

**Análisis Extendido de los Datos del Examen USAF
MGQT de Senter, Waller y Krapohl con el Sistema de
Puntuación Empírica y el Sistema de Puntuación
Objetiva, versión 3**

Raymond Nelson y Benjamin Blalock

Traductor

rodolfo@poligrafia.com.mx

Análisis Extendido de los Datos del Examen USAF MGQT de Senter, Waller y Krapohl con el Sistema de Puntuación Empírica y el Sistema de Puntuación Objetiva, versión 3

Raymond Nelson y Benjamin Blalock

Abstract

Los autores proporcionan un análisis extendido de los datos de validación para el formato USAF MGQT para pruebas de PDD, utilizando datos de archivo de un estudio de laboratorio anterior. El análisis multivariado no encontró diferencias significativas entre las puntuaciones totales y subtotales de la muestra de laboratorio y de una muestra de casos de investigación de campo que utilizan la misma técnica. Se muestra el perfil de precisión de criterio para los modelos TDA ESS y OSS-3, que incluyen la media, las desviaciones estándar y los intervalos de confianza estadísticos para la precisión de la decisión con y sin resultados inconclusos, errores para casos veraces y con engaño, y tasas de inconclusos para casos de engaño y veraces.

Introducción

La Técnica de Preguntas General Modificada por la Fuerza Aérea de los Estados Unidos (USAF-MGQT) (DoDPI, 2006) es una variante moderna de la familia de Técnicas con Preguntas de Comparación (CQT) que surgió por las modificaciones de la Técnica de Preguntas General (Reid, 1947) y de la Técnica de Comparación de Zonas (Backster, 1963). El USAF-MGQT se ajusta a los principios válidos generalmente aceptados para la construcción de pruebas de detección psicofisiológica de engaño (PDD) (Krapohl, 2006), y existe en dos versiones estrechamente relacionadas para las cuales no hay evidencia publicada ni hipótesis convincentes que sugieran que las diferencias sean sustantivas o podrían tener algún efecto sobre la exactitud del criterio. El USAF-MGQT a menudo se utiliza en contextos de investigación de facetas múltiples, para investigar múltiples roles o niveles de participación en un solo incidente o alegato conocido, y en contextos de selección de asuntos múltiples, para los cuales es posible que un examinado pueda estar involucrado en uno o más de los comportamientos distintos de preocupación mientras está no involucrado por completo en otros.

Senter, Waller y Krapohl, (2008), utilizando un escenario simulado de bombardeo en la carretera, informaron un nivel de precisión de criterio con puntuación de 7 posiciones a ciegas promedio de .849, excluyendo los resultados inconclusos. El presente estudio es una investigación sobre el uso del USAF-MGQT cuando se califica mediante un protocolo de puntuación basado en evidencia, el Sistema de Puntuación Empírica (ESS por sus siglas en inglés) (Blalock, Cushman & Nelson, 2009; Handler, Nelson, Goodson & Hicks, 2011; Krapohl, 2010; Nelson, Blalock, Oelrich & Cushman, 2011a; Nelson et al., 2011b; Nelson, Krapohl & Handler, 2008) y el algoritmo computarizado Sistema de Puntuación Objetiva, versión 3 (OSS-3 por sus siglas en inglés) (Nelson, Krapohl & Handler, 2008).

Método

Los datos de archivo se obtuvieron del estudio de laboratorio (N = 69) de Senter et al. (2008) y se calificaron mediante una versión automatizada del ESS y con el algoritmo computarizado OSS-3. El artículo de Senter et al. (2008) proporciona una descripción completa de la metodología de su estudio, y aquí solo se presenta una breve descripción. La muestra consistió en exámenes de laboratorio confirmados, para los cuales 36 de los exámenes se realizaron en personas programadas como veraces, mientras que 33 exámenes se realizaron en personas programadas como engañosas. Todos los exámenes se realizaron con la USAF-MGQT. Todos los exámenes constaron de dos preguntas objetivo de investigación, una relacionada con la participación directa y otra relacionada con la participación secundaria o indirecta en el incidente simulado, junto con tres preguntas de comparación.

Calificamos los datos utilizando una versión automatizada del modelo TDA del ESS, que incluye la medición automatizada de las características fisiológicas, la transformación automatizada de las puntuaciones de los puntos enteros y la ejecución automatizada de las reglas de decisión. Las puntuaciones del ESS se asignaron mediante la comparación de cada pregunta objetivo de la investigación contra la pregunta de comparación anterior o posterior que presentara la mayor reacción para cada sensor componente. Las puntuaciones de corte de decisión se establecieron con un alfa = .05 para clasificaciones de engaño y un alfa de = .1 para clasificaciones veraces.

Las puntuaciones de corte correspondientes para estos niveles alfa fueron -3 y +1, lo que significa que cualquier puntuación subtotal de -3 o inferior sería estadísticamente significativa para engaño ($p < .05$), mientras que los resultados de las pruebas en los que todas las puntuaciones subtotales son +1 o mayores serían estadísticamente significativas para veracidad ($p < .1$). La corrección Bonferonni para la puntuación de corte alfa para clasificaciones de engaño, no se utiliza en los exámenes PDD para los cuales se supone que las preguntas objetivo de investigación son independientes. Sin embargo, una inversa de la corrección de Sidak para problemas independientes se usa para corregir la deflación del alfa que ocurre al calcular la probabilidad normativa de que un examinado produzca un resultado estadísticamente significativo de veracidad para todos los objetivos de la investigación al tiempo que miente a uno o más de los asuntos independientes.

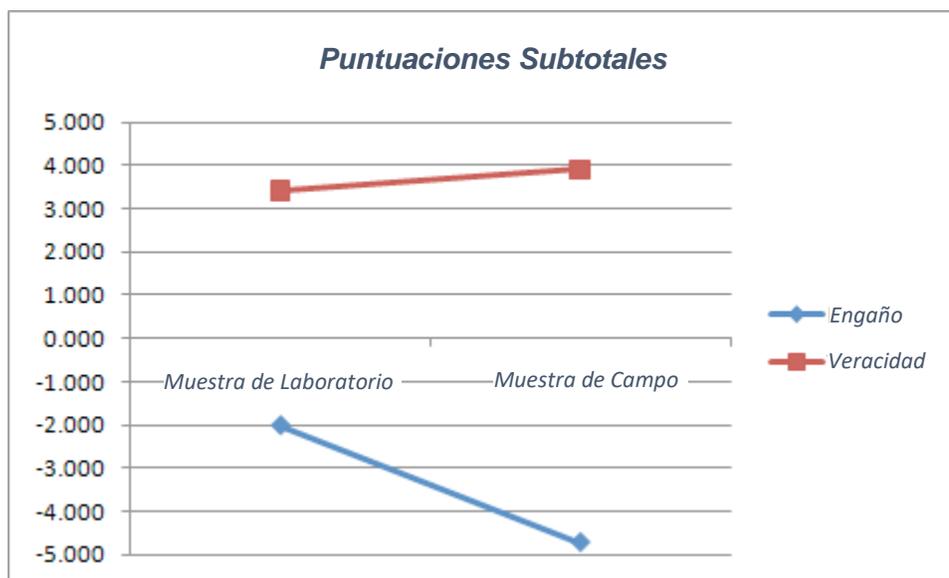
Los datos de muestra también se calificaron utilizando el algoritmo computarizado OSS-3 con un alfa = .05 para clasificaciones de engaño y un alfa = .1 para clasificaciones veraces. Mientras que los procedimientos de puntuación manuales y automatizados para el ESS involucran la asignación de puntuaciones enteras al comparar la fuerza de reacción de cada pregunta relevante contra la más fuerte de las preguntas de comparación vecinas, el algoritmo OSS-3 compara cada pregunta relevante contra el promedio de todas las preguntas de comparación, para cada sensor de componente, para asignar puntajes numéricos en forma de **ratios registrados** estandarizados. La regla de decisión para el modelo ESS automatizado y del OSS-3 era la regla de puntuación por spot (SSR) (Departamento de Defensa, 2006), y es idéntica a la que se usa al puntuar de forma manual los exámenes de facetas múltiples y de asuntos múltiples.

Resultados

Todos los resultados se evaluaron con un nivel de significancia establecido con un alfa $=.05$

Puntuaciones ESS. La puntuación subtotal media para casos de engaño fue de -2.00 (DS = 5.03), y el subtotal medio para los casos veraces fue de 3.42 (DS = 3.47). Se utilizó un ANOVA no balanceado bidireccional (muestra x estatus del caso) para comparar la distribución de las puntuaciones del ESS con las de un estudio basado en una muestra de casos confirmados (N = 22) utilizando el USAF-MGQT (Nelson, Blalock y Handler, 2011). Hubo un efecto de interacción significativo para las distribuciones de la muestra y las puntuaciones subtotales ($F_{1,87} = 541.557, p < .001$). La Figura 1 muestra que la muestra de campo (N = 22) produce puntuaciones subtotales de mayor valor absoluto que la muestra de laboratorio (N = 66), cuando se sometió a una medición automatizada. Los ANOVAs de una vía desbalanceados post-hoc demostraron que las diferencias no fueron estadísticamente significativas para las puntuaciones subtotales. Estos resultados sugieren que las distribuciones muestrales de puntajes de laboratorio de engaño y veracidad no difieren significativamente a pesar de que los puntajes de la muestra de campo estuvieron más lejanos del cero.

Figura 1. Puntuaciones subtotales medias para muestras de laboratorio y de campo USAF MGQT.



Validez de Criterio. La Tabla 1 muestra los porcentajes medios, las desviaciones estándar y los intervalos de confianza estadísticos para un perfil dimensional de precisión de criterio para la muestra de laboratorio de los exámenes USAF-MGQT, que incluyen: sensibilidad, especificidad, resultados inconclusos para casos de engaño y veracidad, errores falso-positivo y falso-negativo, valor predictivo positivo, valor predictivo negativo, porcentaje de

decisiones correctas para los casos de engaño y veraces, y los promedios no ponderados del porcentaje de resultados correctos y de inconclusos para casos de engaño y veracidad. Se muestran los datos para las puntuaciones ESS y para el algoritmo computarizado OSS-3.

Media, Desviaciones Estándar e Intervalos de Confianza al 95% para el ESS Manual, ESS Automatizado y las DS y IC del Algoritmo OSS-3		
	ESS Automatizado	OSS-3
Precisión no Balanceada	.838 (.040) {.758 to .918}	.856 (.038) {.78 to .931}
Inc no Balanceados	.153 (.033) {.087 to .220}	.154 (.036) {.082 to .226}
Sensibilidad	.511 (.071) {.371 to .65}	.667 (.068) {.532 to .802}
Especificidad	.940 (.034) {.872 to .999}	.782 (.058) {.667 to .898}
Error FN	.211 (.059) {.094 to .329}	.148 (.050) {.050 to .246}
Error FP	.029 (.024) {.001 to .078}	.092 (.042) {.009 to .175}
Inc D	.277 (.064) {.151 to .402}	.184 (.055) {.075 to .292}
Inc T	.030 (.024) {.001 to .078}	.125 (.045) {.035 to .215}
PPV	.945 (.045) {.857 to .999}	.879 (.054) {.773 to .984}
NPV	.815 (.052) {.712 to .918}	.840 (.053) {.736 to .944}
D Correctos	.707 (.077) {.554 to .859}	.817 (.061) {.697 to .937}
T Correctos	.969 (.025) {.919 to .999}	.894 (.048) {.800 to .988}

Discusión

Estos resultados apoyan el hipótesis de que tanto los puntajes ESS y OSS-3 de las exámenes US-AF-MGQT pueden diferenciar los casos confirmados veraces de los casos confirmados con engaño, con tasas que son significativamente mayores que el azar.

Por supuesto, es realista la generalización de los resultados del estudio a los ambientes de campo, solo cuando el examen se realiza de manera competente, y siempre y cuando los datos del examen sean de calidad satisfactoria, suficientemente libres de artefactos no interpretables.

La limitación principal del presente estudio involucra el pequeño **cohorte** de evaluadores y el tamaño de la muestra relativamente pequeño. A pesar de estas limitaciones, argumentamos que el **cohorte** de evaluadores, que incluyó un examinador experimentado y dos aprendices sin experiencia, representa y generaliza a los entornos de campo del mundo real de manera más efectiva que los resultados de estudios basados en las puntuaciones de un solo experto. Los resultados del análisis multivariado no encontraron diferencias significativas con las distribuciones de veracidad y engaño de las puntuaciones de las muestras de laboratorio y de campo, aunque la muestra de campo produjo puntuaciones de magnitud absoluta más fuerte (es decir, más lejos de cero).

Estudios pasados no habían mostrado soporte para la hipótesis de que el examen multifacético puede diferenciar efectivamente el engaño de la verdad al nivel de pregunta individual (Barland, Honts & Barger, 1989a; Barland, Honts & Barger, 1989b; Podlesny & Truslow, 1993), y este estudio no intentó diferenciar la veracidad del engaño al nivel de pregunta individual. En cambio, los resultados de las pruebas para los casos de muestra se clasificaron a nivel de la prueba como un todo, a pesar del hecho de que las reglas de decisión involucraban los subtotales de las preguntas de prueba individuales. Como suele ser el caso, se necesita investigación adicional en esta área.

Estos resultados sugieren un interés continuo en el formato USAF-MGQT y un interés continuo en los modelos TDA ESS y OSS-3.

Referencias

- Backster, C. (1963). Standardized polygraph notepack and technique guide: Backster zone comparison technique. Cleve Backster: New York.
- Barland, G. H., Honts, C. R. & Barger, S.D. (1989a). The validity of detection of deception for multiple issues . *Psychophysiology*, 26, 13 (Abstract).
- Barland, G. H., Honts, C. R. & Barger, S.D. (1989b). Studies of the accuracy of security screening polygraph examinations. DTIC AD Number A304654. Department of Defense Polygraph Institute.
- Blalock, B., Cushman, B. & Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph*, 38, 281-288.
- Department of Defense (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Reprinted in *Polygraph*, 40(1), 2-66.

- Handler, M., Nelson, R., Goodson, W. & Hicks, M. (2011). Empirical Scoring System: A Cross-cultural Replication and Extension Study of Manual Scoring and Decision Policies. *Polygraph*, 39, 200-215.
- Krapohl, D. (2010). Short Report: A Test of the ESS with Two-Question Field Cases. *Polygraph*, 39, 124-126.
- Krapohl, D. J. (2006). Validated polygraph techniques. *Polygraph*, 35(3), 149-155.
- Nelson, R., Handler M., Morgan C., & O'Burke P. J. (2012). Short Report: Criterion Validity of the United States Air Force Modified General Question Technique and Iraqi Scorers. *Polygraph*. 41(1), 18-28.
- Nelson, R., Blalock, B., Oelrich, M. & Cushman, B. (2011). Reliability of the Empirical Scoring System with Expert Examiners. *Polygraph*, 40(3), 131-139.
- 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., Krapohl, D. & Handler, M. (2008). Brute Force Comparison: A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers. *Polygraph*, 37, 185-215.
- Podlesny, J. A. & Truslow, C.M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78, 788-797.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542-547.
- Senter, S., Waller, J. & Krapohl, D. (2008). Air Force Modified General Question Test Validation Study. *Polygraph*, 37(3), 174-184.