: respiratorio, electrodérmico, cardiovascular y (opcionalmente) vasomotor (Nelson, 2017a). La intención de esta guía es proporcionar información sobre todos los aspectos del TDA que se utilizan en el ESS-M y que incluyen: a) extracción de las características en las respuestas fisiológicas que se correlacionan con la veracidad y el engaño, b) transformación de las respuestas fisiológicas en puntajes numéricos y reducción de datos o el agregado de los puntajes numéricos, c) uso de una *función de verosimilitud* o *datos de referencia* para el cálculo de valores estadísticos y puntajes de corte, y d) procedimiento para el uso de reglas de decisión para la clasificación e interpretación de los valores numéricos y estadísticos que son indicativos de engaño o veracidad. [En el apéndice A figura un breve resumen de los procedimientos del ESS-M].

1.1 Antecedentes. El ESS-M puede pensarse como una modificación o evolución de los métodos de puntuación manual previamente descritos por Utah (Bell et al., 1999) y el Departamento de Defensa (2006), que se remontan a Backster (1979) y también a Kubis (1962; 1965), los cuales hicieron uso de una escala de Likert (1932) para la transformación numérica de valoraciones subjetivas. El ESS-M es una actualización del Sistema de Puntuación Empírico (Nelson et. al., 2011) que utiliza métodos Bayesianos y una distribución de referencia multinomial. El ESS-M pretende generalizarse para los diferentes formatos CQT, como son los descritos por la American Polygraph Association (2011).

1.2 Objetivos. Un objetivo importante del ESS-M es que todos sus procedimientos y supuestos tengan soporte empírico. Otro objetivo es que las conclusiones categóricas de engaño y veracidad estén respaldadas por un análisis reproducible. Es igualmente importante que el *clasificador estadístico* que mide la fuerza de la información probabilística que respalda la conclusión categórica, proporcione información intuitiva sobre las probabilidades prácticas asociadas con estas mismas categorías, lo que hace necesario el uso de métodos Bayesianos. Los objetivos adicionales incluyen la explicación de cómo los procedimientos y el análisis de la prueba se relacionan con la teoría de la prueba poligráfica y la incorporación del sensor vasomotor opcional en el *modelo de referencia* y en los puntajes de corte de decisión. El ESS-M pretende proporcionar un marco intuitivo para comprender los resultados de la prueba poligráfica y garantizar que sus *tamaños de efecto* igualen o superen los del ESS. Un objetivo final es que el ESS-M pueda ejecutarse manualmente o mediante un algoritmo computarizado automatizado.

2. Ámbito de Autoridad

Esta guía de referencia se proporciona únicamente con fines informativos y educativos. No se expresa ni se asume alguna autoridad, y esta guía no pretende ser una política o norma de cumplimiento obligatorio. Sirve únicamente como referencia acerca de los principios y procedimientos del TDA utilizando el ESS-M con datos CQT. Este documento no pretende ser autor ni árbitro de aspectos científicos como son la definición y optimización de los parámetros de extracción de características, o la selección de alfas, previos y otros parámetros, que deben determinarse mediante el empirismo y el análisis. Adicionalmente, esta guía no proporciona información sobre la validez o el *tamaño de los efectos* del ESS - M y del CQT; para obtener dicha información, se remite a los lectores a los estudios publicados. Esta guía de referencia es meramente informativa y no pretende ser una política o norma de cumplimiento obligatorio. Los lectores deben consultar los estudios publicados para obtener información sobre la validez y el *tamaño del efecto* de las pruebas PDD. Cuando esta guía de referencia difiera de las políticas de un organismo, de las normas o leyes locales, aquellos requerimientos deberán prevalecer.

3. Definiciones

Para un uso experto del ESS-M, es fundamental estar familiarizado con un vocabulario conceptual básico de lo que es ciencia, pruebas, clasificación estadística y análisis Bayesiano. Se ofrecen las siguientes definiciones a manera de introducción y referencia.

3.1 Alfa. Dentro del paradigma de *las pruebas de significancia de las hipótesis nulas* (NHST*) (Fisher, 1934; Cohen, 1994; Jaynes, 2003), y con base en una definición *frecuentista* de la probabilidad, el alfa (α) representa el umbral para aceptar el riesgo de un error Tipo I o falso positivo. A menudo se establece un α = .05 que denota la probabilidad máxima aceptable para rechazar una hipótesis nula verdadera. El nivel alfa define el *nivel de significancia* o el nivel de *significancia estadística*. En el ESS-M el alfa se utiliza para definir la amplitud del *intervalo de credibilidad* Bayesiano (1 - α x 100% = 95%), que define el rango dentro del cual esperamos encontrar resultados de pruebas similares. En otras palabras, existe un margen de incertidumbre del 5% de que los resultados de otra prueba similar podrían estar fuera del intervalo esperado.

3.2 Pareja de análisis. En el CQT, un pareja de análisis es la selección de la pareja o combinación de una *pregunta relevante* y una *pregunta comparativa*. Cada pareja de análisis consta de una sola *pregunta comparativa (CQ)* y una sola *pregunta relevante (RQ)*. Cada análisis requiere de la selección de qué CQ se va a aparear contra cada RQ. Algunas técnicas poligráficas permiten la consideración de dos CQ para cada RQ, aunque la comparación final será entre una CQ y una RQ. Dentro de cada pareja de análisis los valores de la respuesta fisiológica ante la pregunta relevante se comparan con los de la pregunta comparativa. A cada pareja de análisis se le asigna una puntuación *multinomial* entera, y posteriormente se analizan estas puntuaciones en busca de una asociación probabilística con el engaño o la veracidad, basado en la teoría de la prueba poligráfica.

3.3 **Artefacto.** En el TDA con datos de prueba PDD, los artefactos son anomalías en los datos fisiológicos, que se desvían de los patrones esperados en la actividad tónica y fásica. Algunos datos con artefactos son involuntarios y, por lo tanto, no pueden interpretarse más allá de su impacto perturbador en los datos registrados. Otro tipo de datos con artefactos pueden producirse por causas voluntarias o involuntarias, y pueden ser el resultado de contramedidas o de distorsiones estratégicas o sistemáticas durante las pruebas. Aunque no se conozca la causa exacta de cualquiera de los artefactos observados, se pueden realizar clasificaciones probabilísticas reproducibles de las distorsiones, mediante la aplicación de métodos estadísticos al patrón de ocurrencia.

3.4 **Análisis Bayesiano (inferencia Bayesiana, estadística Bayesiana).** El análisis Bayesiano (BA*), también conocido como inferencia Bayesiana, se refiere al uso *del teorema de Bayes* y de los datos disponibles, para realizar inferencias probabilísticas sobre un parámetro desconocido de interés (Berger, 1985; Downey, 2012; Efron, 1986; Gelman, et. al., 2014; Lee, 2004; Stone, 2013; Winkler, 1972). Se basa en una definición Bayesiana de la probabilidad. El BA trata el parámetro de interés (es decir, el engaño o la veracidad) como un valor de probabilidad en el que se utilizan los datos de la prueba/experimento, junto con una estimación de *probabilidad a priori*, como la base de información para calcular una probabilidad posterior. Las entradas para el BA son los datos de la prueba (puntajes de prueba), una *función de verosimilitud* y una *probabilidad a priori*. El BA, unido al NHST es uno de los métodos más útiles e importantes para el análisis de datos y para la toma de decisiones científicas. Los resultados del Análisis Bayesiano suelen expresarse como Factor de Bayes.

3.5 **Factor de Bayes.** El factor de Bayes se utiliza para describir los resultados *del Análisis Bayesiano* (Jarosz y Wiley, 2014; Kaas y Raftery, 1993, 1995; Morey y Rouder, 2011). Indica la magnitud del cambio en la fuerza de la información entre la *probabilidad a priori* y la *probabilidad a posteriori*, que se expresa en forma de posibilidades.

3.6 **Teorema de Bayes.** En matemáticas, un teorema es una afirmación que se ha demostrado como verdadera a partir de afirmaciones previamente establecidas que podrían ser otros teoremas y axiomas. Los teoremas se derivan mediante el razonamiento lógico y proporcionan verdades fundamentales que pueden utilizarse para resolver otros problemas matemáticos o para demostrar teoremas adicionales. El teorema de Bayes fue descrito por primera vez por un estadista históricamente famoso llamado Thomas Bayes (Bayes, & Price, 1763; Jeffreys, 1939, Lee, 2004; Stone, 2013). Proporcionó una fórmula matemática para actualizar las probabilidades de diferentes hipótesis posibles mediante la integración del conocimiento previo con nuevas pruebas, y que es esencial para el *Análisis Bayesiano*.

El teorema de Bayes se expresa como:
P(H|E) = [P(E|H) x P(H)] / P(E)
Donde:
- **P(H|E)** es la probabilidad posterior de la hipótesis (H) dada la evidencia (E)
- **P(E|H)** es la posibilidad de observar la evidencia (E) dado que (H) es verdadera

- **P(H)** es la probabilidad a priori de la hipótesis antes de observar la evidencia
- **P(E)** es la posibilidad marginal o la probabilidad de los datos, bajo todas las hipótesis

3.7 Pregunta comparativa. Dentro del contexto CQT, la pregunta comparativa (CQ) está diseñada para obtener una respuesta más fuerte que en la pregunta relevante (RQ) en una persona veraz. Estas preguntas se han utilizado principalmente de dos maneras: como preguntas de mentira probable, en las que se manipula al sujeto para que responda de una forma que se asume es incorrecta o evasiva y como preguntas de mentira dirigida, en las que se instruye al sujeto para que responda de manera incorrecta. [Vea Nelson et al., 2019 para información más detallada]. Las preguntas comparativas son un elemento importante del CQT porque toda actividad fisiológica está asociada con múltiples fenómenos – no existe un tipo de actividad fisiológica que esté asociada exclusivamente con el engaño. También son importantes porque un principio básico de la ciencia es que todas las conclusiones científicas se realizan teniendo en cuenta otras conclusiones posibles. El uso de preguntas comparativas - comparación *ipsativa* con las *preguntas relevantes* - es fundamental para la evaluación de las *preguntas relevantes* y la puntuación numérica de los datos en un CQT.

3.8 Técnica de Preguntas comparativas (prueba de preguntas de comparación). Una Prueba de Preguntas de Comparación (CQT) es una técnica poligráfica comúnmente utilizada en las pruebas PDD. Consiste en comparar las respuestas fisiológicas de las *preguntas relevantes*, que pertenecen directamente al tema investigado, contra las respuestas en las *preguntas comparativas*, que están diseñadas para provocar respuestas basadas en una combinación de cognición, emoción y comportamiento. El procedimiento de prueba y el análisis de datos se basa en la teoría de que los mayores cambios en la actividad fisiológica se cargan ante diferentes tipos de estímulos de prueba en función del engaño o la veracidad en respuesta a los estímulos objetivo-relevantes. El CQT es ampliamente utilizado en contextos forenses y exploratorios.

3.9 Corrección de continuidad. Una corrección de continuidad es un ajuste matemático simple que se realiza cuando se utiliza una distribución de probabilidad escalonada (discreta). Puede realizarse mediante la suma o resta de un valor pequeño y promediar los valores adyacentes, o mediante otros métodos. El uso de una corrección de continuidad puede ayudar a mejorar tanto la interpretación práctica como la precisión de las estimaciones estadísticas, garantizando que la estimación siempre supere el valor real estimado.

3.10 Intervalo de credibilidad (intervalo de confianza Bayesiano). En estadística Bayesiana, un Intervalo de credibilidad (IC) es análogo al intervalo de confianza *frecuentista* (Jaynes, 1976; 2003; Rice, 2006). Pero mientras los intervalos de credibilidad se basan en una teoría/definición *frecuentista* de la probabilidad, los Intervalos de Confianza se basan en la teoría Bayesiana de la probabilidad. Esto tiene importantes implicaciones en cuanto a la manera en que se interpretan estos intervalos - el significado y la información que pueden transmitir. *Los* intervalos de confianza *frecuentista* son la porción estimada de otros intervalos de confianza (basados en otros conjuntos de datos de muestreo) que

contendrán el parámetro desconocido de interés (Agresti, 1998; Newcombe, 1998; Neyman, 1937; Rosner, 2010; Smithson, 2003). Los Intervalos de Confianza pueden interpretarse como el rango dentro del cual se encuentra el valor verdadero de un parámetro desconocido con una probabilidad especificada (por ejemplo, 90%, 95% o 99%). A diferencia de los intervalos *frecuentistas*, los Intervalos de Confianza Bayesianos proporcionan una interpretación más directa de la probabilidad de que el parámetro desconocido de interés se encuentre dentro del intervalo. Los Intervalos de Confianza del ESS- M para las probabilidades posteriores, se calculan utilizando el método descrito por Clopper y Pearson (1934). Se seleccionó este método por la ventaja práctica de que los intervalos resultantes nunca subestiman la probabilidad de que el parámetro desconocido esté dentro del intervalo, y los límites superior e inferior del intervalo nunca superan los límites teóricos [0, 1]. El límite inferior del Intervalo de Confianza 1 - α se muestra en las tablas de referencia del ESS-M y se utiliza para calcular puntuaciones numéricas que garantizan que las clasificaciones de engaño y veracidad alcanzan el nivel deseado de *significancia estadística*.

3.11 Estado del criterio. El estado del criterio se refiere a la condición o estado real de lo medido en una prueba o experimento. En el contexto de las pruebas PDD, el estado del criterio es la condición verdadera del examinado, como por ejemplo si es realmente veraz o realmente mentiroso.

3.12 Engaño Indicado (DI) / Engaño No Indicado (NDI). Engaño Indicado (DI*) y Engaño No Indicado (NDI*) son etiquetas categóricas utilizadas para describir los resultados de los exámenes *diagnósticos* cuando las puntuaciones numéricas son *estadísticamente significativas* para el engaño (DI) o veracidad (NDI). Describen lo que sugieren los datos de la prueba sin que esto implique incorrectamente que se midió o detectó directamente el engaño o la veracidad. Estos términos son análogos a los términos categóricos más generales de "positivo" y "negativo" que se utilizan al presentar de resultados en otros tipos de pruebas científicas.

3.13 Examen diagnóstico. Los exámenes diagnósticos se realizan en respuesta a un problema conocido, como puede ser un incidente, un alegato o síntomas en el caso de los exámenes médicos y psicológicos (Centros para el Control y la Prevención de Enfermedades, 2012; James y Nordby, 2009). Los resultados de los exámenes diagnósticos tienen por objeto proporcionar una base para llevar a cabo una acción que aborde el problema identificado. Estos exámenes están diseñados para tener una alta sensibilidad y especificidad para asegurar una precisión alta. Los exámenes PDD logran esto mediante la formulación de preguntas de prueba de asunto único, para las que se asume que el *estado del criterio* de las RQ *no es independiente.*

3.14 Valores adimensionales. Los valores adimensionales (Maxfield, 2009) no dependen de unidades de medición físicas estandarizadas. Según el teorema Pi de Buckingham (Buckingham, 1914), se pueden realizar cálculos válidos utilizando cualquier unidad de medida adimensional, siempre que se utilicen de forma consistente. Una ventaja de las

unidades adimensionales es que los datos procedentes de distintas fuentes pueden transformarse en valores unitarios similares y, a continuación, compararse o combinarse. Los valores adimensionales son muy útiles en muchas aplicaciones científicas y de evaluación. En el CQT, durante la *extracción de características,* se pueden asignar valores adimensionales a las diferentes fuentes de datos (por ejemplo, electrodermales, cardiovasculares, etc.) y éstos pueden transformarse y agregarse fácilmente sin estar restringidos a unidades físicas, que normalmente se utilizan para expresar fenómenos físicos.

3.15 Tamaño del efecto. El tamaño del efecto es una medida cuantitativa de la magnitud de un fenómeno o efecto en la investigación o en las pruebas científicas (Cohen, 1988; Ellis, 2010). Es una descripción numérica o estadística de la fuerza, de la relación o del grado esperado de cambio o de la diferencia resultante de una hipótesis o procedimiento especificado. Los tamaños del efecto proporcionan una forma estandarizada de comprender la utilidad práctica de los resultados independientemente del tamaño de la muestra, y a menudo se utilizan para comparar los resultados entre estudios.

3.16 Inferencia frecuentista (estadística frecuentista). La inferencia frecuentista utiliza métodos estadísticos para hacer inferencias o sacar conclusiones acerca de un parámetro desconocido basándose en los datos observados. Este enfoque asume que la realidad es fija (es decir, que la realidad existe de una sola manera) y que los datos incluyen una variabilidad inherente (es decir, error aleatorio de medición). Los métodos frecuentistas se basan en la frecuencia de los sucesos observados para hacer inferencias y predicciones sobre parámetros desconocidos en la población [Consulte Laplace (1812) y Fisher, (1934)].

3.17 Inconcluso (INC) / No opinión (NO). Etiqueta categórica utilizada para describir el resultado de una prueba cuando sus datos y las puntuaciones numéricas no son *estadísticamente significativas* para el engaño o la veracidad, cuando los datos de la prueba no respaldan alguna opinión.

3.18 Independencia (varianza independiente). Varianza independiente significa que las respuestas ante los estímulos de la prueba (RQ en las pruebas PDD) no tienen una fuente compartida de variación - que la variación en respuesta a un ítem no puede influir en la variación de otros ítems. Por ejemplo, el lanzamiento de una moneda es independiente porque cada lanzamiento no está influido por los otros lanzamientos (suponiendo que se trate de una moneda no sesgada). Del mismo modo, se requiere que los casos seleccionados para las muestras de investigación sean independientes e idénticamente distribuidos (IID), lo que significa que cada miembro de una población tiene la misma probabilidad de ser incluido en la muestra. En los exámenes *exploratorios* de asunto múltiples PDD, se asume que el *estado de criterio* de las RQ es independiente. Cuando no se cumple el requisito de independencia - cuando se concibe que los ítems de la prueba tienen una fuente de variación compartida - se dice que los ítems *no son independientes*. Mientras que la suposición de independencia es un requisito riguroso, la suposición de no independencia no lo es. En otras palabras, *la no independencia* no requiere una suposición explícita de dependencia; basta con saber que una suposición de independencia no es

razonable. En los exámenes PDD *de diagnóstico de* asunto único, se asume que el *estado del criterio* de las RQ *no es independiente*.

3.19 Ipsativa (comparación ipsativa). La comparación ipsativa involucra la comparación de datos de un individuo con otros datos obtenidos en el mismo individuo (Baron, 1996; Hicks, 1970). El paradigma CQT hace uso de métodos ipsativos al contrastar las respuestas en las *preguntas relevantes* contra las respuestas a las *preguntas comparativas* en el momento en que se asignan puntuaciones *multinomiales,* que luego pueden compararse contra un *modelo normativo de referencia* y con el *Análisis Bayesiano* de la fuerza probabilística de la información que apoya las hipótesis de engaño o veracidad como posibles causas de los datos observados.

3.20 Función de verosimilitud. En clasificaciones estadísticas, una función de verosimilitud, puede pensarse como cualquier dispositivo utilizado para obtener un valor estadístico en los datos observados (Casella y Berger, 2002). Puede tratarse de una fórmula, cuyos cálculos se realizan manualmente y también mediante una aplicación informática que puede ejecutar cálculos complejos de forma rápida y confiable. Una forma sencilla y útil de una función de verosimilitud es una tabla de referencia de los valores numéricos que representan una distribución de probabilidad (una lista de todos los valores posibles y las probabilidades asociadas a cada uno de ellos). La forma más sencilla de una función de verosimilitud o *modelo de referencia* es un puntaje de corte numérico, seleccionado con el objetivo de alcanzar *los tamaños de efecto de* probabilidad deseados.

3.21 Multinomial (puntuaciones multinomiales, distribución multinomial). Multinomial se refiere a información en la que los valores finales de un resultado pueden involucrar más de dos posibilidades (Abramowitz y Stegun, 1972). En contraste, las puntuaciones binomiales - cuando sólo hay dos resultados posibles - es un caso especial de lo multinomial y suele incluirse en las lecciones introductorias sobre la teoría de la probabilidad. Las distribuciones multinomiales se calculan mediante la combinación de fórmulas (Chen y Koh, 1992; Skiena, 1990), y también pueden caracterizarse o calcularse mediante una simulación Monte Carlo. La distribución multinomial de referencia para los puntajes ESS-M en los datos CQT se basa en la teoría analítica de la prueba poligráfica – los mayores cambios en la actividad fisiológica se cargan ante los diferentes tipos de estímulos de prueba en función del engaño o veracidad en respuesta a los estímulos objetivo-relevantes. Los puntajes ESS-M de los datos CQT son multinomiales porque la puntuación numérica para cada *pareja de análisis* tiene tres valores posibles [+, 0, -], correspondientes a las posibilidades de que se observe un mayor cambio en la actividad fisiológica en la *pregunta comparativa*, en la *relevante* o en ninguna de las dos. Bajo la hipótesis nula, estas posibilidades son equivalentes y su frecuencia de aparición puede caracterizarse como aleatoria o calcularse mediante una fórmula combinada. Los puntajes ESS-M multinomiales de alguna manera son similares a los puntajes Likert (1932). Sin embargo, mientras que las puntuaciones Likert están pensadas para la calificación y transformación numérica de información subjetiva, las puntuaciones multinomiales se asignan en condiciones objetivas, lo que los hace más confiables y más susceptibles a suposiciones interválicas (Stevens, 1946; 1951).

3.22 Multiplicidad (error acumulativo, alfa inflado, el problema de las comparaciones múltiples). Los efectos de multiplicidad o del error acumulativo surgen dentro del contexto de la clasificación y la decisión estadística cuando múltiples decisiones estadísticas contribuyen para la obtención de una sola conclusión. Este efecto puede inflar las tasas de error y reducir la exactitud o precisión de los resultados (Abdi, 2007, McDonald, 2009, Miller, 1981; Sidák, 1967). Cuando se utilizan puntuaciones subtotales en un CQT, esto puede crear un aumento de errores falsos positivos e *inconclusos*. Cuando es posible, los efectos de multiplicidad pueden mitigarse evitando o reduciendo el uso de comparaciones estadísticas múltiples, seleccionando métodos de análisis que sean menos vulnerables a estos efectos de multiplicidad (es decir, ANOVA y otros métodos ómnibus) y mediante el uso cuidadoso de *correcciones estadísticas*.

3.23 No independiente (varianza no independiente). La varianza no independiente se refiere a aquellas respuestas ante los estímulos de prueba - como las RQ en las pruebas de PDD – que tienen un factor común de influencia subyacente o una fuente de variación compartida. Esto significa que la variación en la respuesta a un ítem puede verse afectada por la variación en las respuestas a otros ítems haciendo que las respuestas sean interdependientes. En los exámenes PDD de asunto único (*diagnóstico*), se asume que la varianza del criterio de las RQ no es independiente, lo que indica que las respuestas no están aisladas unas de otras, sino más bien vinculadas por un *estado de criterio* común y a una fuente común de varianza.

3.24 Datos de referencia normativos (modelo de referencia, datos de referencia, datos normativos). Los datos de referencia normativos se refieren al uso de un modelo de referencia que describe las observaciones o resultados esperados (Cohen *et al.*, 2013). En el CQT, los métodos normativos utilizan la comparación de las puntuaciones *ipsativas* de cada individuo contra los datos obtenidos en una muestra o distribución de referencia, que describe lo que se espera bajo una hipótesis especificada. Los datos de referencia para el ESS-M son *multinomiales*, calculados a partir de la teoría básica del CQT - que los mayores cambios en la actividad fisiológica se producen con mayor frecuencia ante diferentes tipos de estímulos de prueba (*preguntas relevantes* y *preguntas comparativas*) en función del engaño o veracidad en respuesta a la preguntas objetivo-relevantes. La forma más sencilla de datos normativos de referencia es un puntaje de corte numérico calculado a partir de un modelo estadístico de referencia.

3.25 Prueba de significancia de la hipótesis nula (NHST*). La prueba de significancia de la hipótesis nula (NHST; Fisher, 1934; Cohen, 1994; Jaynes, 2003; Neyman, 1937; Neyman & Pearson, 1933; Mayo, 1981) es uno de los dispositivos más importantes y útiles para la toma de decisiones científicas. Sin embargo, no está exenta de cierta controversia debido a la rigurosidad de las definiciones en la teoría de la probabilidad, (Colquhoun, 2014; Nickerson, 2000). La NHST requiere del establecimiento de un nivel *alfa* (α) o tolerancia al error antes de realizar una prueba o experimento científico, seguido del cálculo del rango esperado de variabilidad debido a los datos disponibles. La NHST requiere hacer una elección forzada entre una hipótesis nula (a menudo basada en la idea de que los datos

pueden caracterizarse efectivamente como aleatorios) y una hipótesis alternativa. Aunque suele ser difícil (o imposible) calcular la distribución esperada para la hipótesis alterna, suele ser más sencillo calcularla con una hipótesis nula. El rechazo de la hipótesis nula se produce cuando el valor-p calculado es menor o igual que *alfa* ($p \leq \alpha$), lo que indica que la probabilidad de obtener los datos observados bajo la hipótesis nula es suficientemente baja. En esos casos, se dice que los datos dan soporte a la hipótesis alterna.

3.26 Probabilidad a priori (posibilidades a priori). En el A*nálisis Bayesiano*, las estimaciones de probabilidad a priori representan una estimación inicial de las probabilidades de los diferentes resultados antes de realizar una prueba o experimento científico, y se basan en los conocimientos o supuestos existentes sobre una situación. La información probabilística puede transformarse y presentarse de diferentes formas, como probabilidades decimales, porcentajes y posibilidades. La información previa forma parte de la base del *Análisis Bayesiano*. Las probabilidades a priori pueden calcularse a partir de las tasas de la sensibilidad de la prueba, junto con los resultados observados (Raskin, 1986; 1987) [Véase un procedimiento detallado en Nelson (2016b)]. Los valores de probabilidad pueden transformarse a posibilidades mediante esta fórmula: posibilidades = p / (1 - p). A la inversa, las posibilidades pueden transformarse en probabilidades de esta forma: p = posibilidades / (1 + posibilidades).

3.27 Probabilidad posterior (posibilidades posteriores). En el *Análisis Bayesiano,* la probabilidad posterior es la probabilidad actualizada de un resultado después de analizar los datos junto con la *probabilidad previa y el Teorema de Bayes*. Refleja la probabilidad revisada de una hipótesis dada la evidencia. La información sobre probabilidades puede transformarse y presentarse de distintas formas, como probabilidades decimales, porcentajes y posibilidades. A menudo, los resultados *del Análisis Bayesiano* se presentan en forma de posibilidades porque permite discutir la información de la probabilidad utilizando números enteros e ilustra explícitamente que todas las probabilidades indican la fuerza de cierta información con respecto a la fuerza de otra información.

3.28 Segmento post-estímulo. El segmento post-estímulo (también periodo post-estímulo) es un breve periodo de tiempo después de la *Ventana de Evaluación*, durante el cual se espera que los datos vuelvan a un nivel tónico. Junto con la *Ventana de Evaluación*, influirá en la duración de las preguntas durante la prueba.

3.29 Segmento pre-estímulo. El segmento pre-estímulo es un breve periodo de tiempo previo al inicio del estímulo. El segmento pre-estímulo de cada pregunta es también el periodo post-estímulo de la pregunta anterior.

3.30 Probabilidad. La probabilidad *frecuentista* se refiere al número de veces que se observa la ocurrencia de algo en comparación con el número de veces que podría ocurrir (Fisher, 1934; Laplace 1812; Jaynes, 2003). Aunque es muy útil, esta definición tiene algunas limitaciones en su aplicación con fenómenos que no pueden observarse (por ejemplo, sucesos futuros y sucesos pasados lejanos) y sucesos que no se repiten. En cambio, la probabilidad B*ayesiana* hace referencia al grado de certeza ante diferentes conclusiones y

tiene una amplia aplicación práctica en una amplia variedad de fenómenos (Bayes y Price, 1763; Berger, 1985; Gelman et al., 2014; Jeffreys, 1939; Stone, 2013; Winkler, 1972). Existen otras definiciones de probabilidad. En general, la probabilidad se refiere a la fuerza cuantitativa de la evidencia para diferentes conclusiones posibles en condiciones de incertidumbre.

3.31 Pregunta relevante. En el contexto del CQT, una pregunta relevante (RQ) describe el posible involucramiento de un examinado en un comportamiento bajo investigación. Aunque ninguna respuesta fisiológica está asociada de forma exclusiva a una conducta o actividad concreta, pueden evaluarse las respuestas ante las preguntas relevantes y las respuestas a las *preguntas comparativas*, para obtener su correlación estadística con el engaño y la veracidad.

3.32 Periodo de Latencia de la Respuesta (periodo de latencia). El Periodo de Latencia de la Respuesta (RLP*) es el tiempo a partir del inicio del estímulo, durante el cual no se interpreta el inicio de una respuesta. Idealmente, la duración del periodo mínimo de latencia se determina mediante el análisis y optimización estadística.

3.33 Ventana de Inicio de Respuesta. La Ventana de Inicio de Respuesta (ROW*) es el periodo de tiempo, que comienza al final del *Periodo de Latencia de la Respuesta,* durante el cual se espera que comience una respuesta. Las respuestas que comienzan fuera del ROW no se utilizan (no se puntúan). La longitud del ROW se determina idealmente mediante análisis y optimización estadística.

3.34 Saliencia (saliencia diferencial). El concepto de saliencia se refiere a la prominencia, o cualidad de un elemento para sobresalir dentro de un conjunto de elementos. [Véase Handler yNelson (2007) para un análisis de la saliencia dentro del contexto CQT]. En el contexto CQT, la saliencia diferencial se refiere a la observación de un mayor cambio fisiológico en *una pregunta relevante* o *comparativa* dentro de una *pareja de análisis,* y también se refiere a si se cargan mayores cambios en la actividad fisiológica en las *preguntas comparativas* o en las *relevantes.*

3.35 Examen exploratorio. En la medicina y la psicología, las pruebas exploratorias se realizan en ausencia de síntomas conocidos o dificultades funcionales observadas (Centers for Disease Control and Prevention, 2012). Dentro del contexto forense y de investigación, las pruebas exploratorias se llevan a cabo en ausencia de un problema conocido o señalado, como podría ser un incidente o alegato (James, & Nordby, 2009). Estas pruebas tienen por objeto identificar problemas no conocidos y proporcionar información para la gestión y evaluación de riesgos. En PDD, los exámenes exploratorios pueden realizarse como exámenes de asuntos múltiples, para los cuales se asume que el estado del criterio de las RQ es *independiente*, y también pueden realizarse como exámenes de asunto único para los que se supone que el estado del criterio *no es independiente*.

3.36 Reacciones Significativas (SR) / Sin Reacciones Significativas (NSR). Reacciones Significativas (SR) y Sin Reacciones Significativas (NSR) son etiquetas categóricas que se utilizan para describir los resultados de las *pruebas* poligráficas *exploratorias* cuando los

puntajes numéricos son *estadísticamente significativos* para el engaño (SR) o veracidad (NSR). Estos términos indican la conclusión obtenida debido a los datos de prueba, sin que esto implique erróneamente que se midieron o detectaron determinados acontecimientos o comportamientos de manera directa. Estos términos son análogos a los términos categóricos generales de "positivo" y "negativo" utilizados en otros tipos de evaluación científica y también son análogos a los términos categóricos de *Engaño Indicado* (DI) y *Engaño No Indicado* (NDI) utilizado con los exámenes poligráficos *de diagnóstico*.

3.37 Clasificador estadístico. Un clasificador estadístico es un modelo matemático o un algoritmo que es utilizado para categorizar datos en distintas clases o grupos con base en propiedades estadísticas. Se analizan los datos de entrada y se les asigna una categoría predefinida evaluando sus características y patrones dentro de los datos. Los clasificadores estadísticos se utilizan normalmente en el aprendizaje automatizado y en el análisis de datos para hacer predicciones o tomar decisiones, como cuando se determina si un correo electrónico es spam o no, o clasificar imágenes médicas sanas o con enfermedades. Ejemplos de clasificadores estadísticos incluyen los árboles de decisión, las máquinas de soporte de vectores, los clasificadores Bayesianos, redes neuronales y otros métodos. Un clasificador estadístico establece un resultado categórico, pero no necesariamente proporciona información sobre la fuerza práctica o el tamaño del efecto de ese resultado. Por ejemplo, un valor-p puede indicar la significancia estadística, pero no transmite implicaciones prácticas o reales asociadas con el resultado.

3.38 Significancia estadística (significativo, estadísticamente significativo, nivel de significancia). La significancia estadística es una medición que indica si los resultados de un análisis probablemente se deban al azar o que reflejan un efecto o relación verdaderos. Un resultado se considera estadísticamente significativo si la probabilidad de obtenerla al azar es muy baja, normalmente por debajo de un umbral predeterminado (nivel *alfa*, a menudo fijado en 0.05). Esto significa que es improbable que el efecto o la diferencia observada se hayan producido por casualidad. Los resultados del ESS-M son estadísticamente significativos, cuando la fuerza de las posibilidades posteriores de engaño o de veracidad difieren de las posibilidades previas en un nivel que improbablemente se deba al azar (variación inexplicable o aleatoria) y es probable que se observen o repliquen con otros datos de muestreo.

3.39 Ventana de evaluación. La Ventana de Evaluación (WOE*) es el periodo de tiempo fijo durante el cual se produce la *extracción de características*. Incluye el periodo *de latencia de respuesta y la ventana de inicio de la respuesta*.

4. Análisis Global

En el ESS-M, el análisis de los datos de prueba comienza con el análisis global. También denominado evaluación global, trata de una valoración cualitativa holística y exhaustiva de los datos registrados en la prueba (National Center for Credibility Assessment, 2017). Esto puede incluir las observaciones realizadas durante la entrevista, el nivel de funcionamiento durante el proceso de evaluación, características en la salud física y mental, la selección de objetivos, la formulación de preguntas y los hechos conocidos del caso. El objetivo principal del análisis global es determinar si los datos registrados son suficientes para el análisis numérico y estadístico, para que se garantice que los datos pueden representarse razonablemente en el *modelo de referencia* o *datos de referencia*. Además, el análisis global pretende identificar los datos de calidad problemática, como son los segmentos *con artefactos* que podrían requerir de una exclusión selectiva y estadística para valorar la probabilidad de falsificación estratégica. El análisis global es un enfoque clínico (es decir, un proceso semiestructurado que se basa en el profesionalismo, la experiencia, la formación y la habilidad, frente a un proceso estructurado, numérico, estadístico o actuarial). Puede involucrar la inspección y discusión de los detalles que se encuentran fuera del proceso de puntuación numérica objetiva.

5. Extracción de Características.

La extracción de características involucra la evaluación de la calidad de los datos, las marcas de los puntos de inicio y fin de la respuesta, y calcular los valores numéricos que representan a los cambios en la actividad fisiológica. Se espera que las respuestas comiencen al tiempo en que inicia el estímulo, después de éste, o después del *Periodo Mínimo de Latencia de la Respuesta* (*LPR*), y son evaluadas hasta el valor de pico máximo o al final de la *Ventana de Evaluación* (WOE). Los valores de la extracción de las características pueden expresarse como unidades de medición físicas o como unidades *adimensionales*. El ESS-M solamente utiliza respuestas primarias (Dutton, 2000; Kircher y Raskin, 1988; Raskin *et al.*, 1988, Krapohl & McManus, 1999; Nelson, 2020a; Nelson, *et al.*, 2011), y no incluye características de respuestas fisiológicas secundarias.

5.1 Actividad electrodérmica (EDA). La respuesta de interés en el EDA es la amplitud vertical máxima (eje - y) en el incremento de los datos que comienza dentro de la ROW, después de la RLP y hasta el final de la respuesta (inicio de un segmento de pendiente negativa) dentro del WOE. Se mide en unidades *adimensionales* que no están asociadas con las unidades de resistencia o conductividad reales.

5.2 Actividad cardiovascular. La actividad cardiovascular se mide por su amplitud vertical máxima (eje-y) en el incremento de los datos desde el inicio del ROW y hasta el final de la respuesta (inicio de un segmento de pendiente negativa) dentro de la WOE. Se mide en *unidades adimensionales* que no están asociadas con los valores reales de la presión arterial.

5.3 Actividad respiratoria. La característica respiratoria de interés es la reducción o supresión del movimiento asociado con la actividad respiratoria. Puede medirse en unidades *adimensionales* (no asociadas a mediciones del volumen respiratorio o mediciones lineales de la circunferencia torácica o abdominal) como son la excursión de la línea respiratoria (RLE*), que es la suma de las diferencias en los valores sucesivos de las muestras del eje-y desde el inicio del estímulo hasta el final del WOE.

5.3.1 Patrones respiratorios. Al calificar manualmente con el ESS-M, las respuestas respiratorias también pueden extraerse mediante el reconocimiento visual de patrones de respuesta que se correlacionan con la RLE.

5.3.1.a Disminución de la tasa respiratoria.

5.3.1.b Supresión o reducción de la amplitud respiratoria (asociada con el volumen tidal).

5.4 Actividad vasomotora. La característica de la reacción vasomotora de interés es una reducción o supresión del pulso vasomotor. Esto se evalúa calculando la proporción de la amplitud del pulso durante el *segmento pre-estímulo* contra la amplitud de pulso durante el estímulo después del *periodo mínimo de latencia de respuesta* vasomotora. Los cambios en la amplitud del pulso vasomotor indican cambios en el flujo sanguíneo capilar, y pueden medirse en *unidades adimensionales*.

5.5 Actividad física. Los valores de extracción de características de la actividad física no está incluida en los *modelos de referencia* ESS-M, y estos datos no se incluyen en el clasificador Bayesiano ESS-M. La actividad física se evalúa a través del reconocimiento de patrones, y no está sujeta a una medición real.

5.6 Rechazo de artefactos. Para cada sensor de registro poligráfico, la extracción de características se restringe a los segmentos donde la respuesta observada se puede atribuir de manera razonable a los estímulos de prueba. Las respuestas deben ser oportunas ante los estímulos, y deben producirse en ausencia de otra causa posible observable (por ejemplo, *artefacto por* conducta, estímulo externo o por error en el procedimiento) distinta a los estímulos de la prueba.

5.6.1 Respuesta o actividad a destiempo. Respuestas que se producen antes del inicio del estímulo o antes del final de la RLP y que pueden deberse a causas desconocidas y se excluyen de la extracción de características. Los cambios fisiológicos que comienzan después del ROW también se excluyen de la extracción de características.

5.6.2 Otros estímulos. Cuando se observe otra posible causa, los cambios en la actividad fisiológica se excluyen de la extracción de características.

5.6.2.a Comportamiento no esperado. Movimiento, hablar, respiración profunda u otra actividad no esperada durante el segmento del estímulo.

5.6.2.b Estímulo exterior. Cualquier movimiento, ruido o actividad no esperada en el entorno cercano que pueden generar cambios en la actividad fisiológica por novedad o sorpresa.

5.6.2.c Error del examinador. Desviaciones imprevistas o fortuitas de los procedimientos de prueba que se presentan en el momento en que se observa un cambio en la actividad fisiológica.

5.7 Valores de extracción de características ausentes. Cuando un segmento de estímulo tiene un cambio a destiempo en la actividad fisiológica, no se asigna un valor de extracción de características, lo que significa que no puede identificarse un inicio de respuesta dentro del ROW. También pueden tenerse valores ausentes cuando los datos fisiológicos registrados son inestables al grado de que la observación de la actividad tónica y fásica no es fiable. Adicionalmente, no se puede obtener un valor para la extracción de características cuando el cambio en la actividad fisiológica ocurre al tiempo del estímulo de prueba, pero también con la observación de un *artefacto* (por ejemplo, comportamiento no esperado, estímulo externo). En tales casos, el cambio en la actividad fisiológica no puede atribuirse razonablemente al estímulo debido a la influencia confusa del *artefacto*. Cuando no se obtiene ningún valor de extracción de características, se permite dejar el lugar de entrada de datos en blanco o vacío. Sin embargo, es preferible introducir NP* (sin puntuación) o NA (no aplica) para evitar ambigüedades o confusión sobre si el valor que falta se debe a un descuido o a un error.

De cierto modo, la transformación numérica en el ESS-M es la culminación de la extracción de características. Consiste en convertir los datos registrados y las características de respuesta en valores numéricos mediante el uso de mediciones físicas o *valores adimensionales*. Luego, los valores de la extracción de características se combinan en *parejas de análisis* (pares de RQ y CQ). Después se asignan puntuaciones enteras *multinomiales* a cada pareja de análisis. La reducción de datos es un proceso definido y estructurado para agregar los valores numéricos en un conjunto más pequeño de valores, como son las puntuaciones de subtotal y gran total. Existen diferentes métodos para agregar datos que incluyen la suma, la media, los puntajes z y las técnicas de ajuste de modelos. La agregación de los puntajes del ESS-M se realiza mediante la suma.

6.1 Selección de las parejas de análisis RQ y CQ. Cada *pareja de análisis* consta de una sola RQ yuna sola CQ para cada sensor de registro. Dentro de cada *pareja de análisis*, se evalúa la *saliencia* diferencial de la RQ y de la CQ en función de la magnitud del cambio en la actividad fisiológica de la RQ y la CQ. A continuación, se agregan las puntuaciones numéricas y se analizan utilizando un *modelo de referencia* o un *clasificador estadístico*. La selección *de las parejas de análisis* entre la RQ/CQ están dictados por la técnica PDD utilizada, convirtiéndose entonces en una tarea de memorización para los profesionales de campo. También se pueden seleccionar *las parejas de análisis* para prácticamente cualquier técnica PDD utilizando la siguiente heurística (Nelson, 2017b):

6.1.1 CQ precedente. Cuando hay una sola CQ adyacente precediendo la RQ, sin una CQ adyacente posterior a la RQ, evalúe la RQ con la CQ adyacente anterior a la RQ.

6.1.2 CQ precedente no utilizable. Si la CQ que precede a la RQ no es utilizable debido a un *artefacto*, o si no hay un valor de extracción de características para la CQ precedente, la RQ puede evaluarse contra la CQ subsecuente. Esto está permitido incluso cuando hay una única pregunta neutral dentro de la secuencia de preguntas entre la RQ y la CQ posterior.

6.1.3 CQs como corchetes. Cuando sea posible, seleccione la *pareja de análisis* utilizando las dos CQ adyacentes (es decir, inmediatamente anterior e inmediatamente posterior a) contra la RQ. En otras palabras, cuando hay una CQ adyacentes anterior y una posterior a una sola RQ se selecciona la CQ que produce el mayor cambio en la actividad fisiológica.

6.1.4 Pares de RQs con CQs como corchetes. Los pares de RQs adyacentes pueden ser evaluadas contra la CQ que se encuentra inmediatamente antes o después del par de las RQ. Cada una de las RQs emparejadas es un *pareja de análisis*, y cada RQ puede compararse contra la CQ anterior o posterior. Para cada una de las RQ, seleccione la CQ que produzca el mayor cambio en la actividad fisiológica.

6.2 Regla de "más grande es mejor". Dentro de cada *pareja de análisis*, determine si la RQ o la CQ es la que produjo el mayor cambio en la actividad fisiológica. Para asignar un puntaje manual, cualquier diferencia perceptible es suficiente (National Center for Credibility Assessment, 2017). Este proceso se basa en el principio de que incluso pequeñas diferencias en la actividad fisiológica pueden contribuir a aumentar los *tamaños de los efectos* de interés cuando se agregan los datos. [Véase un ejemplo en Nelson (2020b).] Las reacciones fisiológicas de mayor magnitud son provocadas por los estímulos que son más salientes para el examinado debido a factores emocionales, cognitivos o condicionados por el comportamiento. *La saliencia* no se infiere con cada presentación de cada estímulo de prueba, sino que puede observarse o inferirse después de que se agregan las puntuaciones de todas las respuestas

6.3 Puntuación de tres posiciones. Asigne puntuaciones *multinomiales* de tres posiciones [+1,0, -1] para cada *pareja de análisis* en cada sensor de registro. Estas puntuaciones se asignan utilizando el criterio de más grande es mejor, que utiliza diferencias perceptibles a simple vista (National Center for Credibility Assessment, 2017; Van Herk, 1990) cuando se califica manualmente, y esto puede generar criterios más objetivos cuando las puntuaciones se asignan por medio de un algoritmo mecánico o automatizado. Se asignan puntuaciones positivas (+) a la *pareja de análisis* cuando hay un mayor cambio en la actividad fisiológica en respuesta a la CQ, y se asignan puntuaciones negativas (-) cuando se produce un mayor cambio en la actividad fisiológica en respuesta a la RQ. Se asignan puntajes de cero (0) cuando no hay diferencias perceptibles en la magnitud de cambio en la actividad fisiológica en respuesta a la RQ y la CQ.

6.4 Puntuaciones ponderadas del EDA (duplicadas) . Duplicar todos los puntajes del EDA a [+2, 0 o -2]. Esto fue demostrado por primera vez por Krapohl y McManus (1999), y se realiza porque

los estudios publicados han demostrado que los datos de la actividad electrodérmica están más fuertemente asociados con las diferencias en las clasificaciones PDD [Véase Nelson (2017c) para una revisión bibliográfica de coeficientes de correlación para las señales poligráficas]. Los puntajes de la respiración, cardiovascular, y la actividad vasomotora son [+1, 0 o -1].

6.5 Puntuaciones ausentes. No asigne una puntuación *multinomial* ESS-M cuando no exista una RQ o una CQ con valor de extracción de características en una *pareja de análisis*. Esto puede ocurrir cuando no hay un inicio de la respuesta durante la ROW para la RQ o la CQ, o cuando hay un *artefacto de* comportamiento/movimiento, estímulo externo u otro error. La hoja de puntuación puede ser codificada para indicar el valor ausente (por ejemplo, NS, \\, sin puntuación, u otra indicación). Cuando esto ocurre el valor numérico resultante es igual a cero (0) porque el valor en las puntuaciones subtotales y de gran total ausente es de cero (0).

6.6 Hoja de puntuación. Registre las puntuaciones ESS-M en una hoja física o electrónica que esté asociada de forma confiable a cada examen. La hoja de puntuación debe diseñarse de manera que registre todas las puntuaciones de los sensores para todos las *parejas de análisis*. [Véase en el apéndice B para ver un ejemplo de hoja de puntuación manual]. La hoja de puntuación debe incluir documentación sobre la regla de decisión y el resultado categórico.

6.7 Una sola puntuación respiratoria para ambos sensores de respiración. Combine las puntuaciones numéricas para los sensores de respiración torácica y abdominal mediante el procedimiento descrito por Nelson y Krapohl (2017). [La hoja de puntuación manual del apéndice B, celdas de introducción de datos para las puntuaciones de los sensores torácicos y abdominales y la puntuación combinada para cada presentación de cada RQ].

6.7.1 Valores de signo opuesto. Si ambas puntuaciones de respiración son distintas a cero y tienen signos opuestos, la puntuación combinada es cero (0).

6.7.2 Valores de signo similar. Si los dos puntajes en la respiración son iguales, la puntuación combinada es igual a cualquiera de los dos puntajes.

6.7.2.a Ambas puntuaciones de respiración son cero (0). Si ambas puntuaciones de respiración son cero (0), la puntuación combinada es cero (0).

6.7.2.b Ambas puntuaciones de respiración son positivas (+1). Si ambas puntuaciones de respiración son positivas (+1), la puntuación combinada es positiva (+1).

6.7.2.c Ambas puntuaciones de respiración son negativas (-1). Si ambas puntuaciones de respiración son negativas (-1), la puntuación combinada es negativa (-1).

6.7.3. Una puntuación de respiración distinta de cero. Cuando una de las puntuaciones de respiración es distinta de cero y la otra puntuación es cero (0), la puntuación combinada se toma de la puntuación distinta a cero.

6.7.3.a. Una puntuación positiva (+1). Si la puntuación de una respiración es cero (0) y la otra es positiva (+1), la puntuación combinada es positiva (+1).

6.7.3.b. Una puntuación negativa (-1). Si la puntuación de una respiración es cero (0) y la otra es negativa (-1), la puntuación combinada es negativa (-1).

6.8 Reducción de datos mediante suma. Agregue las puntuaciones *multinomiales* de cada *pareja de análisis* (cada RQ) mediante una suma. [Puede utilizarse la hoja de puntuación manual que figura en el Apéndice B para registrar las puntuaciones *multinomiales* de cada sensor. También puede utilizarse para registrar las puntuaciones subtotales para cada RQ, y la puntuación de gran total para los exámenes de asunto único].

6.8.1 Puntuaciones subtotales. Sumar las puntuaciones *multinomiales* de todos los sensores y en todas las presentaciones de cada RQ.

6.8.2 Puntuación de gran total. Sume las puntuaciones de los subtotales de cada RQ para obtener lapuntuación de gran total.

7. Función de verosimilitud / modelo de referencia.

Obtenga las posibilidades a posteriori (y probabilidades a posteriori) de engaño o veracidad utilizando las distribuciones de referencia *multinomiales* que figuran en la Tabla 1, Tabla 2 y Tabla 3. Véase el apéndice C-1, C-2, y C-3 para obtener información más completa sobre las distribuciones multinomiales de referencia, a partir de las cuales se obtienen los puntajes del ESS-M. [Véase también Nelson (2017a) para un conjunto íntegro de tablas *multinomiales* para el ESS-M]. Las tablas de referencia son una forma conveniente de *función de verosimilitud* para el análisis de datos, la investigación y la práctica de campo. Las tablas de referencia ESS-M pueden utilizarse con todo el conjunto de sensores de registro validados. Las tablas ESS-M se calculan bajo una igualdad a priori (posibilidades = 1 a 1) para el engaño o veracidad con un alfa = 0.05 utilizando el conjunto estándar de sensores de registro PDD: respiratorio, cardiovascular y electrodérmico. Estas tablas pueden utilizarse con o sin el sensor vasomotor opcional. El uso de las tablas de referencia comienza con la selección de la tabla correcta para los datos del examen. La tabla 1 y la tabla 2 corresponden a exámenes de asunto único (exámenes de diagnóstico y exámenes exploratorios de un solo tema). La Tabla 3 es para exámenes *exploratorios* de asuntos múltiples.

Tabla 1. Puntuaciones ESS-M de gran total. (probabilidades posteriores y límite inferior de 1-α) intervalo de credibilidad.**

puntuación*	Posibilidades (probabilidad)†	LL05‡
-20	>100 (.992)	17
-19	87 (.989)	15
-18	63 (.984)	14
-17	47 (.979)	13
-16	35 (.972)	11
-15	26 (.963)	9.3
-14	20 (.953)	8.1
-13	15 (.939)	6.8
-12	12 (.924)	5.7
-11	9.5 (.905)	4.9
-10	7.5 (.883)	4.1
-9	6.0 (.858)	3.4
-8	4.9 (.829)	2.9
-7	3.9 (.797)	2.4
-6	3.2 (.762)	2.0
-5	2.6 (.723)	1.7
-4	2.2 (.682)	1.4
-3	1.8 (.639)	1.2
-2	1.5 (.594)	<1
-1	1.2 (.547)	<1
0	1.0 (.500)	<1
1	1.2 (.547)	<1
2	1.5 (.594)	<1
3	1.8 (.639)	1.2
4	2.2 (.682)	1.4
5	2.6 (.723)	1.7
6	3.20 (.762)	2.0
7	3.9 (.797)	2.4
8	4.9 (.829)	2.9
9	6.0 (.858)	3.4
10	7.5 (.883)	4.1
11	9.5 (.905)	4.9
12	12 (.924)	5.7
13	15 (.939)	6.8
14	20 (.953)	8.1
15	26 (.963)	9.3
16	35 (.972)	11
17	47 (.979)	13
18	63 (.984)	14
19	87 (.989)	15
20	>100 (.992)	17

* Se omiten las puntuaciones extremas.
† Probabilidades posteriores de obtener la puntuación observada bajo la hipótesis nula a la teoría del CQT (área de probabilidad bajo la curva).
‡ Límite inferior del intervalo de credibilidad del 95% (intervalo de confianza bayesiano). Se utiliza para calcular puntuaciones de corte numéricas.
** Reimpreso de Nelson, Handler, Coffey, Prado y Blalock (2019).

Tabla 2. Puntajes ESS-M subtotales para exámenes de un solo tema: probabilidades posteriores (probabilidad) y límite inferior del intervalo de credibilidad del 95%.*

puntuación	Posibilidades234RQs♣	LL05♠
-15	<10 (.920)	3.3
-14	8.8 (.898)	2.8
-13	6.9 (.873)	2.4
-12	5.52 (.847)	2.1
-11	4.5 (.818)	1.8
-10	3.73 (.789)	1.6
-9	3.13 (.758)	1.4
-8	2.67 (.728)	1.2
-7	2.3 (.697)	1.1
-6	2.01 (.668)	<1
-5	1.76 (.638)	<1
-4	1.56 (.609)	0.74
-3	1.39 (.582)	0.74
-2	1.24 (.554)	0.74
-1	1.11 (.526)	0.74
0	1	0.74

♣ Probabilidades posteriores de obtener la puntuación observada bajo la hipótesis nula a la teoría del CQT (área de probabilidad bajo la curva). Se calcula con corrección estadística de los efectos de error acumulativo cuando se utilizan puntuaciones subtotales de engaño para exámenes de un solo tema.

♠ Límite inferior del intervalo de credibilidad del 95% (intervalo de confianza bayesiano). Se utiliza para calcular puntuaciones de corte numéricas.

* Reimpreso de Nelson, Handler, Coffey, Prado y Blalock (2019).

Tabla 3. Puntuaciones ESS-M subtotales para exámenes de asuntos múltiples: probabilidades posteriores (probabilidad) y límite inferior del intervalo de credibilidad del 95%.*

Puntuación	Probabilidad	Posibilidades234RQs♦	LL05‡	234LL05♥
-14	>500 (.999)		7.6	
-13	328 (.997)		7.3	
-12	168 (.994)		6.8	
-11	91 (.989)		6.1	
-10	52 (.981)		5.2	
-9	31 (.968)		4.8	
-8	19 (.950)		4.1	
-7	12 (.924)		3.3	
-6	8.1 (.890)		2.7	
-5	5.5 (.845)		2.1	
-4	3.8 (.791)		1.6	
-3	2.68 (.728)		1.2	
-2	1.9 (.656)		<1	
-1	1.4 (.580)		<1	
0	1	1	<1	<1
1		2.6 (.725)		1.2
2		7.0 (.875)		2.5
3		19 (.95)		4.1
4		55 (.982)		5.3
5		163 (.994)		6.8
6		>500 (.998)		7.5

♦ Cuotas posteriores calculadas con corrección estadística por efectos de error acumulativo.
‡ Límite inferior del intervalo de credibilidad del 95% (intervalo de confianza bayesiano). Se utiliza para calcular puntuaciones de corte numéricas. Calculado *sin* corrección estadística para la puntuación engañosa de los subtotales con exámenes de preguntas múltiples.
♥ Límite inferior del límite inferior del 95% del intervalo de credibilidad (intervalo de confianza bayesiano). Calculado *con* Corrección estadística de las puntuaciones de subtotal de veraces con exámenes de preguntas múltiples.
* Reimpreso de Nelson, Handler, Coffey, Prado y Blalock (2019).

7.1 Obtener las posibilidades posteriores de engaño y veracidad para exámenes de asunto único. Se pueden obtener las posibilidades posteriores de engaño y veracidad para exámenes de asunto único de las Tablas 1 y 2 (también de los apéndices C-1 y C-2), siguiendo el procedimiento descrito por Nelson y Rider (2018).

7.1.1 Posibilidades posteriores de engaño para exámenes de asunto único. Los exámenes PDD de asunto único pueden clasificarse como engaño utilizando la puntuación de gran total o la puntuación subtotal en la RQ más baja.

7.1.1.a Posibilidades posteriores de engaño utilizando la puntuación de gran total para exámenes de asunto único. Utilizando la información de la Tabla 1 (también Apéndice C-1), localice la fila con la puntuación de gran total en la columna de *puntuación* del lado izquierdo de la tabla. A continuación, obtenga las *posibilidades posteriores* (y probabilidad posterior) de engaño de la misma fila de la tabla en la columna de *posibilidades*.

7.1.1.b Posibilidades posteriores de engaño utilizando la puntuación subtotal para exámenes de asunto único. Utilizando la información de la Tabla 2 (también Apéndice C-2), localice la fila con la puntuación subtotal de la RQ más baja en la columna de *puntuación* en la parte izquierda de la tabla. A continuación, obtenga las *posibilidades posteriores* (y probabilidad posterior) de engaño en la misma fila de la tabla en la Columna *Odds234RQs (Posibilidades234RQs)*. Estas posibilidades se calculan con una corrección estadística por *error acumulativo* (para reducir el aumento potencial de falsos positivos) cuando se utilizan puntuaciones subtotales en exámenes de asunto único. Note que la existencia múltiples subtotales en RQs *no independientes* crea la posibilidad de un efecto de *error acumulativo*, para el cálculo sólo se utiliza la puntuación subtotal más baja.

7.1.2 Posibilidades posteriores de veracidad en exámenes de asunto único. En los exámenes PDD de asunto único, las clasificaciones de veracidad se realizan utilizando la puntuación de gran total. Utilizando la información de la Tabla 1 (también Apéndice C-1), localice la fila del gran total en la columna de *puntuación* de la parte izquierda de la tabla. A continuación, obtenga las *posibilidades posteriores* (y probabilidad posterior) de veracidad en la misma fila de la tabla en la columna de *posibilidades*.

7.2 Obtener las posibilidades posteriores de engaño o veracidad para los exámenes de asuntos múltiples. Se pueden obtener las posibilidades posteriores de engaño y veracidad en los exámenes asuntos múltiples en la Tabla 3 (también Apéndice C-3), utilizando el procedimiento descrito por Nelson y Rider (2018).

7.2.1 Posibilidades posteriores de engaño para exámenes asuntos múltiples. Utilizando la información de la Tabla 3 (también Apéndice C-3), localice la fila con

la puntuación subtotal de la RQ más baja en la columna de puntuación de la parte izquierda de la tabla. A continuación, obtenga las *posibilidades posteriores* de engaño de la misma fila de la tabla utilizando la columna de *posibilidades*. Para evitar una pérdida involuntaria de sensibilidad en las pruebas con RQ independientes para exámenes de asuntos múltiples, las posibilidades posteriores en clasificaciones de engaño se calculan sin corrección para los efectos de *error acumulativo*.

7.2.2 Posibilidades posteriores de veracidad en exámenes de asuntos múltiples. Utilizando la información de la Tabla 3 (también Apéndice C-3), localice la fila de la RQ con la puntuación subtotal más baja en la columna de puntuación de la parte izquierda de la tabla. A continuación, obtenga las *posibilidades posteriores* de veracidad en la misma fila de la tabla utilizando la columna *Odds234RQs (Posibilidades234RQs)*. Para evitar una pérdida involuntaria de especificidad en la prueba con RQs *independientes* en los exámenes de asuntos múltiples, las posibilidades posteriores de las clasificaciones de veracidad se calculan con corrección de los efectos de *error acumulativo*.

7.3 Puntajes de corte numéricos. Un puntaje de corte numérico es la forma más sencilla de una *función de verosimilitud*, calculado o seleccionado para lograr un *tamaño del efecto* o un nivel de significancia estadística. Los puntajes de corte numéricos permiten a los profesionales de campo realizar clasificaciones categóricas basadas en información probabilística sin considerar directamente los efectos de probabilidad. Los puntajes de corte ESS-M se calculan para garantizar que el límite inferior del *intervalo de credibilidad* bayesiano 1 - α para las posibilidades posteriores será igual o superior a las *posibilidades a priori*. Vea la Tabla 4 para las puntuaciones de corte numéricas de ESS-M que pueden generalizarse para los CQTs de asuntos únicos y múltiples con tres a cinco presentaciones de dos a cuatro RQs utilizando el conjunto estándar de sensores de registro PDD, que incluyen los de respiración, electrodérmicos, cardiovascular, y los sensores vasomotores opcionales.

Tabla 4. Cortes de Calificación ESS-M para 3 a 5 presentaciones de 2, 3 o 4 RQ con y sin el sensor vasomotor.*

	Cortes de gran total		Cortes por subtotales	
	Veraz	Engañoso	Veraz	Engañoso
Problema único (diagnóstico o exploratorio de un solo asunto)	+3	-3	-	(-7)‡
Exámenes exploratorios de temas múltiples	-	-	(+1)‡	-3

‡ Las puntuaciones de corte entre paréntesis se calculan con corrección estadística para la multiplicidad [véase Nelson (2014) para obtener información más detallada sobre los efectos del error acumulativo y las correcciones estadísticas).
* Reimpreso de Nelson y Rider (2018).

7.3.1 Cálculo de puntuaciones de corte para exámenes de asunto único. El cálculo de las puntuaciones de corte para los exámenes de asunto único requiere de la Tabla 1 y la Tabla 2 (también Apéndice C-1 y Apéndice C-2). La puntuación de corte es la puntuación mínima para el que el límite más bajo del *intervalo de credibilidad* Bayesiano 1 - α de las posibilidades posteriores es igual o superior a las *posibilidades previas*.

7.3.1.a Puntuación de gran total para las clasificaciones de engaño en exámenes de asunto único. Utilizando la información de la Tabla 1 (también Apéndice C-1), localice la fila de la columna *LL05* en la que el límite inferior de las posibilidades posteriores de engaño supera la igualdad a priori (1 a 1). A continuación, obtenga el puntaje de corte de la columna de *puntuación* utilizando la misma fila en la tabla.

7.3.1.b Puntaje de corte subtotal para clasificaciones de engaño en exámenes de asunto único. Utilizando la información de la Tabla 2 (también Apéndice C-2), localice la fila de la columna *LL05* para el que el límite inferior de las posibilidades posteriores de engaño supera la igualdad a priori (1 a 1). A continuación, obtenga el puntaje de corte la de la columna de *puntuación* utilizando la misma fila de la tabla. Ya que en un examen de asunto único se realizan múltiples intentos probabilísticos, la puntuación de corte por subtotal para las clasificaciones de engaño de exámenes de asunto único se calcula con una corrección estadística del efecto de error acumulativo.

7.3.1.c Puntaje de corte de gran total para clasificaciones veraces de exámenes de asunto único. Utilizando la información de la Tabla 1 (también Apéndice C-1), localice la fila de la columna *LL05* en el que el límite inferior de la posibilidad posterior de veracidad supere la igualdad a priori (1 a 1). A continuación, obtenga el puntaje de corte la de la columna de *puntuación* utilizando la misma fila en la tabla.

7.3.2 Cálculo de puntuaciones de corte para exámenes de asuntos múltiples. El cálculo de las puntuaciones de corte numéricas para los exámenes de asuntos múltiples requiere de la Tabla 3 (también Apéndice C- 3). El puntaje de corte es la puntuación mínima en la que el límite inferior del *intervalo de credibilidad* Bayesiano 1 – α de las posibilidades posteriores será igual o superior a las *posibilidades a priori*.

7.3.2.a Puntaje de corte subtotal para clasificaciones de veracidad de exámenes de asunto múltiple. Utilizando la información de la Tabla 3 (también Apéndice C-3), localice en las columnas *234LL05.* La fila cuyo límite inferior de las posibilidades posteriores de veracidad superen la igualdad a priori (1 a 1). A continuación, obtenga el puntaje de corte de la columna de *puntuación* utilizando la misma fila en la tabla. Dado que las RQ de los exámenes de asuntos múltiples se puntúan e interpretan bajo un supuesto de independencia, los puntajes de corte subtotales para la clasificación de engaño de los exámenes de asuntos múltiples se calculan sin corrección estadística.

7.3.2.b Puntaje de corte subtotal para clasificaciones de engaño en exámenes de asuntos múltiples. Utilizando la información de la Tabla 3 (también Apéndice C-3), localice la fila en las columnas *LL05.* En la que el límite inferior de las posibilidades posteriores de engaño supere la igualdad a

priori (1 a 1). Entonces obtenga el puntaje de corte en la columna de *puntuación* utilizando la misma fila en la tabla.

8. Interpretación de los resultados de prueba.

Cuando se utiliza el ESS-M para la interpretación de los resultados de las pruebas PDD, se llevan a cabo varias funciones que incluyen el uso de reglas de decisión procesales para la clasificación categórica del resultado de la prueba como indicativo de engaño o de veracidad, análisis de los resultados de las preguntas individuales de los exámenes de asunto único o asuntos múltiples, la documentación del resultado analítico y los parámetros para los resultados categóricos, y la discusión y presentación de los resultados estadísticos y categóricos en lenguaje natural.

8.1 Reglas de decisión. En los exámenes poligráficos, las reglas de decisión describen el procedimiento para llegar a conclusiones categóricas con las puntuaciones de gran total y subtotales. Una regla de decisión muy común dentro del paradigma estadístico *frecuentista*, es que un resultado observado es *estadísticamente significativo* cuando el *valor-p* (la *probabilidad* de obtener los datos observados bajo una hipótesis especificada) es menor o igual que el límite *alfa* ($p \leq a = sig$). Para los puntajes ESS-M en los datos de un CQT, las reglas de decisión se basan en el límite inferior del *intervalo de credibilidad* 1 – α para las posibilidades Bayesianas posteriores de engaño o veracidad. Comúnmente el *alfa* se establece a 0.05, y la amplitud del *intervalo de creíble* Bayesiano se calcula a 1 - α × 100% = 95%. Se considera que el resultado de una prueba es *estadísticamente significativo* cuando el límite inferior del *intervalo de credibilidad* del 95% excede las *posibilidades a priori*. Para simplificar la práctica de campo, este procedimiento puede reducirse a comparar las puntuaciones de gran total y subtotales contra las puntuaciones numéricas de corte calculadas para el límite inferior de las posibilidades posteriores (mostradas en la Tabla 4). Las clasificaciones PDD pueden ser sencillas e intuitivas si sólo se utiliza el puntaje de gran total, pero requieren una lógica de procedimiento explícita cuando se utilizan valores múltiples para realizar una clasificación categórica única del resultado global de la prueba PDD. Las diferentes reglas de decisión son útiles si se aplican para los diferentes tipos de exámenes PDD, como son los exámenes de asunto único y múltiple. Estas reglas incluyen la regla del gran total, de puntuación subtotal, de dos etapas entre otras. [Véase Nelson (2018a) para un estudio de reglas de decisión poligráficas].

8.1.1 Regla del gran total. La regla del gran total (GTR) ha sido descrita por varios autores (Bell et al., 1999; Kircher y Raskin, 1988; Senter, 2003; Weaver,1980), y consiste en comparar la puntuación de gran total observada contra los puntajes de corte numéricos. En la Tabla 4 se muestran las puntuaciones de corte para los exámenes de asunto único. De manera más formal, pueden obtenerse las posibilidades posteriores y el límite inferior del *intervalo de credibilidad* del 95% al localizar el que le corresponde a la puntuación de gran total observada en la columna izquierda de la Tabla 1. Esto indica una probabilidad del 95% de que cualquier prueba similar - en el mismo individuo, con el mismo asunto-objetivo, utilizando las mismas preguntas de examen, instrumentación y sensores similares, y el mismo método de análisis - produciría un resultado igualmente indicativo de

engaño. Tradicionalmente, el GTR es utilizado en exámenes poligráficos de *diagnóstico* de asunto único (evento específico), y también se puede utilizar con polígrafos *exploratorios* de asunto único.

8.1.1.a Clasificación de engaño utilizando el GTR. Si el puntaje de gran total general es igual o inferior a la puntuación de corte numérico para clasificaciones de engaño (-3), clasifique el resultado de la prueba como engaño (DI para exámenes de asunto único, o SR para exámenes *exploratorios*) como se muestra en la Tabla 4. Alternativamente, la clasificación de engaño puede realizarse utilizando la Tabla 1 localizando el renglón de la puntuación observada por gran total en la columna de *puntuación* para determinar si el límite inferior del *intervalo de credibilidad* del 95% para las posibilidades posteriores de engaño, mostradas en la columna *LL05* de la misma fila, supera las posibilidades previas de engaño (1 a 1).

8.1.1.b Clasificación de veracidad utilizando el GTR. Si la puntuación de gran total es igual o superior al puntaje de corte numérico para clasificaciones veraces (+3), clasificar el resultado de una prueba de asunto único como veraz (NDI para los exámenes de *diagnóstico*, o NSR para exámenes *exploratorios)* como se muestra en la Tabla 4. Alternativamente, se puede hacer la clasificación veraz localizando la fila correspondiente a la puntuación observada de gran total general en la columna de *puntuación* de la Tabla 1 y determinar si el límite inferior del *intervalo de credibilidad* del 95% para las posibilidades posteriores de veracidad, mostradas en la misma fila en la columna *LL05*, supera las *posibilidades previas* de veracidad (1 a 1).

8.1.1.c Resultado inconcluso (no opinión) utilizando el GTR. Clasificar el resultado de una prueba de asunto único como *inconcluso*, si la puntuación de gran total no alcanza la puntuación numérica ni para engaño ni para veracidad. En este caso, el límite inferior del *intervalo de creíble* del 95% no supera las *posibilidades previas.*

8.1.2 Regla de puntuación de subtotal. Cuando se utiliza la Regla de Puntuación de Subtotal (SSR; Department of Defense, 2006a, Capps C Ansley 1992; Natinal Center for Credibility Assessment, 2017; Senter Waller C Krapohl; 2008), el resultado del examen se determina únicamente por las puntuaciones subtotales de las RQ, sin tener en cuenta la puntuación de gran total. El SSR se utiliza tradicionalmente con polígrafos de asuntos múltiples exploratorios.

8.1.2.a Clasificación de engaño utilizando el SSR. Si la puntuación en la RQ más baja tiene un subtotal que es igual o inferior a la puntuación de corte numérica para clasificaciones de engaño [-3], clasifique los resultados de exámenes poligráficos de asuntos múltiples como engaño (SR),como se muestra en la Tabla 4. Alternativamente, puede hacerse una clasificación de engaño localizando la fila para la puntuación subtotal más baja, en la

columna de *puntuación* de la Tabla 3 y a continuación determinar si el límite inferior del *intervalo de credibilidad* del 95% para las posibilidades posteriores de engaño mostradas en la misma fila en la columna *LL05* supera las posibilidades previas de engaño (1 a 1).

8.1.2.b Clasificación de veracidad utilizando el SSR. Si la puntuación en la RQ más baja por subtotal es igual o superior a la puntuación de corte numérico de veracidad [+1], clasifique los resultados de exámenes poligráficos de asuntos múltiples como veraces (NSR), como se muestra en la Tabla 4. Alternativamente, puede hacerse una clasificación localizando la fila del subtotal más bajo en la columna de *puntuación* de la Tabla 3 y, a continuación, determinar si el límite inferior del *intervalo de credibilidad* del 95% para las posibilidades posteriores de veracidad mostrada en la columna *234LL05* supera las *posibilidades previas* de veracidad (1 a 1). Note que el uso de puntajes subtotales de veracidad para clasificaciones de polígrafos de asunto múltiple produce efectos de *error acumulativo* y por esta razón las posibilidades posteriores de veracidad, el *intervalo de credibilidad*, y la puntuación numérica de corte para clasificaciones veraces utilizando las puntuaciones subtotales se calcula con una *corrección estadística de multiplicidad*.

8.1.2.c Resultado inconcluso (no opinión) utilizando el SSR. Clasifique los resultados de polígrafos asunto múltiple como *inconclusos* si la puntuación subtotal en la RQ más baja no alcanza el puntaje de corte numérico ni para engaño ni para veracidad. En esta situación, el límite inferior del *intervalo de credibilidad* al 95% no supera las posibilidades a priori, y las posibilidades posteriores no son *estadísticamente significativas* al nivel del alfa deseado.

8.1.3 Regla en dos etapas. La regla de dos etapas (TSR) se utiliza para los exámenes PDD de asunto único e incluye tanto la puntuación de gran total como la puntuación subtotal más baja (Senter, 2003; Senter C Dollins, 2003; Handler, Nelson C Blalock, 2008; Krapohl, 2005; Krapohl C Cushman, 2006; Nelson et al., 2011). La fase 1 del TSR es similar al GTR. La fase 2 del TSR es similar al del SSR. Obsérvese que el uso de puntuaciones subtotales para la clasificación de engaño en polígrafos de asunto único produce el efecto de *error acumulativo*, y por esta razón, las posibilidades posteriores, el *intervalo de credibilidad* y la puntuación numérica para clasificaciones de engaño mediante el subtotal se calculan con una *corrección estadística* por *multiplicidad*.

8.1.3.a Etapa 1 del TSR. La etapa 1 del TSR es similar a la del GTR. Clasifica los resultados de la prueba como indicativos de engaño o de veracidad [DI/SR o NDI/NSR] si la puntuación de gran total general es igual o superior a cualquiera de los puntajes de corte numéricas de gran total [+3 o -3] como se muestra en la Tabla 4. Alternativamente, localice la fila correspondiente a la puntuación de gran total de la Tabla 1 y obtenga el límite inferior del

intervalo de credibilidad al 95% para las posibilidades posteriores de engaño o veracidad de la misma fila en la columna *LL05*. El resultado de una prueba puede clasificarse como indicativo de engaño o de veracidad si el límite inferior del *intervalo de credibilidad* Bayesiano supera las posibilidades a priori de engaño o veracidad (1 a 1).

8.1.3.b Etapa 2 del TSR. La Etapa 2 del TSR puede pensarse como similar al SSR, y se utiliza cuando la puntuación de gran total no es concluyente. Clasificar el resultado de la prueba como engaño (DI/SR) si la puntuación subtotal más baja es igual o inferior a la puntuación numérica [-7] para clasificaciones de engaño basadas en las puntuaciones subtotales de los exámenes de asunto único como que figura en la Tabla 4. Alternativamente, puede obtenerse la misma clasificación localizando la puntuación subtotal en la RQ más baja en la columna de puntuación de la Tabla 2 y, a continuación, obtener el límite inferior del *intervalo de credibilidad* al 95% de la misma fila en la columna *LL05*. Un resultado puede clasificarse como engaño si el límite inferior *intervalo de credibilidad* Bayesiano supera las *posibilidades previas* de engaño (1 a 1). Una diferencia importante entre la Etapa 2 del TSR y el SSR es que en esta fase solo se realizan clasificaciones de engaño. En esta fase no se realizan clasificaciones de veracidad.

8.1.3.c Ambigüedad en los resultados totales y subtotales. Cuando se utiliza el TSR, en raras ocasiones una puntuación de gran total podría respaldar una clasificación de veracidad al tiempo que la puntuación subtotal más baja apoya una clasificación de engaño. En estos casos, se permite clasificar el resultado de la prueba como engaño debido a la puntuación subtotal más baja, independientemente de la puntuación de gran total. Este enfoque se justifica por la presencia de variabilidad inesperada o excesiva en los datos de la prueba, lo que daña la confiabilidad de las conclusiones basadas en la puntuación de gran total. Por el contrario, la puntuación subtotal más baja sigue siendo estadísticamente significativa para una clasificación de engaño utilizando un puntaje de corte subtotal corregido estadísticamente [-7]. Por procedimiento, este enfoque del TSR puede realizarse de la siguiente manera 1) clasificar el resultado de la prueba como engaño utilizando la puntuación de corte por gran total [-3] o la puntuación de corte por subtotal [-7], y 2) clasificar el resultado de la prueba como veraz utilizando la puntuación de corte por gran total [+3].

8.2 Análisis de los resultados y de las preguntas de prueba. Los examinadores PDD de campo podrían tener que dar una explicación o reporte de los resultados de las pruebas para cada una de las RQs, además del resultado global de la prueba. Para analizar los resultados de cada una de las RQ de exámenes poligráficos de asunto único y múltiple se hace uso distintos procedimientos (Nelson, Blalock & Handler 2019).

8.2.1 Análisis de los resultados de las preguntas de asunto único. Para los exámenes de asunto único, todas las preguntas heredan el mismo resultado categórico de la prueba en su conjunto.

8.2.1.a Resultado DI/SR en la prueba. Todas las RQ se clasifican y reportan como DI/SR.

8.2.1.b Resultado NDI/NSR en la prueba. Todas las RQ se clasifican y reportan como NDI/NSR

8.2.1.c Resultado INC/NO en la prueba. Todas las RQ se clasifican y reportan como INC/NO

8.2.2 Análisis de los resultados de las preguntas de asuntos múltiples. En el caso de los exámenes de asuntos múltiples, el resultado global se determina por el resultado categórico producido por la RQ con la puntuación subtotal más baja. Los resultados categóricos para las otras RQs se obtienen mediante la comparación del subtotal de cada RQ contra los puntajes de corte subtotales de los exámenes de asuntos múltiples [+1, -3]. Es de destacar que no está permitido informar resultados tanto de engaño como veracidad en el mismo examen.

8.2.2.a El subtotal más bajo es NSR. Clasifique el resultado global de la prueba como NSR cuando la RQ con la puntuación subtotal más baja es *estadísticamente significativa* para veracidad. Cuando esto ocurre, todas las demás puntuaciones subtotales en las RQ también son estadísticamente significativas para veracidad y se clasifican y reportan como NSR.

8.2.2.b El subtotal más bajo es SR. Clasifique el resultado global de la prueba como SR si la puntuación subtotal más baja es *estadísticamente significativa* para el engaño. Compare cada uno de los otros subtotales de las RQ contra el puntaje de corte subtotal para clasificaciones de engaño [-3], y clasifique la puntuación de la RQ como SR si la puntuación subtotal cumple con el corte mínimo para SR. Clasifique la puntuación de cada RQ como INC si no cumple con la puntuación mínima para una clasificación de engaño. No se permite informar resultados de engaño y veracidad en el mismo examen.

8.3 Documentación y reporte de la información y resultados del ESS-M. La interpretación de los resultados de las pruebas ESS-M requiere de una comunicación clara y eficaz de los resultados. Los resultados de las pruebas deben ser comprensibles y útiles para los profesionales que remiten a los evaluados. Además, el análisis de los resultados analíticos debe ser reproducibles por otros expertos en PDD. La reproducibilidad de los resultados analíticos se ha convertido en un tema importante en la ciencia computacional (Peng, 2011), y requiere que se documente y reporte el proceso de análisis que incluye todos los pasos y parámetros de entrada (por ejemplo, fórmulas, priors, alfas, procesamiento de señales, transformaciones, correcciones estadísticas, etc.) que fueron utilizados para

completar el análisis. Véase un ejemplo en el Apéndice D de una hoja de trabajo para registrar los parámetros de análisis y los resultados. La documentación debe incluir el ID del examen, fecha del análisis y nombre del analista, además de la siguiente información:

8.3.1 Puntuaciones numéricas. Proporcione los puntajes numéricos a partir de los cuales se obtuvo el resultado de la prueba, incluyendo:

8.3.1.a Puntaje de las RQ (RQ con subtotal más bajo). Proporcione el puntaje subtotal de cada RQ. Esto ayuda a comprender las respuestas fisiológicas específicas de cada pregunta. Siempre que se usan las puntuaciones subtotales, se utiliza la RQ con el subtotal más bajo para determinar el resultado probabilístico y categórico de la prueba.

8.3.1.b Puntuación de gran total (exámenes de asunto único). Incluya el puntaje de gran total para exámenes de asunto único. Esta puntuación integra todas las respuestas en una medición única para facilitar la interpretación.

8.3.2 Puntajes de corte. Documente los puntajes de corte numéricos utilizados en el análisis. Los puntajes de corte son los umbrales utilizados para clasificar los resultados de las pruebas como indicativos de engaño o de veracidad si la fuerza de la información posterior ha alcanzado el nivel requerido de significancia estadística.

8.3.3 Límites alfa. Reporte los límites *alfa* establecidos para el *intervalo de credibilidad* Bayesiano. Los límites *alfa* indican el nivel de significancia estadística utilizado en el análisis, y también influyen en los puntajes de corte numéricos.

8.3.4 Uso de correcciones estadísticas. Reporte cualquier corrección estadística aplicada durante el análisis. Esto garantiza la transparencia y ayuda a comprender cómo se manejaron las comparaciones múltiples. *La corrección estadística* se utiliza cuando se usan las puntuaciones subtotales para clasificar un resultado de prueba como indicativo de engaño o de veracidad, y no son necesarias cuando sólo se utiliza el puntaje de gran total.

8.3.4.a Corrección estadística para exámenes de asunto único. Cuando se utiliza el TSR, se aplica una *corrección estadística* al cálculo de las posibilidades posteriores de la RQ con el subtotal en más bajo, cuando esta se utiliza como base para la clasificación de engaño. En otras palabras, el puntaje de corte subtotal de ESS-M para clasificaciones de engaño para los exámenes de asunto único se calculan con una *corrección estadística* para los efectos del *error acumulativo*. El objetivo de esta corrección es mitigar el aumento potencial de errores falso positivo.

8.3.4.b Corrección estadística para los exámenes de asunto múltiple. Cuando se utiliza el SSR, se aplica una *corrección estadística* al cálculo de las posibilidades posteriores para la RQ con la puntuación subtotal más baja, cuando esta se utiliza como base para una clasificación de veracidad. Dicho de otro modo, el puntaje de corte subtotal del ESS-M para

clasificaciones veraces de exámenes de asuntos múltiples se calcula con una *corrección estadística* del *error acumulativo*. El objetivo de esto es mitigar el posible aumento de resultados inconclusos.

8.3.5 Regla de decisión. Especifique la regla de decisión aplicada, ya sea Regla de Gran Total (GTR*), Regla de Puntaje Subtotal (SSR*) o Regla de Dos Etapas (TSR*). Esta explica las bases de la clasificación final de los resultados de la prueba.

8.3.6 Resultado categórico. Los resultados categóricos de las pruebas siempre se basan en resultados probabilísticos, pero son más convenientes ya que reducen la ambigüedad que podría resultar de las diferentes interpretaciones de la información probabilística.

8.3.6.a Resultado global de la prueba. Reporte el resultado global de la prueba (por ejemplo, DI/SR o NDI/NSR). Esto proporciona una conclusión clara que se basa en el análisis.

8.3.6.b Resultados en las RQ. Analice y comunique los resultados de cada RQ, utilizando los procedimientos detallados en la Sección 8.2. Este detalle ayuda a reducir la confusión y malentendidos sobre el significado de cada RQs.

8.3.7 Clasificador estadístico. Reporte el *clasificador estadístico*, que incluye el tipo de clasificador y su valor estadístico. Los resultados del ESS-M se basan en el *Análisis Bayesiano* y en el límite inferior de las *posibilidades posteriores* de engaño o veracidad.

8.3.7.a Posibilidades posteriores (probabilidades posteriores). Documente la posibilidad o *probabilidad posterior*. Esto proporciona una medición Bayesiana de las posibilidades de engaño o veracidad.

8.3.8.b Factor de Bayes. Comunique el *Factor de Bayes*, que indica la magnitud del cambio en la fuerza de la información debido a la prueba.

8.3.8 Nivel de significancia (valor-alfa). Describa el *nivel de significancia (alfa)* utilizado en el análisis. Esto ayuda a comprender la precisión y el margen de error para el resultado categórico.

8.3.9 Intervalo de confianza (límite inferior del intervalo de credibilidad de Bayes 1-a x 100%). Incluya el límite inferior del *intervalo de credibilidad* Bayesiano. Esto proporciona un rango dentro del cual se espera que se encuentre el valor real del parámetro desconocido (probabilidad de engaño o veracidad) con un cierto nivel de confianza. También refleja la *probabilidad* de observar resultados similares de prueba.

9. Trabajo con posibilidades previas y valores alfa diferentes.

Las *posibilidades previas* y el *nivel alfa* determinarán el cálculo de las tablas de referencia ESS-M que fueron utilizadas, para establecer las *posibilidades posteriores* y el *intervalo de credibilidad* a posteriori. Para fines prácticos, ambos influyen en los puntajes de corte numéricos requeridos para clasificar un resultado de prueba como DI/SR o NDI/NSR al nivel de *significancia estadística* deseado. Las Tablas 1, 2 y 3 (también los Apéndices C-1, C1 y C-3) se proporcionan utilizando una distribución previa equitativa (1 a 1, o posibilidades equitativas de engaño y veracidad) con el límite inferior del intervalo de credibilidad bayesiano para *alfa* = 0.05 (una cola). El nivel *alfa* es una consideración tanto administrativa como práctica. Reducir el nivel *alfa* puede aumentar el nivel de confianza de que el resultado observado se reproducirá con otras muestras, pero también puede aumentar el puntaje de corte numérico. Incrementar el nivel *alfa* puede reducir los puntajes de corte numéricos y los resultados *inconclusos*, y también puede decrementar la reproductibilidad o la probabilidad de que el resultado observado se deba a una variación aleatoria, inexplicable o no controlada. El cálculo de las posibilidades posteriores con diferentes posibilidades previas también influirá en las posibilidades posteriores. Lo ideal es que las *posibilidades previas* se basen en información objetiva y en los hechos del caso. También es posible evaluar las *posibilidades posteriores* de los resultados de las pruebas PDD para un rango o distribución de posibilidades previas.

9.1 Cálculo de posibilidades posteriores mediante el Teorema de Bayes. Se pueden calcular las posibilidades posteriores de engaño o veracidad utilizando el Teorema de Bayes (Nelson, 2018b). Por conveniencia, la *posibilidad posterior* (y las probabilidades posteriores) serán las mismas que los valores en la columna *cdfConCor* en el Apéndice C, bajo una previa equivalente (1 a 1). Un cambio en las *posibilidades previas* va a cambiar las *posibilidades posteriores*. Cuando las *posibilidades previas* de engaño son bajas, las *posibilidades posteriores* por correspondencia son menores (Nelson, Handler & Thiel, 2021). Por el contrario, las *posibilidades posteriores* de engaño son mayores cuando las *posibilidades previas* son mayores. Se obtiene de manera más fácil el cálculo del Teorema de Bayes y de las posibilidades posteriores mediante un algoritmo informático automatizado. Aquí se muestra el procedimiento. Por conveniencia, la Tabla 5 muestra las *posibilidades posteriores* de engaño para un rango de *posibilidades previas* de 9:1 a 1:9.

Tabla 5. Posibilidades posteriores de culpabilidad con probabilidades previas altas y bajas de culpabilidad (posibilidades previas entre paréntesis).*

	.9 (9:1)	.8 (4:1)	.75 (3:1)	.67 (2:1)	.5 (1:1)	.33 (1:2)	.25 (1:3)	.2 (1:4)	.125 (1:7)	.1 (1:9)
Puntuaciones de Gran total	Probabilidades Predictivas Positivas Posteriores									
-24	4668	2075	1556	1037	519	259	173	130	74	58
-23	3170	1409	1057	704	352	176	117	88	50	39
-22	2184	971	728	485	243	121	81	61	35	27
-21	1527	679	509	339	170	85	57	42	24	19
-20	1084	482	361	241	120	60	40	30	17	13
-19	779	346	260	173	87	43	29	22	12	9.6
-18	567	252	189	126	63	32	21	16	9	7
-17	419	186	140	93	47	23	16	12	6.6	5.2
-16	313	139	104	69	35	17	12	8.7	5	3.9
-15	236	105	79	53	26	13	8.8	6.6	3.8	2.9
-14	181	80	60	40	20	10	6.7	5	2.9	2.2
-13	139	62	46	31	15	7.7	5.2	3.9	2.2	1.7
-12	109	48	36	24	12	6	4	3	1.7	1.3
-11	85	38	28	19	9.5	4.7	3.2	2.4	1.4	1.1
-10	68	30	23	15	7.5	3.8	2.5	1.9	1.1	0.8
-9	54	24	18	12	6	3	2	1.5	0.9	0.7
-8	44	19	15	9.7	4.8	2.4	1.6	1.2	0.7	0.5
-7	35	16	12	7.9	3.9	2	1.3	1	0.6	0.4
-6	29	13	9.6	6.4	3.2	1.6	1.1	0.8	0.5	0.4
-5	24	10	7.8	5.2	2.6	1.3	0.9	0.7	0.4	0.3
-4	19	8.6	6.4	4.3	2.1	1.1	0.7	0.5	0.3	0.2
-3	16	7.1	5.3	3.5	1.8	0.9	0.6	0.4	0.3	0.2
-2	13	5.8	4.4	2.9	1.5	0.7	0.5	0.4	0.2	0.2
-1	11	4.8	3.6	2.4	1.2	0.6	0.4	0.3	0.2	0.1
0	9	4	3	2	1	0.5	0.3	0.3	0.1	0.1

* Reimpreso de Nelson, Handler y Thiel (2021).

9.1.1 Obtenga el valor de probabilidad de la tabla de referencia ESS-M. Primero localice la fila con el puntaje de la prueba (gran total o subtotal) en la columna de *puntuación* utilizando el Apéndice C-1, Apéndice C-2, o Apéndice C-3. A continuación, obtenga el valor de la columna *cdfContCor* en la misma fila de la tabla. Este valor es la función de distribución acumulada (suma acumulativa de los valores discretos de probabilidad) después de aplicar una *corrección de continuidad* para garantizar que el valor calculado siempre supera el valor real. Cuando las probabilidades previas de engaño y veracidad no son iguales, la probabilidad posterior puede calcularse mediante el *Teorema de Bayes*.

9.1.2 Probabilidad posterior. El valor de la probabilidad posterior se calcula como: **postProb = (previa * (1 - lf) / ((previa * (1 - lf)) + ((1 - previa) * lf)), donde**

9.2.1.a previa es el valor de la probabilidad previa

9.2.1.b lf es el valor de la columna *cdfConCor en las tablas de los Apéndices C*.

9.1.3 Posibilidades posteriores de engaño. Las posibilidades posteriores de engaño se calculan como: **postOdds = postProb / (1 - postProb).**

9.1.4 Posibilidades posteriores de veracidad. Las probabilidades posteriores de veracidad se calculan como: **postOdds = 1 - (postProb / (1 - postProb)).**

9.2 Cálculo del intervalo de credibilidad mediante el método de Clopper-Pearson. El límite inferior del intervalo de credibilidad Bayesiano se utiliza para clasificar los resultados de las pruebas PDD como indicativos de engaño o de veracidad. Un resultado de prueba es *estadísticamente significativo* si el límite inferior del *intervalo de credibilidad* para las *posibilidades posteriores* supera las *posibilidades previas*. Por conveniencia, se incluyen los valores inferiores límite en las Tablas 1, 2 y 3 bajo una previa equivalente. Cuando las *posibilidades previas* de engaño y veracidad no son iguales, las Tablas 1, 2 y 3 sólo se utilizan como *función de verosimilitud,* utilizando los datos de la prueba para calcular las posibilidades posteriores. Entonces se utiliza el método de Clopper-Pearson para calcular el *intervalo de credibilidad*. El cálculo del *intervalo de credibilidad* de Clopper-Pearson se obtiene de manera más eficiente mediante el uso de un sistema informático automatizado. A continuación, se muestra el procedimiento.

9.2.1 Grados de libertad para los límites superior e inferior de la distribución F. Para calcular los límites alfa de la distribución se necesitan los grados de libertad F y se calculan con *n* igual a la puntuación ESS-M máxima posible. Las técnicas PDD descritas por la American Polygraph Association (2011) contienen de dos a cuatro RQ repetidas hasta cinco veces. La puntuación máxima posible oscila de 50 a 100, con una mediana de 75. Por esta razón, n = 75 para estos cálculos.

9.3.1.a df1Low = 2 * (n * (1 - p) + 1

9.3.1.b df2Low = 2 * n * p

9.3.1.c df1High = 2 * (n * p + 1)

9.3.1.d df2AHigh = 2 * n * (1 - p)

9.2.2 Límites alfa superior e inferior de la distribución F. Se necesitan los límites alfa de dos colas de la distribución F para calcular la probabilidad posterior.

9.3.2.a Flow = F[(1 - a/2), df1Low, df2Low]

9.3.2.b Fup = F[a/2, df1High, df2High].

9.2.3 Calcule los límites superior e inferior de la probabilidad posterior. Los límites alfa del intervalo de credibilidad para la probabilidad posterior se calculan utilizando el estadístico F.

9.2.3.a Límite inferior = (1 + Flow * ((1 - p) + 1 / n) / p)^-1

9.2.3.b Límite superior = (1 + (1 - p) / (1 / n + p) * Fup)^-1

9.2.4 Calcule el límite inferior de las probabilidades posteriores234LL05. El límite inferior de las posibilidades posteriores es el clasificador estadístico ESS-M. Las posibilidades posteriores son estadísticamente significativas cuando el límite inferior supera las posibilidades previas. El alfa (α = 0,05) en este punto es de una cola.

9.2.4.a LLO5 = Límite inferior / (1 - Límite inferior)

9.3 Cálculo de puntuaciones de corte numéricas a partir del alfa y el prior. Las tablas de referencia como las que se muestran en los Apéndices C-1, C-1 y C-3 se pueden obtener a partir de cualquier valor alfa y cualquier probabilidad previa. Es posible un gran número de distribuciones de referencia. Estos se pueden calcular manualmente, aunque a menudo se realizan por computadora. Las tablas de referencia ESS-M se utilizan para calcular valores posteriores (posibilidades y probabilidades posteriores) y puntuaciones de corte numérico. Las puntuaciones de corte numérico (la puntuación mínima de gran total y subtotal que sea estadísticamente significativa para el *alfa* y las probabilidades previas) se pueden calcular a partir de las tablas de referencia utilizando el procedimiento descrito en la sección 7.3. Para mayor comodidad, el Apéndice E-1 muestra las puntuaciones de corte para exámenes de polígrafo de asunto único bajo varias posibilidades previas y límites alfa. El Apéndice E-2 muestra el rango de puntuaciones de corte para exámenes de polígrafo de asuntos multiples.

Referencias

Abdi, H. (2007). Bonferroni and Šidak corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics.* Sage.

Abramowitz, M. & Stegun, I. A. (1972). Multinomial Coefficients. §24.1.2 [pp. 823-824] in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th printing.* National Bureau of Standards (now National Institute of Standards and Technology).

Agresti, A. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician, 52(2),* 119-126.

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph, 40(4),* 196-305. [Electronic version] Retrieved August 20, 2012, from http://www.polygraph.org/section/research-standards-apa-publications.

Backster, C. (1979). *Standardized Polygraph Examination Notepack.* [Author.]

Baron, H. (1996). Strengths and limitations of ipsative measurement". *Journal of Occupational and Organizational Psychology. 69,* 49–56.

Bayes, T. & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London, 53, 370–418.

Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph, 28(1),* 1-9.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis. (Second edition.)* New York: Springer Verlag.

Buckingham, E. (1914). On physically similar systems; illustrations of the use of dimensional equations. *Physical Review, 4(4),* 345-376.

Casella, G., & Berger, R. L. (2002). *Statistical inference (2nd ed.).* Duxbury.

Capps, M. H. & Ansley, N. (1992). Analysis of federal polygraph charts by spot and chart total. *Polygraph, 21,* 110-131.

Centers for Disease Control and Prevention (CDC). (2012). *Introduction to Epidemiology. Lesson 6: Screening and Diagnostic Tests.* Retrieved from https://www.cdc.gov/csels/dsepd/ss1978/lesson6/section1.html

Chen, C. C. & Koh, K. M. (1992). *Principles and Technique in Combinatorics.* World Scientific.

Clopper, C.; Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 26, 404–413.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1994). The Earth Is Round (p < .05). *American Psychologist, 49(12)*, 997-1003.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). A*pplied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Routledge.

Colquhoun D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science,* 1: 140216.

Department of Defense (2006). *Federal Psychophysiological Detection of Deception Examiner Handbook*. [Author.] Reprinted in Polygraph, 40(1), 2-66.

Downey, A. B. (2012). *Think Bayes. Bayesian Statistics Made Simple.* Green Tea Press.

Duttton,. D (2000). Guide for performing the objective scoring system. *Polygraph, 29(2),* 177-184.

Efron, B. (1986). Why isn't everyone a Bayesian? *American Statistician. 40(1),* 1-5.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes.* Cambridge University Press.

Fisher, R. A. (1934). *Statistical Methods for Research Workers (4th ed.).* Edinburgh: Oliver and Boyd.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtary, A., & Rubin, D. B. (2014). *Bayesian Data Analysis.* CRC Press.

Handler, M., & Nelson, R. (2007). Polygraph terms for the 21st century. *Polygraph, 36(3)*, 157-164.

Handler, M., Nelson, R. & and Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph, 37(2)*, 100-111.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. Psychological Bulletin, 74(3), 167-184.

James, S. H., & Nordby, J. J. (Eds.). (2009). *Forensic science: An introduction to scientific and investigative techniques.* CRC Press.

Jarosz, A. J. & Wiley, J. (2014) What are the odds? a practical guide to computing and reporting Bayes Factors. *Journal of Problem Solving, 7.*

Jaynes, E. T. (1976). Confidence intervals vs Bayesian intervals. In W. L. Harper & C. A. Hooker, (eds.) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science.* Dordrecht: D. Reidel.

Jaynes E. T. (2003). *Probability Theory: The Logic of Science.* Cambridge University Press.

Jeffreys, H. (1939). *Theory of Probability. (3rd edition.).* Clarendon Press, Oxford.

Kaas, R. E. & Raftery, A. E. (1993). *Bayes Factors and Model Uncertainty. Technical Report No. 254.* Department of Statistics: University of Washington.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 99(430),* 773-795.

Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of poly- graph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291- 302.

Krapohl, D. J. (2005). Polygraph decision rules for evidentiary and paired-testing (Marin Protocol) applications. *Polygraph, 34(3),* 184-192.

Krapohl, D. J. & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: a replication. *Polygraph, 35(1),* 55-63.

Krapohl, D., & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph, 28(3),* 209-222

Kubis, J. F. (1962). *Studies in Lie Detection: Computer Feasibility Considerations.* Final report, Air Force contract RADC-TR 62-205 Contract AF 30(602)-2270. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Kubis, J. F. (1965). *Analysis of Polygraphic Data.* Final report, U.S. Air Force contract RADC-TDR- 64-101, AF30(602)-2634. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Laplace, S. P. (1812). Théorie analytique des probabilités. Paris: Courier.

Lee, P. M. (2004). *Bayesian Statistics, an introduction (3rd ed.).* Wiley.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-55.

Maxfield, B. (2009). Units! In, *Essential Mathcad for Engineering, Science and Math. (2nd ed.).* Academic Press.

Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of confidence intervals. *Philosophy of Science, 48(2),* 269–280.

McDonald, J. H. 2009. *Handbook of Biological Statistics (2nd ed.).* Sparky House Publishing, Baltimore, Maryland.

Miller, R. G. (1981). *Simultaneous Statistical Inference (2nd Ed.).* Springer Verlag New York.

Morey, R. D. & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16(4),* 406–419.

National Center for Credibility Assessment. (2017). *Test Data Analysis: Numerical Evaluation Scoring System Pamphlet.* [Author.] (Retrieved from http://antipolygraph.org on 4/13/2019.)

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine, 17(8),* 857–872.

Nelson, R. (2016a). Scientific (analytic) theory of polygraph testing. *APA Magazine, 49(5)*, 69-82.

Nelson, R. (2016b). Five Minute Science Lesson: Prior Probabilities and the Algebra To Calculate 'Em. *APA Magazine, 49(5)*, 55-61

Nelson, R. (2017a). Multinomial reference distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment, 46(2),* 81-115.

Nelson, R. (2017b). Heuristic principles to select comparison and relevant question pairs when scoring any CQT format. *APA Magazine, 50(1),* 73-83.

Nelson, R. (2017c). Literature survey of structural weighting of polygraph signals: why double the EDA? *Polygraph & Forensic Credibility Assessment, 48(2),* 105-112.

Nelson, R. (2018a). Practical polygraph: a survey and description of decision rules. *APA Magazine, 51(2),* 127-133.

Nelson, R. (2018b). Five minute science lesson: Bayes' theorem and Bayesian analysis. *APA Magazine, 51(5),* 65-78.

Nelson, R. (2020a). Accuracy effects for ESS and three-position scores of federal ZCT exams using the grand total rule with traditional/federal and multinomial cutscores. *Polygraph & Forensic Credibility Assessment, 49(2),* 157-175.

Nelson, R. (2020b). Bigger is better for automated scoring: analysis of minimum constraints for RQ/CQ ratios. *Polygraph & Forensic Credibility Assessment, 49(1),* 61-72.

Nelson, R. (2020c). Multinomial cutscores for Bayesian analysis with ESS and three-position scores of comparison question polygraph tests. *Polygraph & Forensic Credibility Assessment, 49(1),* 61-72.

Nelson, R., Blalock, B., & Handler, M. (2019). Practical polygraph: how to parse categorical results for test questions of diagnostic and screening polygraphs. *APA Magazine, 52(3),* 60-65.

Nelson, R. & Handler, M. (2020). ESS-M Cheat Sheet. *APA Magazine, 53(2),* 61-63.

Nelson. R., Handler, M., Prado, R., & Blalock, R. (2020). A Discussion of PLC and DLC question procedure and Ironic Process Theory. *Polygraph & Forensic Credibility Assessment, 49 (1),* 35-52.

Nelson, R., Handler, M., Coffey, T., Prado, R. & Blalock, B. (2019). How to: a step-by-step worksheet for the multinomial ESS. *Polygraph & Forensic Credibility Assessment, 48 (1),* 60-75.

Nelson, R. Handler, M, Shaw, P, Gougler, M, Blalock, B., Russell, C, Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph, 40(2),* 67-78.

Nelson, R., Handler, M., & Thiel, E. (2021). Posterior odds of deception and truth telling for low and high prior probabilities. *Polygraph & Forensic Credibility Assessment, 50(2)* 96-105.

Nelson, R, & Krapohl, D. (2017). Practical polygraph: a recommendation for combining the upper and lower respiration data for a single respiration score. *APA Magazine, 50(6),* 31-41.

Nelson, R., Rider, J. (2018). Practical Polygraph: ESS-M Made Simple. *APA Magazine, 51 (6),* 55–62.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A. (236)* 333–380

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231*(694-706), 289-337.

Nickerson, R. S. (2000). Null hypothesis significance tests: a review of an old and continuing controversy. *Psychological Methods, 5(2),* 241–301.

Peng R. D. (2011). Reproducible research in computational science. *Science, 334(6060),* 1226–1227.

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional, and le- gal issues surrounding acceptance of polygraph evidence. *Utah Law review, 29,* 29-74.

Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioral Science/Revue Canadienne de la Science Computationnelle 19(4),* p. 389-404.

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). *A Study of the Validity of Polygraph Examinations in Criminal Investigation.* Final Report, National Institute of Justice,

Grant No. 85-IJ-CX-0040. [Reprinted in Polygraph & Forensic Credibility Assessment, 2019, 48 (1)].

Rice, J. A. (2006). *Mathematical Statistics and Data Analysis (3rd ed.).* Belmont, CA: Duxbury Press, 2006.

Rosner, B. (2010). *Fundamentals of Biostatistics.* Brooks/Cole: Boston, MA.

Senter, S. (2003). Modified general question test decision rule exploration. *Polygraph, 32,* 251-263.

Senter, S. M. & Dollins, A. B. (2003). *New Decision Rule Development: Exploration of a two-stage approach. Report number DoDPI00-R-0001.* Department of Defense Polygraph Institute

Research Division, Fort Jackson, SC. Reprinted in Polygraph, 37(2), 149-164.

Senter, S., Waller, J. & Krapohl, D. (2008). Air Force Modified General Question Test Validation Study. *Polygraph, 37(3),* 174-184.

Sidak, Z. (1967). Rectangular confidence region for the means of multivariate normal

distributions. *Journal of the American Statistical Association, 62,* 626-633.

Skiena, S. (1990). *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Addison-Wesley, Reading, MA.

Smithson, M. (2003). *Confidence intervals. Quantitative applications in the Social Sciences Series, No. 140.* SAGE Publications.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103(2684).* 677–680.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley.

Stone, J. (2013). *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis.* Sebtel Press.

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(2), 70-79.

Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph, 9,* 94-108.

Winkler, R. L. (1972). *An Introduction to Bayesian Inference and Decision.* Holt McDougal.

Apéndice A: Referencia rápida del ESS-M

1 Extracción de características.

1.1 Supresión o reducción de la actividad respiratoria

1.2 Amplitud del cambio en la actividad electrodérmica desde el inicio de la respuesta hasta el pico de respuesta

1.3 Amplitud del cambio en la actividad cardiovascular desde el inicio de la respuesta hasta el pico de respuesta

1.4 Reducción vasomotora del volumen sanguíneo capilar de la punta de los dedos

2 Transformación numérica y reducción de datos

2.1 Asigne puntuaciones de 3 posiciones [+, 0, -] utilizando la *regla de "cuanto más grande, mejor"*

2.1.a Seleccione los puntos de análisis según el formato de pregunta
2.1.b Duplicar todas las puntuaciones de EDA a +/- 2

2.2 Puntúa solo las reacciones oportunas

2.2.a La ventana de evaluación es desde el inicio de la pregunta hasta 15 segundos después del inicio
2.2.b La ventana de inicio de la respuesta finaliza 5 segundos después de la respuesta verbal
2.2.c El período de latencia es de 1/2 segundo: no puntúe las respuestas que comienzan durante el período de latencia

2. No puntuar segmentos de datos con artefactos

3 Función de verosimilitud

1 Determinar el nivel alfa

1.a Alfa = 0,05 para el engaño
1.b Alfa = 0,05 para veracidad
1.c El alfa se puede ajustar según lo indiquen las prioridades, los riesgos y los objetivos de la misión

3.1 Determinar la fuerza de la información previa (un previo igual es casi siempre el previo óptimo)
3.2 Seleccione la tabla de referencia correcta
3.3 Determinar las puntuaciones de corte (los paréntesis indican el uso de una corrección estadística)

3.3.a Exámenes de asunto único: +3 / - 3 (-7) con corrección estadística para subtotales engañosos en la etapa 2
3.3.b Exámenes exploratorios de asuntos múltiples: (+1) / -3 con corrección estadística para subtotales veraces

4 Interpretar e informar de los resultados

4.1 Puntuaciones numéricas

4.1.a Puntuación de gran total y puntuación de corte de gran total (solo para exámenes de asunto único)
4.1.b Puntuación de subtotal más baja y puntuación de corte de subtotal

4.2 Resultado de la probabilidad – Análisis bayesiano

4.2.a Factor de Bayes (magnitud del cambio en la fuerza de la información)
4.2.b Probabilidades posteriores de engaño o de veracidad (también la probabilidad posterior)
4.2.c Límite inferior del intervalo de credibilidad Bayesiano (1-alfa) x 100%

4.3 Resultado categórico: seleccione la regla de decisión para exámenes de un solo asunto o múltiples asuntos.

4.3.a Examen de un solo asunto= GTR o TSR (cada pregunta hereda el resultado del examen)
4.3.b Examen de asuntos múltiples = SSR (el examen hereda el resultado del subtotal más bajo y, a continuación, analiza los resultados para otras RQ)

[Reimpreso del original de Nelson y Handler (2020).]

Apéndice B: Hoja de puntuación manual de ESS-M

ID del examen:_____ Fecha de análisis: _____

Calificador: _____

Tabla 4. Matriz de hojas de puntuación para todos las gráficas: sensores como filas de casos, incluidas las puntuaciones torácicas y abdominales								
	R1		R2		R3		R4	
ID de pregunta								
Gráfica 1								
Resp. Torácica.								
Resp. Abdominal.								
Electrodérmico								
Cardio								
Vasomotor								
Gráfica 2								
Resp. Torácica.								
Resp. Abdominal.								
Electrodérmico								
Cardio								
Vasomotor								
Gráfica 3								
Resp. Torácica.								
Resp. Abdominal.								
Electrodérmico								
Cardio								
Vasomotor								
Gráfica 4								
Resp. Torácica.								
Resp. Abdominal.								
Electrodérmico								
Cardio								
Vasomotor								
Gráfica 5								
Resp. Torácica.								
Resp. Abdominal.								
Electrodérmico								
Cardio								
Vasomotor								
Subtotales								
Gran Total								
Resultado de la prueba (círcule)				DI / SR	NDI / NSR			
Resultados de la pregunta (círcule)	DI / SR	NDI / NSR	DI / SR	NDI / NSR	DI / SR	NDI / NSR	DI / SR	NDI / NSR

Regla de decisión (marque uno)	GTR	SSR	TSR
Previo			
Alfa (Engaño/Veracidad)			
Puntuaciones de corte			

[Basado en el original de Nelson y Krapohl (2017).]

Apéndice C-1: Tabla de Referencia del ESS-M para Puntajes de Corte de Gran Total con 3 a 5 presentaciones de 2 a 4 Preguntas Relevantes con o sin el sensor vasomotor

Previa= .5 (1 a 1), Alfa = .05 / .05 (veracidad / engaño)

puntuación	Combinaciones	*Pmf*	*Cdf*	cdfContCor	probabilidades	probabilidadesLL05
-24	9915	.0008*	.0023	.0019	518.7 (.998)	21.4
-23	10248	.0011	.0034	.0028	352.2 (.997)	20.18
-22	10572	.0015	.0048	.0041	242.7 (.996)	18.69
-21	10888	.0020	.0069	.0059	169.7 (.994)	16.95
-20	11193	.0027	.0096	.0082	120.4 (.992)	17.25
-19	11488	.0036	.0132	.0114	86.55 (.989)	14.98
-18	11770	.0047	.0179	.0156	63.05 (.984)	13.98
-17	12040	.0061	.0239	.0210	46.52 (.979)	12.51
-16	12295	.0077	.0316	.0280	34.75 (.972)	10.89
-15	12536	.0097	.0411	.0367	26.26 (.963)	9.29
-14	12760	.0119	.0527	.0475	20.06 (.953)	8.05
-13	12970	.0144	.0668	.0607	15.49 (.939)	6.83
-12	13163	.0172	.0835	.0765	12.07 (.924)	5.74
-11	13342	.0202	.1031	.0953	9.5 (.905)	4.85
-10	13504	.0235	.1257	.1172	7.53 (.883)	4.06
-9	13652	.0269	.1514	.1424	6.02 (.858)	3.41
-8	13783	.0303	.1803	.1710	4.85 (.829)	2.85
-7	13900	.0336	.2122	.2030	3.93 (.797)	2.39
-6	14000	.0369	.2471	.2383	3.2 (.762)	2
-5	14086	.0398	.2847	.2766	2.62 (.723)	1.67
-4	14155	.0424	.3247	.3177	2.15 (.682)	1.39
-3	14210	.0446	.3667	.3611	1.77 (.639)	1.16
-2	14248	.0461	.4102	.4064	1.46 (.594)	0.97
-1	14272	.0471	.4548	.4529	1.21 (.547)	0.8
0	14279	.0475	.5000	.5000	1 (.500)	0.67
1	14272	.0471	.5452	.5471	1.21 (.547)	0.8
2	14248	.0461	.5898	.5936	1.46 (.594)	0.97
3	14210	.0446	.6333	.6389	1.77 (.639)	1.16
4	14155	.0424	.6753	.6823	2.15 (.682)	1.39
5	14086	.0398	.7153	.7234	2.62 (.723)	1.67
6	14000	.0369	.7529	.7617	3.2 (.762)	2
7	13900	.0336	.7878	.7970	3.93 (.797)	2.39
8	13783	.0303	.8197	.8290	4.85 (.829)	2.85
9	13652	.0269	.8486	.8576	6.02 (.858)	3.41
10	13504	.0235	.8743	.8828	7.53 (.883)	4.06
11	13342	.0202	.8969	.9047	9.5 (.905)	4.85
12	13163	.0172	.9165	.9235	12.07 (.924)	5.74
13	12970	.0144	.9332	.9393	15.49 (.939)	6.83
14	12760	.0119	.9473	.9525	20.06 (.953)	8.05
15	12536	.0097	.9590	.9633	26.26 (.963)	9.29
16	12295	.0077	.9685	.9720	34.75 (.972)	10.89
17	12040	.0061	.9761	.9790	46.52 (.979)	12.51
18	11770	.0047	.9821	.9844	63.05 (.984)	13.98
19	11488	.0036	.9868	.9886	86.55 (.989)	14.98
20	11193	.0027	.9904	.9918	120.4 (.992)	17.25
21	10888	.0020	.9931	.9941	169.7 (.994)	16.95
22	10572	.0015	.9952	.9959	242.7 (.996)	18.69
23	10248	.0011	.9966	.9972	352.2 (.997)	20.18
24	9915	.0008*	.9977	.9981	518.7 (.998)	21.4
* Valores extremos omitidos						

La *puntuación* (score) es la calificación de Gran Total. *Combinaciones* (ways) es el número de combinaciones de sensor-puntuación que pueden lograr cada puntuación total. *pmf* es la masa de probabilidad para cada puntuación. *cdf* es la suma acumulada de la *pmf*. *cdfContCor* es la *cdf* corregida por continuidad, de modo que la estimación estadística siempre excede el valor estadístico real, que también se utiliza como probabilidad posterior. *probabilidades* (odds) son las probabilidades posteriores de veracidad o engaño, calculadas a partir del *cdfContCor* usando p/(1-p). Además, el *cdfContCor* se puede calcular a partir de las probabilidades usando probabilidades/(1+probabilidades). Las *probabilidadesLL05* (oddsLL05) son los límites inferiores del intervalo de credibilidad (intervalo de confianza bayesiano) para una *previa*=.5 y alfa/2 = .05 para veracidad y engaño.

[Reimpreso de Nelson Handler, Coffey, Prado y Blalock (2019).]

Apéndice C-2: Tabla de Referencia ESS-M para las puntuaciones subtotales de los exámenes de un solo tema con 3 a 5 presentaciones de 2 a 4 RQ con o sin el Vasomotor Sensor

Previa = 0,5 (1 a 1), Alfa = 0,05 / 0,05 (veracidad / engaño): se incluyen todas las correcciones estadísticas

puntuación	*combinaciones*	*pmf*	*cdf*	*Cdf ContCor*	*Probabilidades*	*Probabilidades 234RQs*	*Probabilidades LL05*	*Probabilidades 234LL05*
-15	161	.0005*	.0009	.0007	1517	11.49 (.92)	7.71	3.32
-14	200	.0011	.0020	.0015	682.2	8.8 (.898)	7.56	2.84
-13	243	.0021	.0041	.0030	328.4	6.9 (.873)	7.27	2.42
-12	287	.0037	.0077	.0059	168	5.52 (.847)	6.79	2.07
-11	333	.0062	.0139	.0109	90.88	4.5 (.818)	6.1	1.81
-10	378	.0099	.0236	.0190	51.67	3.73 (.789)	5.22	1.56
-9	423	.0150	.0383	.0315	30.72	3.13 (.758)	4.84	1.37
-8	465	.0216	.0592	.0500	19.01	2.67 (.728)	4.11	1.19
-7	505	.0297	.0875	.0758	12.19	2.3 (.697)	3.3	1.05
-6	540	.0389	.1242	.1104	8.06	2.01 (.668)	2.66	0.93
-5	571	.0489	.1697	.1546	5.47	1.76 (.638)	2.06	0.83
-4	595	.0588	.2236	.2087	3.79	1.56 (.609)	1.58	0.74
-3	615	.0678	.2852	.2720	2.68	1.39 (.582)	1.19	0.66
-2	628	.0750	.3531	.3432	1.91	1.24 (.554)	0.89	0.59
-1	637	.0797	.4254	.4201	1.38	1.11 (.526)	0.65	0.53
0	639	.0814	.5000	.5000	1	1	0.48	0.48
1	637	.0797	.5746	.5799	1.38	2.63	0.65	1.18
2	628	.0750	.6469	.6568	1.91	7.01	0.89	2.45
3	615	.0678	.7148	.7280	2.68	19.17	1.19	4.13
4	595	.0588	.7764	.7913	3.79	54.52	1.58	5.31
5	571	.0489	.8303	.8454	5.47	163.4	2.06	6.77
6	540	.0389	.8758	.8896	8.06	522.8	2.66	7.47
7	505	.0297	.9125	.9242	12.19	1810	3.3	7.73
8	465	.0216	.9408	.9500	19.01	6870	4.11	7.82
9	423	.0150	.9617	.9685	30.72	28990	4.84	7.84
10	378	.0099	.9764	.9810	51.67	137900	5.22	7.84
11	333	.0062	.9861	.9891	90.88	750600	6.1	7.85
12	287	.0037	.9923	.9941	168	4745000	6.79	7.85
13	243	.0021	.9959	.9970	328.4	35400000	7.27	7.85
14	200	.0011	.9980	.9985	682.2	317000000	7.56	7.85
15	161	.0005*	.9991	.9993	1517	3490000000	7.71	7.85
* Valores extremos omitidos								

La *puntuación* (score) es la puntuación subtotal más baja. *Combinaciones (ways)* es el número de combinaciones de sensor-puntuación que pueden lograr cada puntuación subtotal. pmf es la masa de probabilidad para cada puntuación. *cdf* es la suma acumulada de la *pmf*. *cdfContCor* es la *cdf* corregida por continuidad, de modo que la estimación estadística siempre excede el valor estadístico real, que también se utiliza como probabilidad posterior. Las *probabilidades (odds)* son las probabilidades posteriores de veracidad o engaño para un único subtotal (sin corrección estadística), calculadas a partir del *cdfContCor* usando p/(1-p). Además, el *cdfContCor* se puede calcular a partir de las probabilidades usando probabilidades/(1+probabilidades). *probabilidadesLL05 (oddsLL05)* es el límite inferior del intervalo de credibilidad (intervalo de confianza bayesiano) para una previa=.5 y alfa/2 = .05 para veracidad y engaño. *probabilidades234RQ* (odds234RQ) son las probabilidades medias de engaño con corrección estadística para 2, 3 y 4 RQ. Probabilidades234LL05 (Odds234LL05) son los límites inferiores medios del intervalo de credibilidad (intervalo de confianza bayesiano) para una previa=.5 y alfa/2 = .05 para veracidad y engaño con 2, 3 o 4 RQ.

[Reimpreso de Nelson Handler, Coffey, Prado y Blalock (2019).]

Apéndice C-3: Tabla de referencia ESS-M para las puntuaciones subtotales de los exámenes de asuntos múltiples de 3 a 5 presentaciones de 2 a 4 RQ con o sin el sensor vasomotor

Previa = 0,5 (1 a 1), Alfa = 0,05 / 0,05 (veracidad/ engaño): se incluyen todas las correcciones estadísticas

puntuación	Combinaciones	*pmf*	*fcd*	Cdf ContCor	probabilidades	probabilidades234RQs	probabilidadesLL05	probabilidades234LL05
-15	161	.0005*	.0009	.0007	1517 (>.999)	11.49	7.71	3.32
-14	200	.0011	.0020	.0015	682.2 (.999)	8.8	7.56	2.84
-13	243	.0021	.0041	.0030	328.4 (.997)	6.9	7.27	2.42
-12	287	.0037	.0077	.0059	168 (.994)	5.52	6.79	2.07
-11	333	.0062	.0139	.0109	90.88 (.989)	4.5	6.1	1.81
-10	378	.0099	.0236	.0190	51.67 (.981)	3.73	5.22	1.56
-9	423	.0150	.0383	.0315	30.72 (.968)	3.13	4.84	1.37
-8	465	.0216	.0592	.0500	19.01 (.950)	2.67	4.11	1.19
-7	505	.0297	.0875	.0758	12.19 (.924)	2.3	3.3	1.05
-6	540	.0389	.1242	.1104	8.06 (.890)	2.01	2.66	0.93
-5	571	.0489	.1697	.1546	5.47 (.845)	1.76	2.06	0.83
-4	595	.0588	.2236	.2087	3.79 (.791)	1.56	1.58	0.74
-3	615	.0678	.2852	.2720	2.68 (.728)	1.39	1.19	0.66
-2	628	.0750	.3531	.3432	1.91 (.656)	1.24	0.89	0.59
-1	637	.0797	.4254	.4201	1.38 (.580)	1.11	0.65	0.53
0	639	.0814	.5000	.5000	1	1	0.48	0.48
1	637	.0797	.5746	.5799	1.38	2.63 (.725)	0.65	1.18
2	628	.0750	.6469	.6568	1.91	7.01 (.875)	0.89	2.45
3	615	.0678	.7148	.7280	2.68	19.17 (.95)	1.19	4.13
4	595	.0588	.7764	.7913	3.79	54.52 (.982)	1.58	5.31
5	571	.0489	.8303	.8454	5.47	163.4 (.994)	2.06	6.77
6	540	.0389	.8758	.8896	8.06	522.8 (.998)	2.66	7.47
7	505	.0297	.9125	.9242	12.19	1810 (>.999)	3.3	7.73
8	465	.0216	.9408	.9500	19.01	6870 (>.999)	4.11	7.82
9	423	.0150	.9617	.9685	30.72	28990 (>.999)	4.84	7.84
10	378	.0099	.9764	.9810	51.67	137900 (>.999)	5.22	7.84
11	333	.0062	.9861	.9891	90.88	750600 (>.999)	6.1	7.85
12	287	.0037	.9923	.9941	168	4745000 (>.999)	6.79	7.85
13	243	.0021	.9959	.9970	328.4	3,540000000 (>,999)	7.27	7.85
14	200	.0011	.9980	.9985	682.2	3.1700000000 (>.999)	7.56	7.85
15	161	.0005*	.9991	.9993	1517	3,49000000000 (>,999)	7.71	7.85

* Valores extremos omitidos

La *puntuación (Score)* es la puntuación del subtotal más baja. *Combinaciones* (*Ways)* es el número de combinaciones de puntuación de sensor que pueden lograr cada puntuación de subtotal. *pmf* es la masa de probabilidad para cada puntuación. *cdf* es la suma acumulada del *PMF*. *cdfContCor* es la cdf corregida de continuidad, de modo que la estimación estadística siempre supera el valor estadístico real, que también se utiliza como probabilidad a posteriori. *Las probabilidades (odds)* son las probabilidades posteriores de veracidad o engaño para un solo subtotal (sin corrección estadística), calculadas a partir del *cdfContCor* usando *p*/(1-p). Además, el *cdfContCor* se puede calcular a partir de las probabilidades utilizando *odds*/(1+odds). *probabilidadesLL05 (oddsLL05)* es el límite inferior del intervalo de credibilidad (intervalo de confianza bayesiano) para prior=.5 y alpha/2 = 0,05 para veracidad y el engaño. *probabilidades234RQ* (odds234RQ) mediana de probabilidades de engaño con corrección estadística para 2, 3 y 4 RQ. *Las probabilidades234LL05 (Odds234LL05)* son los límites inferiores de la mediana del intervalo de credibilidad (intervalo de confianza bayesiano) para prior=.5 y alpha/2 = .05 para veracidad y el engaño con 2, 3 o 4 RQ.

[Reimpreso de Nelson Handler, Coffey, Prado y Blalock (2019).]

Apéndice D: Hoja de trabajo de análisis ESS-M

1.	**ID Examen / Nombre de Examinado**	
2.	**Fecha de Examen (fecha de análisis)**	
3.	**Examinador / analista**	
4.	Examen Diagnóstico o Exploratorio	Diagnóstico | Exploratorio
5.	Regla de Decisión	GTR | TSR | SSR
6.	Correción Estadística para Subtotales	NA | Subtotales de Engaño | Subtotales de Veracidad
7.	Niveles Alfa (nivel de sig. de 1-cola): Veracidad | Engaño	|
8.	Posibilidades previas de engaño = p / (1 - p)	
9.	Probabilidad previa de engaño = odds / (1 + odds)	
10.	Puntajes de corte: Veracidad/Engaño (Subtotal)	| ( )
11.	Número de RQs	2RQs 3RQs 4RQs
12.	RQ: ID pregunta / puntaje	|
13.	RQ: ID pregunta / puntaje	|
14.	RQ: ID pregunta / puntaje	|
15.	RQ: ID pregunta / puntaje	|
16.	Gran total (coloque NA cuando use el SSR)	
17.	Puntaje Subtotal menor: ID pregunta / puntaje	|
18.	**Resultado**	**DI/SR NDI/NSR INC/NO**
19.	**Clasificado por**	**Gran Total Subtotal Menor**
20.	**Tabla de Referencia ESS-M (circule uno)**	**B-1 (GTR & TSR) | B-2 (TSR) | B-3 (SSR)**
21.	**Posibilidades posteriores = p / (1 – p) (error sistemático estadístico**	
22.	**Probabilidad posterior = odds / (1 + odds)**	
23.	**Límite-inferior de Intervalo de credibilidad Posterior (oddsLL05 o odds234LL05)**	
24.	**Factor Bayes = Posibilidades posteriores / Posibilidades previas**	
25.	**Tolerancia Alfa/Nivel sig. (error aleatorio est.)**	**DI/SR NDI/NSR**
26.	**(1 - *alpha*) x 100% interval creíble Bayesiano**	

[Reimpreso de Nelson Handler, Coffey, Prado y Blalock (2019).]

Apéndice E-1: Puntuaciones de corte de ESS-M para exámenes de un solo tema

Puntuaciones ESS-M / Examen de evento específico

Probabilidades previas de engaño	Probabilidad previa	Alfa (veracidad/engaño)								
		.01/.01	.01/.05	.01/.10	.05/.01	.05/.05	.05/.10	.10/.01	.10/.05	.10/.10
9 a 1 (9 en 10)	.90	+14 / -9 (ninguno)	+14 / -6 (ninguno)	+14 / -4 (-21)	+13 / -9 (ninguno)	+13 / -6 (ninguno)	+13 / -4 (-21)	+13 / -9 (ninguno)	+13 / -6 (ninguno)	+13 / -4 (-21)
8 a 1 (8 en 9)	.89	+13 / -8 (ninguno)	+13 / -5 (ninguno)	+13 / -4 (-20)	+13 / -8 (ninguno)	+13 / -5 (ninguno)	+13 / -4 (-20)	+12 / -8 (ninguno)	+12 / -5 (ninguno)	+12 / -4 (-20)
7 a 1 (7 en 8)	.88	+13 / -7 (ninguno)	+13 / -5 (-22)	+13 / -4 (-18)	+12 / -7 (ninguno)	+12 / -5 (-22)	+13 / -4 (-18)	+12 / -7 (ninguno)	+12 / -5 (-22)	+12 / -4 (-18)
6 a 1 (6 en 7)	.86	+12 / -7 (ninguno)	+12 / -4 (-20)	+12 / -4 (-16)	+11 / -7 (ninguno)	+11 / -4 (-20)	+11 / -4 (-16)	+11 / -7 (ninguno)	+11 / -4 (-20)	+11 / -4 (-16)
5 a 1 (5 en 6)	.83	+11 / -6 (ninguno)	+11 / -4 (-17)	+11 / -3 (-15)	+10 / -6 (ninguno)	+10 / -4 (-17)	+10 / -3 (-15)	+10 / -6 (ninguno)	+10 / -4 (-17)	+10 / -3 (-15)
4 a 1 (4 en 5)	.80	+10 / -5 (-21)	+10 / -4 (-15)	+10 / -3 (-14)	+10 / -5 (-21)	+10 / -4 (-15)	+10 / -3 (-14)	+9 / -5 (-21)	+9 / -4 (-15)	+9 / -3 (-14)
3 a 1 (3 en 4)	.75	+9 / -5 (-16)	+9 / -3 (-13)	+9 / -3 (-12)	+8 / -5 (-16)	+8 / -3 (-13)	+8 / -3 (-12)	+8 / -5 (-16)	+8 / -3 (-13)	+8 / -3 (-12)
2 a 1 (2 en 3)	.67	+7 / -4 (-13)	+7 / -3 (-11)	+7 / -2 (-10)	+6 / -4 (-13)	+6 / -3 (-11)	+6 / -2 (-10	+6 / -4 (-13)	+6 / -3 (-11)	+6 / -2 (-10)
1 a 1 (1 en 2)	**.50**	**+4 / -4 (-9)**	**+4 / -3 (-7)**	**+4 / -2 (-6)**	**+3 / -4 (-9)**	**+3 / - 3 (-7)**	**+3 / -2 (-6)**	**+2 / -4 (-9)**	**+2 / -3 (-7)**	**+2 / -2 (-6)**
De 1 a 2 (1 en 3)	.33	+4 / -7 (-11)	+4 / -6 (-9)	+4 / -6 (-8)	+3 / -7 (-11)	+3 / -6 (-9)	+3 / -6 (-8)	+2 / -7 (-11)	+2 / -6 (-9)	+2 / -6 (-8)
De 1 a 3 (1 en 4)	.25	+5 / -9 (-11)	+5 / -8 (-9)	+5 / -8 (-8)	+3 / -9 (-11)	+3 / -8 (-9)	+3 / -8 (-8)	+3 / -9 (-11)	+3 / -8 (-9)	+3 / -8 (-8)
De 1 a 4 (1 en 5)	.20	+5 / -10 (-12)	+5 / -10 (-10)	+5 / -9 (-9)	+4 / -10 (-12)	+4 / -10 (-10)	+4 / -9 (-9)	+3 / -10 (-12)	+3 / -10 (-10)	+3 / -9 (-9)
De 1 a 5 (1 en 6)	.17	+6 / -11 (-12)	+6 / -10 (-10)	+6 / -10 (-10)	+4 / -11 (-12)	+4 / -10 (-10)	+4 / -10 (-10)	+3 / -11 (-12)	+3 / -10 (-10)	+3 / -10 (-10)
De 1 a 6 (1 en 7)	.14	+7 / -12 (-13)	+7 / -11 (-11)	+7 / -11 (-10)	+4 / -12 (-13)	+4 / -11 (-11)	+4 / -11 (-10)	+4 / -12 (-13)	+4 / -11 (-11)	+4 / -11 (-10)
De 1 a 7 (1 en 8)	.13	+7 / -12 (-13)	+7 / -12 (-11)	+7 / -11 (-10)	+5 / -12 (-13)	+5 / -12 (-11)	+5 / -11 (-10)	+4 / -12 (-13)	+4 / -12 (-11)	+4 / -11 (-10)
De 1 a 8 (1 en 9)	.11	+8 / -13 (-13)	+7 / -13 (-11)	+8 / -12 (-10)	+5 / -13 (-13)	+5 / -13 (-11)	+5 / -12 (-10)	+4 / -13 (-13)	+4 / -13 (-11)	+4 / -12 (-10)
De 1 a 9 (1 en 10)	.10	+9 / -14 (-13)	+9 / -13 (-12)	+9 / -13 (-11)	+6 / -14 (-13)	+6 / -13 (-12)	+6 / -13 (-11)	+4 / -14 (-13)	+4 / -13 (-12)	+4 / -13 (-11)

Los paréntesis indican el uso de una corrección estadística para los efectos de multiplicidad.

[Reimpreso de Nelson (2020c).]

Apéndice E-2: Puntuaciones de corte de ESS-M para exámenes asuntos múltiples

Puntuaciones ESS-M / Examen de asuntos múltiples

		Alfa (veracidad/engaño)								
Probabilidades previas de engaño	Probabilidad previa	.01/.01	.01/.05	.01/.10	.05/.01	.05/.05	.05/.10	.10/.01	.10/.05	.10/.10
9 a 1 (9 en 10)	.90	(+8) / ninguno	(+8) / ninguno	(+8) / -9	(+7) / ninguno	(+7) / ninguno	(+7) / -9	(+7) / ninguno	(+7) / ninguno	(+7) / -9
8 a 1 (8 en 9)	.89	(+7) / ninguno	(+7) / ninguno	(+7) / -6	(+7) / ninguno	(+7) / ninguno	(+7) / -6	(+7) / ninguno	(+7) / ninguno	(+7) / -6
7 a 1 (7 en 8)	.88	(+7) / ninguno	(+7) / -10	(+7) / -6	(+7) / ninguno	(+7) / -10	(+7) / -6	(+7) / ninguno	(+7) / -10	(+7) / -6
6 a 1 (6 en 7)	.86	(+7) / ninguno	(+7) / -8	(+7) / -5	(+6) / Ninguno	(+6) / -8	(+6) / -4	(+6) / Ninguno	(+6) / -8	(+6) / -5
5 a 1 (5 en 6)	.83	(+6) / -14	(+6) / -5	(+6) / -4	(+6) / -14	(+6) / -5	(+6) / -4	(+6) / -14	(+6) / -5	(+6) / -4
4 a 1 (4 en 5)	.80	(+6) / -9	(+6) / -5	(+6) / -4	(+5) / -9	(+5) / -5	(+5) / -4	(+5) / -9	(+5) / -5	(+5) / -4
3 a 1 (3 en 4)	.75	(+5) / -6	(+5) / -4	(+5) / -3	(+5) / -6	(+5) / -4	(+5) / -3	(+4) / -6	(+4) / -4	(+4) / -3
2 a 1 (2 en 3)	.67	(+4) / -5	(+4) / -3	(+4) / -3	(+3) / -5	(+3) / -3	(+3) / -3	(+3) / -5	(+3) / -3	(+3) / -3
1 a 1 (1 en 2)	**.50**	**(+2) / -4**	**(+2) / -3**	**(+2) / -2**	**(+1) / -4**	**(+1) / -3**	**(+1) / -2**	**(+1) / -4**	**(+1) / -3**	**(+1) / -2**
De 1 a 2 (1 en 3)	.33	(+1) / -6	(+1) / -5	(+1) / -4	(0) / -6	(0) / -5	(0) / -4	(0) / -6	(0) / -5	(0) / -4
De 1 a 3 (1 en 4)	.25	(0) / -7	(0) / -6	(0) / -6	(0) / -7	(0) / -6	(0) / -6	(0) / -7	(0) / -6	(0) / -6
De 1 a 4 (1 en 5)	.20	(+1) / -7	(0) / -7	(0) / -6	(0) / -7	(0) / -7	(0) / -6	(0) / -7	(0) / -7	(0) / -6
De 1 a 5 (1 en 6)	.17	(+4) / -8	(+4) / -7	(+4) / -7	(0) / -8	(0) / -7	(0) / -7	(0) / -8	(0) / -7	(0) / -7
De 1 a 6 (1 en 7)	.14	(Ninguno) / -8	(Ninguno) / -8	(Ninguno) / -7	(0) / -8	(0) / -8	(0) / -7	(0) / -8	(0) / -8	(0) / -7
De 1 a 7 (1 en 8)	.13	(Ninguno) / -9	(Ninguno) / -8	(Ninguno) / -8	(0) / -9	(0) / -8	(0) / -8	(0) / -9	(0) / -8	(0) / -8
De 1 a 8 (1 en 9)	.11	(Ninguno) / -9	(Ninguno) / -8	(Ninguno) / -8	(Ninguno) / -9	(Ninguno) / -8	(Ninguno) / -8	(0) / -9	(0) / -8	(0) / -8
De 1 a 9 (1 en 10)	.10	(Ninguno) / -9	(Ninguno) / -9	(Ninguno) / -8	(Ninguno) / -9	(Ninguno) / -9	(Ninguno) / -8	(0) / -9	(0) / -9	(0) / -8

Los paréntesis indican el uso de una corrección estadística para los efectos de multiplicidad.

[Reimpreso de Nelson (2020c).]

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
Traductores jpadilla@deltasis.co y rodolfo@poligrafia.com.mx