# Reporte Breve:

**Análisis Bootstrap de los Efectos de Pruebas en Serie Utilizando Datos de Laboratorio**

# Raymond Nelson

**ABSTRACT**

Se analizó la precisión de criterio para la *prueba-exploratoria-de-asunto-único-en-serie* (SSIST) utilizando datos de laboratorio de dos publicaciones previas en las que se adaptó el formato de la prueba exploratoria con mentira dirigida (DLST) al contexto de prueba de asunto único. El contexto de pruebas en serie implica la adición de varios exámenes de asunto único - cada uno con múltiples presentaciones de un conjunto de RQs no independientes, junto con preguntas de procedimiento - en un solo examen. Los resultados SSIST son negativos - no requieren de entrevistas, pruebas o investigaciones adicionales - cuando los resultados de todas las series son negativos. Los resultados SSIST son positivos - lo que puede requerir entrevistas, pruebas o investigaciones adicionales - cuando los resultados de cualquier serie son positivos. Resultados previos que utilizaron métodos Monte Carlo, sugirieron que las estrategias de pruebas en serie no evitan los efectos bien conocidos de *multiplicidad* estadística, también conocidos como efectos de *alfa inflado* (Nelson, 2023). También se denominan efecto de *error acumulativo* y, de forma más general, efecto de *comparación múltiple* o *problema de comparaciones múltiples.* Losresultados analíticos de este proyecto dan una mayor ilustración de los efectos de error acumulativo que pueden derivarse del uso de estrategias de pruebas en serie y resaltan la necesidad de investigación adicional sobre el cálculo y manejo de los efectos de error acumulativo tanto para las metodologías de asunto múltiple como para las SSIST.

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.

Traductor. rodolfo@poligrafia.com.mx

La prueba-exploratoria-de-asunto-único-en-serie (SSIST) es un enfoque exploratorio de asuntos múltiples que intenta aprovechar los efectos de precisión de criterio de un formato de prueba de asunto único, dentro del contexto de las pruebas1 de asuntos múltiples. O'Burke et al., (2022) propusieron el uso de una prueba de exploratoria de secuencia única con asunto único (es decir, varias iteraciones de las preguntas de prueba en un solo registro de secuencia) y plantearon la hipótesis de que una serie de pruebas exploratorias de secuencia única puede tener ventajas en comparación con los formatos de pruebas poligráficas exploratorias de asunto múltiple.

O'Burke (2022) describió el uso del enfoque SSIST para cada tema objetivo dentro de un examen con múltiples series. Describieron que cada serie (identificada por un caracter alfabético) abordaba un asunto objetivo de examinación diferente: serie-A uso ilegal de drogas, serie-B delitos graves, serie-C violencia doméstica y serie-D delitos sexuales. Los examinados que aprueban todas las series se clasifican como veraces o que pasan la prueba (es decir, los resultados de las pruebas se reportan como *respuestas no significativas no indicativas de engaño*). Los examinados se clasifican como engaño (es decir, los resultados de las pruebas se reportan como *respuestas significativas indicativas de engaño*) si obtienen un resultado positivo en cualquiera de las series.

El análisis del enfoque SSIST utilizando métodos Monte Carlo (Nelson, 2023) sugirió que las ventajas hipotetizadas pueden ser elusivas, porque las estrategias de pruebas en serie no evitan los ya conocidos efectos estadísticos de *multiplicidad* - también conocidos como el efecto de *alfa inflado*, o efecto del *error acumulativo*, y de manera generalizada como efecto de *comparación múltiple* o *problema de las comparaciones múltiples.*

Los efectos de multiplicidad son las razones por las que los investigadores y los estadistas están atentos a los problemas potenciales en torno al cálculo de resultados estadísticos múltiples - incluyendo los efectos de multiplicidad derivados de un análisis primario y a los que están alrededor del análisis *post hoc* (denominados por los profesionales de la poligrafía del campo como *break-out* testing) en general. El conocimiento y la experiencia asociada a los efectos de multiplicidad son la base, cuando es posible, para el uso rutinario del método analítico ómnibus (es decir, la familia de ANOVAs en la investigación científica, y los formatos de prueba de asuntos múltiples en el contexto de la exploración poligráfica), y del uso de correcciones estadísticas (por ejemplo, Bonferroni y Sidak) cuando es práctico [Véase Nelson (2015) para una introducción y discusión].

1 Las pruebas de diagnóstico se diferencian de las exploratorias por la presencia de un problema conocido (es decir, una acusación o incidente) en el que una persona es sospechosa de involucramiento. Las pruebas poligráficas exploratorias se diferencian de los exámenes diagnósticos de otra manera importante - Los polígrafos exploratorios a menudo se realizan como exámenes de asuntos múltiples y se interpretan bajo un supuesto de no independencia (APA, 2011; 2023).

Un análisis previo utilizando métodos Monte Carlo (Nelson, 2023) proporcionó una visión sobre los efectos esperados del enfoque SSIST en comparación con el enfoque tradicional de pruebas ómnibus (es decir, de asunto múltiple). Los resultados del análisis Monte Carlo revelaron un aumento en la sensibilidad de la prueba al engaño y una disminución de los errores falso negativo, junto con una disminución desproporcionada de la especificidad de la prueba con el correspondiente aumento de los errores falsos positivos y de resultados inconclusos. Los efectos acumulados mostraron una disminución general de la precisión en la clasificación en comparación con los efectos de las pruebas de asuntos múltiples descritos por APA (2011). Este proyecto pretende explorar más a fondo los efectos esperados, y complementar la creciente base de evidencia y conocimientos científicos asociados al SSIST.

# Método

Se obtuvieron datos de dos estudios de laboratorio (Prado, et al., 2015a; 2015b) de polígrafos de asunto único que constaban de dos preguntas relevantes (RQ) no independientes presentadas cuatro veces en una sola secuencia. Los formatos de prueba diagnóstica descritos en los estudios de Prado et al. son una adaptación de la *prueba-exploratoria-de-mentira-dirigida* (DLST) de asunto múltiple. Aunque este nombre no se incluyó en las dos publicaciones, este formato de prueba de asunto único a veces es denominado por los practicantes de polígrafo de campo como la *prueba-diagnóstica-de-mentira-dirigida* (DLDT)2. Las puntuaciones de los dos estudios de laboratorio se agregaron y se sometieron a un re-muestreo Bootstrap para modelar los efectos de las series de polígrafos exploratorios de asunto único. La puntuación media para los casos con criterio de culpabilidad fue de -3.67 (sd=5.76) y la puntuación media para los casos con criterio de inocencia fue de 5.26 (sd=5.48). El modelo Bootstrap permitió conocer con certeza el estado de criterio de cada serie (cada asunto objetivo) - porque se estableció en el experimento de laboratorio controlado - y se utilizaron estados de criterio independientes para cada serie en cada caso dentro del modelo Boostrap3.

2Las pruebas diagnósticas son aquellas pruebas realizadas en respuesta a un problema conocido (APA, 2023), y hacen uso de formatos de prueba de asunto único para los que los problemas de multiplicidad pueden evitarse o manejarse fácilmente. Las pruebas exploratorias son las que se realizan en ausencia de un problema conocido (por ejemplo, un incidente o una acusación conocida). Aunque los objetivos prácticos podrían ser diferentes, la validación tanto de las pruebas de diagnóstico como de las exploratorias se consigue mediante el análisis de la sensibilidad, la especificidad y las tasas de error de las pruebas. Cuando éstas se conocen con una confianza razonable, es aceptable generalizar los tamaños de los efectos esperados de un contexto de uso a otro.

3Los estados de criterio múltiple para el polígrafo de campo se interpretan bajo el supuesto de independencia, pero se sabe que no son completamente independientes. Hay una variedad de factores que limitan potencialmente la independencia entre los objetivos de polígrafos de asuntos múltiples, incluyendo los recursos atencionales limitados del examinado y por el hecho de que algunos comportamientos pueden covariar (por ejemplo, delitos graves, agresión sexual).

# Análisis

Se utilizaron métodos de re-muestreo Bootstrap para calcular los intervalos de confianza estadística de las tasas de sensibilidad de la prueba (verdadero positivo o VP) y especificidad (verdadero negativo o VN), además de las tasas de error de falso negativo (FN) y falso positivo (FP). Estas estadísticas se utilizaron para calcular otros parámetros de precisión del criterio, como las tasas de resultados inconclusos en los casos de criterio inocente (I-INC) y de criterio culpable (C-INC), la proporción de clasificaciones correctas (no inconclusos) en los casos de criterio culpable (CPC) y de criterio inocente (IPC), y las tasas de precisión no ponderada (ACCY) y de resultados inconclusos no ponderados (INC). También se calculó el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN). Para calcular los errores estándar se utilizó un procedimiento Bootstrap doble, consistente en 1 millón de iteraciones de los datos de muestreo, y los intervalos de confianza se calcularon como los cuantiles .025 y .975 del espacio Bootstrap.

# Resultados

La tabla 1 muestra las estimaciones de punto y los intervalos de confianza del 95% para los efectos de precisión utilizando los datos agregados de los estudios de Prado et al., junto con modelos SSIST que incluyen dos, tres y cuatro asuntos objetivo independientes. En el enfoque SSIST, cada asunto objetivo independiente se presenta como una serie de prueba independiente. Cada serie consta de varias presentaciones/repeticiones de las RQs junto con otras preguntas de examen. En los exámenes asunto único, las RQ se interpretan asumiendo que no son independientes. Sin embargo, un examen de series múltiples puede constar de varios conjuntos diferentes de preguntas de prueba. Un SSIST completo consta de todas las series (objetivos). Se dice que los examinados producen resultados negativos en la prueba si los resultados son negativos (es decir, estadísticamente significativos de veracidad) en todas las series de preguntas, y se dice que producen resultados positivos si los resultados son positivos (es decir, estadísticamente significativos de engaño) para cualquier serie de preguntas. El resultado global de la prueba es correcto si los resultados de todas las series son correctos, y el resultado de la prueba es incorrecto si el resultado de alguna serie es incorrecto. El resultado global de la prueba es inconcluso si ninguna de las series es positiva y una o más series no son estadísticamente significativas para veracidad. De este modo, la interpretación de los resultados del SSIST es conceptualmente similar a los formatos poligráficos de asunto múltiple. También se muestran en la Tabla 1 las estimaciones de puntos y los intervalos de confianza para las técnicas poligráficas de asunto múltiple (APA, 2011), y de los resultados agregados de los estudios publicados sobre la DLST (Prado et al., 2015a; 2015b).

**Tabla 2. Resumen de la precisión para exámenes en serie de asunto único con estados de criterio independiente (mixto), incluyendo Media (SE) [IC 95%].**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Métrica | DLDT(Datos Prado et al, 2015a, 2015b) | SSIST2 series separadas (2 Temas Objetivo\*) | SSIST3 series separadas (3 Temas Objetivo \*) | SSIST4 series separadas (4 Temas Objetivo \*) | Asunto Múltiple Técnicas PDD (APA, 2011) |
| Sensibilidad (TP) | .717 (.065) | .751 (.061) | .762 (.060) | .767 (.059) | .771 (.072) |
|  | [.592 to .836] | [.630 to .859] | [.642 to .868] | [.648 to .871] | [.630 to .911] |
| Especificidad (TN) | .779 (.059) | .605 (.070) | .471 (.071) | .366 (.069) | .719 (.047) |
|  | [.659 to .889] | [.468 to .736] | [.337 to .605] | [.235 to .497] | [.626 to .811] |
| Falso negativo (FN) | .154 (.052) | .139 (.049) | .132 (.048) | .131 (.048) | .113 (.058) |
|  | [.061 to .264] | [.053 to .238] | [.048 to .228] | [.047 to .226] | [.001 to .226] |
| Falso positivo (FP) | .085 (.039) | .163 (.053) | .235 (.060) | .300 (.065) | .144 (.039) |
|  | [.019 to .167] | [.071 to .267] | [.128 to .352] | [.184 to .428] | [.066 to .221] |
| Inconcluso culpable (G-INC) | .129 (.047) | .111 (.044) | .106 (.043) | .102 (.042) | .112 (.051) |
|  | [.041 to .228] | [.034 to .200] | [.031 to .194] | [.029 to .189] | [.013 to .212] |
| Inconcluso inocente (I-INC) | .136 (.048) | .232 (.061) | .294 (.065) | .334 (.068) | .136 (.031) |
|  | [.048 to .235] | [.124 to .351] | [.175 to .421] | [.209 to .446] | [.076 to .196] |
| Inconcluso no ponderado (INC) | .132 (.034) | .171 (.038) | .200 (.039) | .218 (.040) | .125 (.029) |
|  | [.070 to .200] | [.104 to .246] | [.128 to .277] | [.143 to .297] | [.068 to .183] |
| Porcentaje correcto culpable(GPC) | .823 (.058) | .844 (.055) | .852 (.053) | .855 (.053) | .873 (.066) |
|  | [..700 to .929] | [..734 to .940] | [.747 to .946] | [..749 to .948] | [.744 to .999] |
| Porcentaje correcto inocente (IPC) | .902 (.046) | .788 (.067 | .667 (.079 | .549 (.087) | .831 (.043) |
|  | [.805 to .978] | [.654 to .905] | [.512 to .812] | [.381 to .549] | [.746 to .915] |
| Precisión no ponderada (ACCY) | .862 (.037) | .816 (.044) | .760 (.048) | .702 (.051) | .850 (.039) |
|  | [.786 to .930] | [.729 to .893] | [.666 to .848] | [.603 to .796] | [.773 to .926] |
| Valor predictive positivo (PPV) | .894 (.049) | .822 (.057) | .765 (.060) | .719 (.062) | .828 (.059) |
|  | [.789 to .976] | [.706 to .921] | [.646 to .873] | [.599 to .833] | [.712 to .943] |
| Valor predictive negativo (NPV) | .836 (.054) | .813 (.065) | .780 (.077) | .737 (.090) | .878 (.049) |
|  | [.718 to .933] | [.683 to .927] | [.628 to .917] | [.558 to .899] | [.782 to .973] |

\* Las probabilidades a priori se calcularon utilizando la ecuación de Sidak 1-(1-a)^(1/n) para mantener la tasa base a priori agregada constante en 0.5 (1 a 1) para facilitar un contraste intuitivo entre las condiciones 2RQ, 3RQ y 4RQ.

# Discusión

Los resultados de este proyecto, utilizando datos de un estudio de laboratorio controlado, ilustran aún más el potencial de los efectos de multiplicidad del SSIST, que llevan a tasas globales de precisión del criterio que difieren notablemente de la precisión tanto de los exámenes poligráficos de asunto múltiple como de los polígrafos de asunto único con una sola serie. Estos resultados, junto con los resultados previamente reportados por Nelson (2023), sugieren que las estrategias de pruebas seriadas pueden generar problemas de error acumulado, y no ofrecen una solución conveniente y simple al complejo problema de la exploración poligráfica de asunto múltiple. Análisis previos (Barland, Honts & Barger, 1989; Nelson, 2023) sugieren que los efectos de error acumulado no se evitan con este enfoque, sino que simplemente se reubican. Los datos de este análisis muestran que puede disminuir la precisión global como resultado de esta propuesta y sugieren la necesidad de una mayor exploración y estudio antes de embarcarse en la sustitución a gran escala de los formatos de prueba poligráfica se asunto múltiple para los que se pueden aplicar convenientemente correcciones estadísticas a las puntuaciones de corte y a los métodos de análisis.

Dos factores principales influyen en la precisión de las pruebas: 1) si se asume que los temas objetivo de las pruebas son o no independientes, y 2) el número de preguntas RQs (Nelson y Handler, 2017). Es importante destacar que el uso de más RQs puede aumentar la precisión o exactitud de los exámenes de asunto único, pero puede disminuir la precisión o exactitud de los exámenes asuntos múltiples. El presente estudio aborda la ambigüedad del SSIST - ¿el SSIST puede mantener la precisión de los polígrafos asunto único dentro del contexto de las pruebas de asuntos múltiples? Hasta el momento no hay evidencia publicada que apoye la idea de que esto es posible. Por el contrario, la evidencia disponible sugiere que los polígrafos en serie de una sola pregunta comparten fenómenos estadísticos bien conocidos en los exámenes de preguntas múltiples. Los polígrafos de asunto único en serie pueden estar sujetos a los efectos del error acumulativo, similares a los de los exámenes de asunto múltiple con las correspondientes disminuciones en la exactitud de la prueba en comparación con el uso de un examen de asunto único en serie. Hay algunos contextos en los que el aumento de la sensibilidad de la prueba y la reducción de los errores falsos negativos pueden ser un efecto deseable, incluso teniendo en cuenta la reducción de la especificidad de la prueba y la disminución global de la precisión.

Las pruebas exploratorias de asunto único no son una idea nueva dentro de la profesión poligráfica (Barland, Honts, & Barger, 1989), y en otros contextos de evaluación. Se han propuesto ya otras posibilidades para un polígrafo exploratorio de asunto único (Barland, Honts, & Barger, 1989; Krapohl et al. 2024)4. El interés de los profesionales de campo y los directores de programas justifica una mayor investigación sobre el SSIST, y sobre otros formatos poligráficos exploratorios de asunto único, utilizando estudios de laboratorio controlados para los que se pueda conocer el criterio para objetivos de prueba múltiples. La investigación futura debería abordar los asuntos relacionados con la selección de probabilidades previas apropiadas, el impacto de diferentes probabilidades previas sobre las probabilidades posteriores, y el uso potencial de correcciones estadísticas para manejar los efectos del error acumulativo. Se requiere más información para comprender las diferentes ventajas y desventajas del enfoque SSIST en comparación con los métodos más comunes de análisis y de pruebas ómnibus.

4 La diferencia importante entre la prueba exploratoria de asunto único descrita por Krapohl et al., 2020 y el SSIST es que Krapohl et al. proponen el uso de un enfoque exploratorio de asunto único en un sola serie, que evita el potencial del efecto de error acumulativo.

Aunque el objetivo del SSIST puede ser aprovechar los tamaños del efecto de precisión de un formato de prueba de asunto único en el contexto de la exploración de asuntos múltiples, en este momento los datos indican que los tamaños del efecto de precisión para los exámenes de series múltiples no son equivalentes a los tamaños del efecto para los exámenes de asunto único en una sola serie.

# Referencias

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph, 2011, 40(4):* 195 305. [Electronic version] Retrieved January 15, 2023, from <http://www.polygraph.org.>

American Polygraph Association. (Aug. 25, 2023). *APA Standards of Practice.* Retrieved from [https://www.polygraph.or](http://www.polygraph.org/)g on Nov. 7, 2023.

Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the Accuracy of Security Screening Polygraph Examinations.* Department of Defense Polygraph Institute.

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

Krapohl, D. J. Grubin, D., Benson, T., & Morris B. (2020). ModiÀcation of the AFMGQT to accommodate single-issue screening. the British one-issue screening test. *Polygraph & Forensic Credibility Assessment, 49(2):* 176-183.

Nelson, R. (2015). Bonferroni and Sidak corrections for multiplicity effects with subtotal scores of comparison question polygraph tests. *Polygraph, 44(2):* 162-167.

Nelson, R. (2023). Multiplicity Effects in the Serial Single-issue Testing Situation: When is a Single-issue Test Not a Single-issue Test? *Polygraph & Forensic Credibility Assessment, 52(2):* 129-147.

Nelson, R., & Handler, M. (2017). Practical polygraph: how to select a polygraph test format.

O’Burke, J. P., Westerman, D., & Goodson, W. (2022). Using single-issues in screening examinations. *APA Magazine, 55(6):* 51-59.

Prado. R., Grajales, C., & Nelson, R. (2015). Laboratory study of directed lie polygraphs with Spanish speaking examinees. *Polygraph, 44(1):* 79-90

Prado. R., Grajales, C., & Nelson, R. (2015). Laboratory study of a diagnostic polygraph technique in a single sequence: a replication study. *Polygraph, 44(1):* 1-12.