

Reporte del Comité de Investigación APA:  
Propuesta de Uso para el Formato de Prueba AFMGQT<sup>1</sup> de Evento-  
Específico

Raymond Nelson, Mark Handler, Marty Oelrich y Barry Cushman

Rodolfo Prado Pelayo

Traductor

[rodolfo@poligrafia.com.mx](mailto:rodolfo@poligrafia.com.mx)

## Reporte del Comité de Investigación APA:

### Propuesta de Uso para el Formato de Prueba AFMGQT<sup>1</sup> de Evento-Específico

Raymond Nelson, Mark Handler, Marty Oelrich y Barry Cushman

#### Abstract

Los miembros del comité de investigación de la APA responden un cuestionamiento acerca de las modificaciones al formato AFMGQT para exámenes de evento específico. Con respecto a la posibilidad de generalizar los conocimientos científicos disponibles actuales con respecto a los formatos de prueba de evento específico con dos o tres preguntas relevantes, revisamos la evidencia publicada acerca de la validez de la hipótesis de super-dampening, y de la efectividad de las preguntas de asunto-externo o "sintomáticas" para reducir los resultados inconclusos, aumentar la precisión de la prueba, o detectar la presencia de asuntos externos. Tomamos en consideración si es que la inclusión de preguntas de asunto-externo afecta tanto a las distribuciones de los resultados de prueba o a la precisión de criterio de los resultados de prueba. La evidencia empírica acerca de la efectividad de las preguntas de asunto-externo es confusa y por lo tanto no informativa. La investigación sobre preguntas de asunto-externo y de la hipótesis del super-dampening no ha demostrado de manera concluyente sus efectos hipotéticos y los formatos de preguntas de prueba que no incluyen preguntas de asunto-externo han demostrado que producen tasas de precisión media equivalentes o superiores a las de los formatos de preguntas de prueba que sí incluyen estas preguntas. No se encontraron bases o evidencias que apoyen la conclusión de que nuestro conocimiento y datos normativos actuales no puedan generalizarse para el AFMGQT si se Outiliza como un formato de prueba de evento-específico con dos o tres preguntas relevantes.

Los miembros del Comité de Investigación de la American Polygraph Association (APA) han recibido preguntas acerca de la validez del uso de la Técnica Air-Force Modified General Question Technique (AFMGQT) - Técnica Modificada de Preguntas Generales de la Fuerza Aérea – con formatos de preguntas V1 y V2 (Department of Defense, 2006a) como formato de preguntas de evento-específico de asunto-único. En concreto, estas preguntas se han centrado el uso potencial de los formatos de preguntas AFMGQT de manera similar al formato Federal You-Phase -Federal Fase-Usted-(You-Phase) y del formato Federal Zone Comparison Test (ZCT) – Prueba Federal de Zonas de Comparación (ZCT) (Department of Defense, 2006a). Estas preguntas se refieren a la posibilidad de generalizar los datos normativos actualmente disponibles y la evidencia publicada de la validez de criterio. Estas consultas se dirigen a si nuestro conocimiento actual es suficiente para cuantificar el grado de incertidumbre y el nivel de confianza para los resultados de prueba obtenidos mediante el formato AFMGQT como un formato de prueba de evento específico de asunto-único. En otras palabras, ¿Podrían generalizarse nuestros datos normativos de un ZCT a un formato AFMGQT?

<sup>1</sup>Este reporte es un producto del Comité de Investigación de la APA. Se presentó a la Mesa Directiva de la APA, quien lo aceptó. Deberá considerarse como un reporte de investigación, y no representa una posición oficial de la American Polygraph Association -Asociación Americana de Poligrafía-.

El uso propuesto del AFMGQT incluye dos (2) preguntas relevantes (RQs) rodeadas en secuencia por tres (3) preguntas de comparación (CQs), (es decir, CQ, RQ, CQ, RQ, CQ). El formato también incluye preguntas no-diagnósticas en forma de una pregunta neutral y una relevante de sacrificio. Las dos RQs del formato propuesto deberán ser formuladas para dirigirse a los detalles del asunto único de preocupación, de manera similar a la selección de objetivos y practicas comunes de formulación de preguntas de la técnica Federal You-Phase. La You-Phase utiliza una secuencia similar de preguntas calificables, con la adición de dos preguntas de asunto-externo.

Se ha sugerido una propuesta de secuencia similar, utilizando el formato AFMGQT para contextos de prueba de evento-específico con tres (3) RQs en una secuencia con tres (3) CQs, (es decir, CQ, RQ, CQ, RQ, CQ, RQ). El enfoque propuesto para la selección del objetivo y de la formulación de pregunta sería similar a la del ZCT, una secuencia de preguntas de prueba que puede incluir preguntas relevantes tanto primarias (es decir, participación directa) secundaria (es decir, participación indirecta), y preguntas de procedimiento que constan de una pregunta neutral y una relevante de sacrificio. Las preguntas de prueba podrían seleccionarse y formularse utilizando los principios comúnmente empleados con el formato ZCT con la excepción de la exclusión de las preguntas de asunto-externo.

Los formatos You-Phase y ZCT se utilizan en contextos de diagnóstico de evento-específico. También se utilizan en contextos exploratorios de asunto-único, como un formato exploratorio independiente de asunto-único, o como una prueba de asunto-único de desglose para ser utilizado después de un examen exploratorio de asuntos-múltiples no resuelto. Ya sea que se utilice en un contexto exploratorio de un asunto-único o en un contexto de investigación de eventos-específicos, las preguntas de la prueba de los formatos You-Phase y ZCT se interpretan bajo el supuesto de un criterio de varianza no-independiente (Department of Defense, 2006b). Eso significa que la conducta que afecta el estado de criterio (es decir, culpable o inocente) de cualquier pregunta objetivo individual, también podría afectar el estado de criterio de otras preguntas objetivo. Las suposiciones acerca de la varianza independiente y no independiente se expresan en las prácticas de campo mediante la selección de reglas de decisión y datos normativos que se utilizan para analizar el resultado del examen y para determinar el nivel de significancia estadística.

El uso tradicional del AFMGQT ha sido el de un formato multi-faceta para la investigación de problemas conocidos o alegados, y como un formato exploratorio de asunto-múltiple. La calificación e interpretación de los datos de prueba del AFMGQT tradicionalmente se ha realizado bajo el supuesto de que el estado de criterio de cada pregunta objetivo varía de forma independiente (American Polygraph Association, 2001; Department of Defense, 2006b). Esto tiene como significado la suposición de que el estado de criterio de las preguntas objetivo individuales es distinto al de los estados de criterio de otros objetivos del examen (inocente y culpable). Una diferencia importante en el uso propuesto del formato AFMGQT es la selección y formulación de las preguntas objetivo del AFMGQT en el cual se asume que el estado de criterio de los estímulos objetivo varía de forma no-independiente y por lo tanto se evalúan utilizando la puntuación de gran total.

Las investigaciones no han abordado específicamente este uso particular del AFMGQT, por lo que cuestionamientos científicos acerca de la exactitud de criterio y datos normativos dependerán de una evaluación de nuestro conocimiento con respecto a los formatos existentes de examen de evento-específico. ¿Es científicamente razonable y responsable el generalizar el conocimiento de los formatos You-Phase y ZCT para la propuesta de uso del formato AFMGQT? Los resultados de las pruebas poligráficas podrían jugar un papel en la toma de decisiones, que podrían afectar la seguridad de la comunidad, las libertades individuales, las oportunidades futuras, y la seguridad de la información. Cualquier conclusión relacionada con estos asuntos estará mejor informada mediante una evaluación cuidadosa de la evidencia científica y no por mera especulación. La importancia práctica de esta investigación consiste en si los profesionales de campo que utilicen esta modificación, cumplen con las expectativas de una práctica basada en evidencia y del uso de técnicas poligráficas validadas. Si no es así, el público podría tener una expectativa razonable para la notificación de su uso como una técnica experimental, como un componente del consentimiento informado y de la autorización para la prueba.

El Comité de Investigación de la APA adopta la posición de que la respuesta a las preguntas relacionadas con la validez del uso propuesto del AFMGQT se encuentra en la evidencia perteneciente a dos asuntos fundamentales: 1) si es que las preguntas de asunto externo incrementan la efectividad de la prueba poligráfica, y 2) si es que el conocimiento actualmente disponible acerca de la exactitud de criterio y de las distribuciones normativas de puntuaciones de personas inocentes y culpables son generalizables y suficientes para cuantificar el grado de incertidumbre asociado con los resultados de la prueba.

### Discusión

En el núcleo de la presente investigación se encuentra la cuestión fundamental con respecto a si nuestro conocimiento actual es suficiente para permitirnos generalizar el grado de incertidumbre y el nivel de confianza para el rango de precisión esperado en los exámenes You-Phase y ZCT. La respuesta a esta pregunta se determina de mejor manera por la presencia o ausencia de diferencias significativas entre el uso propuesto de los formatos AFMGQT y los formatos más tradicionales del ZCT y You-Phase. Una perspectiva cautelosa sostendría que la existencia de una diferencia significativa requeriría de un estudio adicional antes de proceder con la aprobación del uso propuesto del AFMGQT. En este caso, la noción de diferencias "significativas" debe definirse como cualquier diferencia para la que la evidencia ha demostrado que afecta tanto a la distribución de las calificaciones numéricas de la prueba, como a la precisión de criterio de los resultados de prueba. Adicionalmente, cualquier hipótesis claramente articulada que defina un mecanismo plausible cuyos resultados numéricos de prueba o su exactitud criterio se esperaría que pudieran diferir, podría ser la base para ser cautelosos y realizar estudios adicionales antes de proceder a aprobar el uso propuesto del AFMGQT.

El uso propuesto del AFMGQT como un formato de prueba de evento específico depende de la cuestión de si los datos normativos disponibles para los exámenes You-Phase y ZCT se pueden generalizar para el AFMGQT si es que la selección de objetivos, la formulación de preguntas y los métodos analíticos de datos de prueba son similares a los de los formatos con dos o tres preguntas

relevantes. Los datos normativos se utilizan para calcular el grado de incertidumbre asociado con un resultado de prueba, y para calcular los cortes numéricos de decisión que determinan el nivel de significancia estadística de conclusiones categóricas basadas en datos de prueba probabilísticos. Existen datos normativos, tanto para el formato Federal You-Phase y para los formatos de prueba Federal Zone Comparison Test, incluyendo estadísticas normativas para calificaciones de 7-posiciones (American Polygraph Association, 2011) y calificaciones del Sistema de Puntuación Empírica (ESS) (American Polygraph Association, 2011; Nelson et al., 2011), basados en el supuesto del criterio de varianza no independiente entre las preguntas objetivo.

Existen algunas diferencias en el uso tradicional del AFMGQT y de los formatos You-Phase y ZCT. Dentro de estas diferencias está el enfoque de la selección y formulación de las preguntas objetivo, y las suposiciones acerca del criterio de varianza de independencia o de no independencia cuando se califican los datos de los exámenes. Se encuentran disponibles datos normativos para el formato AFMGQT cuando las preguntas de estímulo de prueba se califican e interpretan bajo el supuesto de una varianza de criterio independiente (American Polygraph Association, 2011). El uso propuesto eliminaría esta diferencia importante.<sup>2</sup>

Se mantendrían las diferencias en las posiciones numéricas de cada pregunta dentro de la secuencia<sup>3</sup>, aunque no existe alguna razón plausible que sugiera que cualquier diferencia en los números de las preguntas o en los cambios de ubicación propuestos podrían afectar las características del rendimiento de la prueba.

Otra diferencia entre el uso propuesto del formato AFMGQT y el uso tradicional de los formatos You-Phase y ZCT es la exclusión de las preguntas de asunto-externo, referidas a los examinadores como preguntas "sintomáticas". Dos ejemplos de preguntas de asunto-externo pueden encontrarse en las publicaciones del Department of Defense (2006), que incluyen: 1) "¿Cree usted que solamente le voy a hacer las preguntas que revisamos?" y 2) "¿Hay algo más que le preocupe que pudiera yo preguntarle?"<sup>4</sup>

La exclusión de las preguntas de asunto-externo de las secuencias propuestas de preguntas de prueba de evento específico AFMGQT merece una mayor discusión.

<sup>2</sup> El uso propuesto del AFMGQT no intenta prevenir la continuación del uso del AFMGQT en su forma tradicional, cuando las circunstancias indiquen que sea útil el supuesto la varianza de criterio independiente.

<sup>3</sup> Las preguntas relevantes se localizan en la posición 5 y 7 en los formatos You-Phase y en las posiciones 5, 7 y 10 en los formatos ZCT, mientras que estas preguntas deberán de ser localizadas en las posiciones 4, 6 y 8 en el uso propuesto del AFMGQT.

<sup>4</sup> Se han descrito además otras versiones de las preguntas de asunto externo.

La inclusión de preguntas de asunto-externo en la secuencia de preguntas de prueba es la expresión práctica de la incorporación de la hipótesis<sup>5</sup> del super-dampening en el modelo de preguntas de comparación.<sup>6</sup>

La hipótesis del super-dampening sugiere que si los asuntos fuera del alcance del (los) objetivo(s) de investigación, son de mayor preocupación para el examinado que las preguntas objetivo de investigación, serán la causa de resultados inconclusos (Backster, 1962). La exclusión de las preguntas de asunto-externo representa la exclusión de la hipótesis de super-dampening del modelo. Un corolario de la hipótesis del super-dampening es que el uso de preguntas de prueba de asunto-externo, permitirá a un examinador poder detectar la existencia de un asunto externo. Se encuentra implícita en cada hipótesis la expectativa de incrementar la eficiencia del modelo. En la presente investigación, se debe observar la evidencia acerca de una mayor efectividad del modelo en términos de la dimensión de la exactitud de criterio (por ejemplo, sensibilidad, especificidad, tasas de error, o la precisión de decisión global) como resultado de un decremento en la tasa de resultados inconclusos. Adicionalmente, está implícita en la inclusión de cualquier hipótesis la descripción de cómo operacionalizar y cuantificar cualquier efecto anticipado.

Una hipótesis es valiosa, y por lo tanto aceptable, si ofrece cualquier incremento en la efectividad del modelo en el que se aplica.

<sup>5</sup> La hipótesis del super-dampening es referida por Backster (2001) como el “concepto super-dampening concept.” El término “concepto” en general no se usa de esta forma en el ámbito científico. Para que cualquier idea, explicación o sugerencia tenga el mérito de ser considerada como hipótesis científica debe declararse para que pueda ser falseada y probada – lo que requiere una forma de medición objetiva, o al menos una tendencia hacia una medición objetiva en la forma de una observación clínica publicada que pueda ser fácilmente replicada. Las hipótesis que se ofrecen de manera tal que no pueden ser probadas y por lo tanto no pueden ser falseadas, son vistas como hipótesis defectuosas, que por lo tanto son hipótesis falsas – además del riesgo de involucrarnos en prácticas pseudo-científicas. Las hipótesis son preguntas experimentales o de investigación, que tienen que ser tomadas con precaución hasta que exista evidencia replicada que de soporte a su validez. Las teorías científicas son aquellas hipótesis de trabajo para las que existe evidencia replicada que da soporte a una idea o explicación. Entonces, las teorías son hipótesis donde la evidencia sugiere que son correctas y que no son inconsistentes con otra evidencia. Ya que no es posible saberlo todo, los científicos están obligados a seguir aprendiendo. El rol de la ciencia es seguir estudiando nuestras ideas y comparar nuestro conocimiento actual con los supuestos de una base de datos siempre en expansión. La exploración curiosa de los límites del conocimiento y de la precisión de nuestros supuestos, eventualmente darán como resultado la observación de evidencia que es inconsistente con las explicaciones de nuestras teorías de trabajo actuales – momento en el que la obligación será modificar la hipótesis de manera tal que pueda explicar tanto a la evidencia nueva como a la vieja, o simplemente descartar la hipótesis como incorrecta y volver a comenzar con la hipótesis nueva que explica tanto a la evidencia vieja como la nueva. Las leyes de la ciencia son aquellas ideas para las que la evidencia y las réplicas son tan fuertes y robustas que se espera que no habrá evidencia nueva que controvierta su validez. El “concepto” super-dampening en realidad es una hipótesis de super-dampening.

<sup>6</sup> El uso de este término transmite que la prueba poligráfica no evalúa mentiras per se, y que en su lugar prueba la significancia estadística de error mediante el registro y calificación de data proxy, en forma de diferencia en la fuerza de las respuestas fisiológicas, como indicadores de la saliencia, ante los diferentes tipos de estímulo de prueba que están combinados de forma matemáticamente óptima para representar o “modelar” el acto de engañar o decir la verdad con un alto rango de precisión, que es significativamente mayor que el azar. El objetivo de las pruebas científicas – cuando no es posible una observación o medición del fenómeno de interés – es cuantificar el grado de incertidumbre asociado con una conclusión.

Las hipótesis falsas no tienen valor alguno y deben ser eliminadas y reemplazadas o sometidas a modificaciones hasta que una nueva hipótesis sea soportada por evidencia replicable. Este es un aspecto inherente al método científico. El requisito para la aceptación de una hipótesis es simple, aunque no necesariamente fácil de lograr: debe existir evidencia replicable para demostrar que los fenómenos o datos observados realmente funcionan como se espera.

En el caso presente, la hipótesis de super-dampening pretende explicar la ocurrencia de resultados de pruebas inconclusos, y las preguntas de asunto-externo pretenden ya sea detectar o reducir la interferencia que podrían imponer los asuntos externos en los datos de la prueba y en el resultado de la prueba. La medición global de la validez de la hipótesis de super-dampening y de las preguntas de asunto-externo es si la precisión de criterio de los resultados de la prueba poligráfica se mejoran en términos de precisión de discriminación, (es decir, sensibilidad de prueba y especificidad de prueba) o de errores (es decir, errores falso positivo y falso negativo). Un tema relacionado es si se puede esperar que la incorporación de la hipótesis del super-dampening y de las preguntas de asunto externo dentro de la estructura de la prueba pueda dar cuenta de un cambio causal en las distribuciones de las puntuaciones de los culpables o inocentes.

Si la hipótesis del super-dampening es una hipótesis verdadera o correcta, se podría esperar entonces que el rechazo de la hipótesis del super-dampening y de las preguntas de asunto-externo de la estructura de prueba, pudiera afectar negativamente la distribución de las puntuaciones numéricas y la precisión de criterio de los resultados de prueba. Esto entonces afectaría la generalización de nuestro conocimiento acerca de la distribución normativa de calificaciones de personas culpables e inocentes en formatos de prueba que no incluyen preguntas de asunto-externo. Si la hipótesis del super-dampening es falsa o incorrecta (es decir, las preguntas de asunto-externo no tienen efecto en los resultados de prueba o en la exactitud de prueba), entonces se podría esperar que la exclusión de las preguntas de asunto-externo de la secuencia de preguntas de prueba no tendría ningún efecto sobre las distribuciones de las puntuaciones numéricas ni en el criterio de precisión de los resultados de prueba. Por lo tanto, es fundamental entender las bases de la evidencia científica de la hipótesis del super-dampening y de las preguntas de asunto-externo.

Revisión de la literatura científica acerca de la hipótesis del super-dampening.

Capps, Knill y Evans (1993) llevaron a cabo un estudio de las preguntas de asunto-externo con tres examinadores experimentados con licencia en sus respectivos estados, y miembros de la American Polygraph Association. Los examinadores, con experiencia promedio de más de 5 años, acordaron llevar a cabo cualquier examinación utilizando un formato ZCT que incluyera preguntas de asunto externo o sustituyéndolas con preguntas neutrales. Se recolectaron un total de N = 150 exámenes de campo, que incluyeron 75 exámenes ZCT con preguntas de asunto externo y 75 exámenes ZCT sin preguntas de asunto externo. Los datos fueron evaluados por los examinadores originales quienes reportaron un total de 12 casos inconclusos cuando se excluyeron las preguntas de asunto externo y cuatro (4) resultados inconclusos cuando se incluyeron preguntas de asunto externo.

Capps, Knill y Evans (1993) reportaron los resultados de bondad de la prueba de ajuste chi cuadrada ( $X^2$ ) sugiriendo una diferencia estadísticamente significativa ( $X^2 [1] = 4.48, p = 0.034$ ).

Sin embargo, no reportaron qué métodos subjetivos o no cuantitativos utilizaron los examinadores originales para disminuir sus resultados inconclusos cuando incluyeron preguntas de asunto-externo. Los resultados de los examinadores originales son confusos por el hecho de que los examinadores pudieron haber confiado en alguna información extrapoligráfica, en la forma de observaciones interpersonales o de información de antecedentes del caso, para la formulación de sus conclusiones. La causa exacta de la reducción de preguntas inconclusas no puede ser concluyentemente atribuible a la formulación de las preguntas de asunto-externo. Capps, Knill y Evans reportaron también los resultados utilizando un algoritmo estadístico automatizado, que generó 8 resultados inconclusos cuando se excluyeron las preguntas de asunto-externo y 5 resultados inconclusos cuando se incluyeron preguntas de asunto-externo, pero no reportaron los resultados de la prueba estadística de las conclusiones utilizando las mediciones objetivas del algoritmo de calificación automatizado.

Honts, Amato y Gordon (2001) reanalizaron los datos de campo reportados por Capps, Knill y Evans (1993) y mostraron que las preguntas de asunto-externo dieron cuenta de aproximadamente el 3% de la variabilidad entre los resultados concluyentes vs. inconclusos. Debido a que Capps, Knill y Evans no reportaron los resultados de la prueba estadística de los resultados de examinación automatizados, Krapohl y Ryan (2001) re-evaluaron los resultados reportados y calcularon la estadística chi cuadrada utilizando el mismo método de contingencia 2x2, el cual mostró que la diferencia no fue estadísticamente significativa ( $X^2 [1] = 0.758, p = 0.384$ ), indicando que las diferencias en las tasas de inconclusos, cuando se incluyen o excluyen preguntas de asunto externo, no difieren del azar. Una diferencia importante entre los resultados automatizados y las calificaciones manuales es que el método de puntuación automatizado no podía hacer uso de la información extrapoligráfica, y se basó únicamente en las respuestas fisiológicas registradas. También es posible que los examinadores participantes pudieron saber o adivinar el propósito del experimento, y que los resultados observados fueron influenciados por estas expectativas. Aunque los resultados de Capps, Knill y Evans son de interés, estos resultados confusos no pueden ser aceptados como una base suficiente para concluir que las preguntas de asunto externo realmente ofrecen las ventajas hipotéticas.

Los resultados del estudio de Capps, Knill y Evans (1993) no fueron suficientes para rechazar la hipótesis nula de que la inclusión de preguntas de asunto externo no contribuye de forma significativa en la reducción de resultados inconclusos con base en las respuestas fisiológicas registradas. Tampoco podría el estudio de Capps, Knill y Evans rechazar la hipótesis nula con respecto a la capacidad de detectar la presencia o ausencia real de un asunto externo. Es muy importante, que ellos no investigaron la presencia o ausencia real de un asunto externo y por lo tanto no pudieron investigar si las preguntas de asunto-externo contribuyeron con la capacidad de discriminar la presencia o ausencia de asuntos externos. Por último, el diseño de la muestra impidió cualquier posibilidad de investigar si la hipótesis del super-dampening y de las preguntas de asunto-externo

incrementan la precisión de criterio de los resultados de la prueba poligráfica o de la distribución de los puntajes de los veraces y de los que engañan.

Honts, Amato y Gordon (2001) estudiaron la pregunta de asunto-externo utilizando un diseño factorial que incluyó personas programadas como inocentes y culpables, la existencia o inexistencia de un asunto externo, y la inclusión o exclusión de preguntas de asunto-externo dentro de la secuencia de preguntas de prueba. Habían  $N = 92$  participantes asignados al azar a las ocho condiciones distintas. Esta investigación fue diseñada para investigar el efecto de la hipótesis del super-dampening y de las preguntas de asunto-externo en la precisión de criterio de la prueba, y si es que existen diferencias en la forma en que las personas culpables e inocentes responden a las preguntas de asunto externo. También se dirigió a si los examinadores federales experimentados podrían utilizar las preguntas de asunto-externo para detectar la existencia de asuntos externos, y si es que las preguntas de asunto-externo podrían ser sujetas a un análisis formal. Los datos fueron analizados por tres instructores del Department of Defense Polygraph Institute (ahora llamado National Center for Credibility Assessment).

Honts, Amato y Gordon (2001) reportaron que la existencia de un asunto externo tuvo poco efecto en los resultados de las pruebas de las personas culpables, pero se asoció con más calificaciones negativas de prueba en los participantes inocentes. Los calificadores no tuvieron la capacidad de discriminar la existencia de asuntos externos con tasas superiores que el azar. Los investigadores calcularon un coeficiente de eficiencia de la decisión (Kircher, Horowitz y Raskin, 1988) para determinar la efectividad de la pregunta de asunto-externo para detectar asuntos externos, pero fueron incapaces de desarrollar una solución estadística o numérica formal para calificar o interpretar eficientemente las preguntas de asunto externo. Honts, Amato y Gordon reportaron que solamente se identificó una función discriminativa estadísticamente significativa en la forma de una diferencia entre las respuestas a las preguntas de asunto-externo y a las preguntas relevantes. Ellos reportaron una correlación canónica de .215 con una precisión de clasificación de .667 para los participantes con ausencia de asunto externo y .645 para los participantes con presencia de asunto externo, y reportaron que no esperaban que los resultados sobrevivieran a una validación cruzada.

Honts, Amato y Gordon (2001) reportaron que las preguntas de asunto-externo no contribuyeron en el incremento de la precisión criterio o en el aumento de la detección de la presencia de asuntos externos. Los resultados de una ANOVA multivariada indicaron que las preguntas de asunto-externo podrían ser utilizadas como preguntas de comparación válidas, aunque no existe un incidente anterior o posterior de esta propuesta en los protocolos de prueba de campo. Los resultados del estudio de Honts, Amato y Gordon no apoyaron la hipótesis del super-dampening y anotaron que "la preocupación de que los delitos no descubiertos podrían opacar las respuestas a las preguntas de prueba relevantes, parece ser infundada" (p. 73).

Un estudio de Krapohl y Ryan (2001) involucró  $N = 100$  exámenes de campo que fueron seleccionados de un archivo de casos confirmados por el Department of Defense y que fueron sujetos a una puntuación numérica de 7-posiciones a ciegas por parte de un examinador experimentado,

quien también realizó una anotación subjetiva acerca de la presencia o ausencia de asuntos externos para cada canal registrado y para cada pregunta. Los datos también fueron evaluados utilizando un método automatizado de orden de rangos aplicado a las preguntas relevantes, de comparación y de asunto externo. Krapohl y Ryan reportaron que prácticamente no hay correlación entre el número de reacciones a las preguntas de asunto-externo y las puntuaciones numéricas totales ( $r = .07$ ), y que prácticamente no hay correlación entre la puntuación de diferencia de rango (es decir, la diferencia en los rangos sumados de las preguntas relevantes y de comparación) y las puntuaciones de rango de las preguntas de asunto-externo ( $r = .06$ ). Igualmente, no hubo correlación entre las puntuaciones de diferencia de rango y las preguntas de asunto-externo para los casos de inocentes ( $r = .04$ ). Se encontró una correlación ligeramente más fuerte entre las puntuaciones de diferencia de rango y las preguntas de asunto-externo para los casos de los culpables ( $r = -.18$ ),<sup>7</sup> pero la relación no fue estadísticamente significativa. No hubo evidencia de amortiguación en las puntuaciones numéricas en función de las respuestas a las preguntas de asunto-externo y ni soporte para la hipótesis de super-dampening. Krapohl y Ryan llegaron a la conclusión de que su análisis controvierte la hipótesis del super-dampening.

### Conclusión

Esta revisión de la literatura publicada reveló que las preguntas de asunto externo han demostrado ser ineficientes para lograr el objetivo implícito en la redacción de estas preguntas (es decir, evaluar la presencia de asuntos externos que podrían explicar o reducir la aparición de un resultado de prueba inconcluso). No encontramos evidencia que apoye la conclusión de que las puntuaciones numéricas o la exactitud de la prueba varían en función de las respuestas a las preguntas de asunto-externo. Por consiguiente, concluimos que nuestro actual conocimiento con respecto a la precisión de criterio y a las distribuciones normativas de las puntuaciones se puede generalizar para los otros formatos de prueba mencionados si es que la selección del objetivo de la prueba y la formulación de preguntas, la presentación de estímulos y los protocolos de análisis de datos de prueba son similares. Concluimos además que nuestra base de conocimiento actual es suficiente para cuantificar el grado de incertidumbre de los resultados obtenidos con el uso propuesto del AFMGQT dentro del rango de efectividad mostrado en otras técnicas poligráficas existentes de evento específico con dos o tres preguntas relevantes, independientemente de la inclusión o exclusión de preguntas de asunto-externo que no son calificadas.

La aceptación de la hipótesis del super-dampening y la aprobación de los supuestos de que las preguntas de asunto-externo juegan un papel causal en los resultados inconclusos, en la exactitud de la prueba, o en la distribución de los puntajes de prueba, requiere de evidencia no confusa y replicable que apoye el rechazo de la hipótesis nula de que no tienen efecto. Sin embargo, la evidencia que sustenta la hipótesis del super-dampening y de las preguntas de asunto-externo es prácticamente inexistente hasta el momento. En la actualidad, las bases de la evidencia publicada ascienden a un pequeño número de estudios que en conjunto no logran apoyar la validez de la hipótesis de super-dampening o del uso de preguntas de asunto-externo para lograr el efecto hipotético.

<sup>7</sup> De alguna manera, esto es consistente con los hallazgos de Honts, Amato y Gordon (2001) de que las preguntas de asunto-externo podrían funcionar como una forma de preguntas comparativas.

Observamos también que ha habido un cambio o movimiento en los efectos hipotéticos, y que ninguno de los beneficios sugeridos están avalados por evidencia empírica.

En la actualidad no existe un método publicado para el análisis cuantitativo de las preguntas de asunto-externo y no hay datos normativos publicados con los que se intente cuantificar el grado de incertidumbre con respecto a cualquier interpretación del significado de las respuestas a estas preguntas. De igual manera, no hay una descripción publicada acerca de la eficacia de cualquier procedimiento clínico estructurado o de algún método cualitativo para la interpretación de las respuestas a estas preguntas. La interpretación de las respuestas a las preguntas de asunto-externo, sigue siendo un proceso clínico no cuantificado y desestructurado. Aunque alguna evidencia sugiere que la existencia de asuntos externos podría estar relacionada con más puntuaciones negativas en personas inocentes, no se sabe si esto está en función del incremento de la respuesta en las preguntas relevantes o del decremento de la respuesta en las preguntas de comparación. En cualquier caso, no encontramos bases para cualquier conclusión de que la generalización de nuestro conocimiento de la precisión de criterio y de las distribuciones normativas en los puntajes de personas culpables e inocentes debería de ser considerada en función de la inclusión o exclusión de las preguntas de asunto-externo por sí mismas.

Después de publicaciones previas relacionadas con las preguntas de asunto-externo, se generó cierta controversia y discusión, principalmente en respuesta a los desacuerdos expresados por los creadores y defensores de la hipótesis del super-dampening. Estos desacuerdos no ofrecieron pruebas adicionales a la discusión y se basaron en conjeturas y en opiniones personales. Si el pasado es indicativo de un futuro potencial, se pueden esperar ciertas objeciones para las actuales conclusiones. Si tales objeciones se acompañan de evidencias no confusas que les den soporte, entonces deberían de ser tomadas en serio y esta materia debería ser reconsiderada. Las objeciones y los argumentos ofrecidos sin evidencia son una forma de hipótesis que requiere de estudio. No hay ningún beneficio para la profesión al ignorar la evidencia disponible a favor de hipótesis sin confirmación o no probada, ya que esta forma de confirmación basada en prejuicios sólo sirve para satisfacer un objetivo secundario, al tiempo que interfiere con la integridad científica.

Ejemplos de objeciones previas ante la evidencia científica acerca de la hipótesis del super-dampening y de las preguntas de asunto-externo se pueden encontrar en el registro publicado. Backster (2001) objetó el uso de los datos en bruto "agrupados", y sugirió que cada gráfica de prueba es "única" en cuanto a su capacidad para predecir el éxito del examinador. En esto, el uso "único" requiere de una definición – ya que parece distinto al asunto científico y estadístico de la independencia, que se refiere al grado de covarianza entre los elementos (gráficos) que se espera que tengan covarianza -. Las gráficas de la prueba poligráfica dentro de un solo examen no son independientes debido a que incluyen los mismos estímulos, dirigidos a los mismos objetivos de estímulo para el mismo examinado. Si las preguntas de estímulo en un gráfico se ven afectadas por un comportamiento pasado, entonces es soporte para pensar que se podría esperar que este mismo comportamiento pasado vaya a influir en las reacciones de las gráficas de prueba subsecuentes.

Backster (2001) sugirió utilizar solamente aquellos casos en los que había una respuesta a las preguntas de asunto-externo y donde no había respuesta a las preguntas relevantes ni a las preguntas de comparación. Hacer esto, sería una violación grave del protocolo de investigación, ya que equivaldría a "regatear preguntas" o "seleccionar-cerezas" en los datos para conformar una conclusión previa. Sería el equivalente científico de escribir una historia para adaptarla a un final deseado. Además, Backster sugirió que la ausencia de evidencia no es base para excluir las preguntas de asunto externo de la estructura y administración de la prueba, y ofreció la analogía de que la eliminación de las preguntas de asunto externo del formato de preguntas, debido a la ausencia de evidencia sería el equivalente a retirar detectores de humo debido a la ausencia de fuego. Ante esto, tenemos que señalar que la analogía falla porque los detectores de humo han demostrado ser capaces de detectar incendios, mientras que las preguntas de asunto externo no han demostrado que funcionan como se hipotetizó. Fundamentalmente, es importante adherirse a los principios básicos de las pruebas científicas éticas y de las prácticas basadas en evidencia: normalmente debería exigirse la evidencia antes de incluir una hipótesis experimental en las prácticas de campo que afectan tanto a las personas como al público. Notamos que las discusiones acerca de la práctica basada en evidencia no comenzaron a aclararse dentro de la profesión poligráfica en la época de los trabajos de Backster acerca de la hipótesis del super dampening y de las preguntas de asunto-externo.

Backster (2001) comenzó a re-enmarcar el propósito de las preguntas de asunto externo cuando escribió lo siguiente: "Nuevamente, estas preguntas están diseñadas para indicar el éxito del examinador en la obtención de, al menos, una escasa confianza del examinado de que no se harán preguntas no revisadas" (Pág. 213). Los complejos supuestos, como la habilidad de evaluar la confianza del examinado en el examinador, normalmente requieren de evidencia y réplicas antes de ser incorporados en procedimientos que afectan a los individuos y a la comunidad. La sugerencia de que las preguntas de asunto externo proporcionan evidencia al nivel de la confianza que el examinado tiene en el examinador, o que el grado de confianza se presenta en los resultados inconclusos, es en esencia, una hipótesis altamente especulativa y en gran medida no comprobable, y por lo tanto no científica, dentro del medio de prueba poligráfica.

No se ha ofrecido evidencia que apoye la idea de que las preguntas de asunto externo pueden de hecho evaluar la "confianza del examinado." Del mismo modo, no existe descripción alguna acerca de la base psicológica para suponer que la calidad o la "confianza", como un atributo del rapport entre el examinado y el examinador, se manifiesta en las mismas respuestas autónomas que se utilizan para calcular el nivel normativo de la significancia estadística de la carga diferencial de la atención y de las respuestas ante los estímulos de prueba comparativos y relevantes. Honts, Amato y Gordon (2001) escribieron que al preguntar "¿Hay algo más que te preocupe que yo te pregunte acerca de?" es el equivalente funcional de "Antes de 1998 ¿Alguna vez hiciste algo que fuera deshonesto o ilegal?" Una hipótesis más plausible y comprobable que la hipótesis de la confianza podría sugerir que las respuestas fisiológicas observadas funcionan como una combinación de la carga cognitiva y emocional relacionada con la meta de pasar la prueba poligráfica y con los comportamientos pasados referidos por las preguntas de estímulo de prueba. Esta forma de hipótesis puede ser estudiada con mayor facilidad en el contexto de las respuestas diferenciales ante diferentes tipos de estímulos de prueba, y puede explicar más fácilmente los fenómenos y observaciones ya conocidos de que hacer preguntas de asuntos-externos puede estimular respuestas por asuntos externos. Se necesitan investigaciones adicionales en esta área.

En 1962, Backster aún no había introducido la hipótesis de la confianza, y describió la hipótesis del super-dampening como una amortiguación de todas las respuestas que ocurren normalmente:

El asunto externo en dicha persona, acerca de lo que él se encuentra muy aprensivo, es mucho más importante para él - o afecta más directamente a su bienestar - de lo que lo hace el motivo de la prueba poligráfica, provocando entonces un "super-dampening" en todas las respuestas que habrían ocurrido normalmente, incluyendo la amortiguación en todas las respuestas de las preguntas de estimulación revisadas cuestionadas al sospechoso inocente (Pág. 65).

En esta declaración se encuentran incrustados una serie de supuestos esotéricos y corolarios de la hipótesis del super-dampening, algunos de los cuales se han encontrado subsecuentemente con evidencia controversial. Lo central en esto es la noción de aprensión o miedo acerca del propio bienestar como la base de las respuestas ante los estímulos relevantes y de comparación - una hipótesis que en repetidas ocasiones se ha demostrado como falsa mediante la elevada precisión de criterio de las técnicas poligráficas que hacen uso de preguntas comparativas no-manipuladas de mentira-directa.<sup>8</sup>

Se expresaron otras objeciones por parte de Mate (2001a; 2001b), quien sugirió que Krapohl y Ryan (2001) habían dejado de reconocer las limitaciones del tamaño de la muestra de Capps, Knill y Evans (N = 150). Es altamente improbable que se pasara por alto el tamaño de la muestra, ya que Krapohl y Ryan comentaron sobre el tamaño de la muestra en la publicación del informe. Adicionalmente, habría sido imposible pasar por alto el tamaño de la muestra al tiempo de calcular la estadística chi cuadrada que Mate criticó, a pesar de que Capps, Knill y Evans (1993) utilizaron la misma estadística. Quizá, hoy en día sería seleccionado un método estadístico diferente - posiblemente un método estadístico robusto como el bootstrapping, el cual no depende de supuestos acerca de la forma de la distribución de los datos, o una estadística de Poisson, que es útil para eventos poco frecuentes - pero la estadística chi-cuadrada, probablemente era el método preferido al tiempo del estudio de Capps, Knill y Evans, y tiene cierto sentido el utilizar el mismo método para calcular la estadística faltante.

Mate (2001b) argumentó que los resultados confusos de Capps, Knill y Evans (1993) son adecuados para rechazar la hipótesis nula, y además re-enmarcó el propósito de las preguntas de asunto externo en torno a los efectos hipotéticos de proporcionar seguridad y discriminación con respecto a la confianza cuando escribió lo siguiente:

Para comenzar, se debe entender que la revisión y la incorporación subsecuente de las preguntas sintomáticas en la Técnica de Zona de Comparación está diseñada principalmente para asegurarle al examinado que no se harán preguntas no revisadas durante la administración de la prueba poligráfica durante la recolección de datos fisiológicos, y en segundo lugar para identificar al examinado que no está convencido de que no se le harán preguntas sin revisar durante la prueba. (p 220)

<sup>8</sup> Fue publicado por la American Polygraph Association, 2011, un sumario de estudios en técnicas poligráficas utilizando preguntas comparativas de mentira directa o probable.

Además del hecho de que esta hipótesis del aseguramiento se le agregó a las preguntas de asunto- externo solamente después de la aparición de evidencia controversial acerca la eficacia de estas preguntas, la base psicológica para el efecto de la hipótesis del aseguramiento no se puede determinar. Mate también sugirió el uso de preguntas de asunto-externo como la última pregunta en la serie, y ofreció el argumento de que los examinados podrían tener un "alivio"<sup>9</sup> en la última pregunta anticipando el final de la prueba. Esto representa una diferencia sustancial con respecto al propósito y fundamento sugerido previamente de las preguntas de asunto-externo. En cualquier caso, los hallazgos de Capps, Knill y Evans se mantienen como confusos e insuficientes para apoyar una conclusión con respecto a la validez de las preguntas de asunto-externo. Un enfoque científico requiere que no rechacemos de forma simple información incómoda, y que no nos centremos de forma simple en evidencia halagadora.

En los últimos 13 años desde la publicación de Honts, Amato y Gordon (2001) y Krapohl y Ryan (2001), no se ha hecho un nuevo intento por parte de los defensores de las preguntas de asunto externo para resolver los resultados confusos de Capps, Knill y Evans (1993). Desde entonces no se ha agregado evidencia a nuestra base de conocimientos científicos con respecto a las hipótesis del súper-dampening o de las preguntas de asunto-externo.

En respuesta a cualquier sugerencia potencial de que se espera que las preguntas de asunto-externo no evalúan la influencia de un asunto externo, pero que corrigen dicha influencia, notamos que no hay razón alguna en la literatura publicada en psicología, fisiología, psicofisiología, evaluaciones científicas, o en la teoría de la decisión estadística que sugiera que se podría esperar que la presentación de un estímulo de prueba pueda corregir un problema que no se puede cuantificar de manera confiable. Un enfoque más racional sería reforzar el principio general de que los examinadores deben esforzarse para conducir una entrevista y obtener datos de manera tal que el examinado no se distraiga. Honts, Amato y Gordon (2001) abordaron esto en su observación de que los asuntos externos podrían estar relacionados con los errores falso positivos, y que esto podría ser exacerbado por el uso de formatos de examinación ineficientes o enfoques de entrevista acusatorios que podrían incrementar el nivel de distracción debido a asuntos externos. Ellos sugirieron investigaciones futuras en esta área. Krapohl y Ryan (2001) señalaron que las preguntas de asunto externo pueden generar distracción en la prueba cuando ellos hablaron de preocupaciones reportadas de que las preguntas de asunto-externo generan sospechas en algunas personas. También señalaron que la inclusión de preguntas de asunto-externo cambia un examen explícito de asunto único en un examen explícito de asunto múltiple.

En consideración del hecho de que los principios subyacentes validados para el uso propuesto del formato AFMGQT son idénticos a los de otros formatos de evento específico, y de que la evidencia sugiere que las preguntas de asunto-externo y de los supuestos relacionados con ellas son hipótesis falsas, la única diferencia potencial remanente entre el uso propuesto del AFMGQT con otros formatos de prueba de evento específico, involucra la hipótesis de que la sola inclusión de preguntas

<sup>9</sup> El constructo de "descanso" en este uso no se encuentra en la literatura psicofisiológica, fuera de las discusiones acerca de las técnicas poligráficas de Backster. No está claro si es que este término se refiere a una forma de reacción debido al alivio a un retorno al nivel tónico.

de asunto-externo podría provocar ciertos cambios fundamentales indescriptibles en función de la efectividad de los estímulos objetivo de prueba.

En la actualidad no hay evidencia alguna, ni lógica plausible que apoye esta especulación, que debería, al igual que otras hipótesis, estar razonablemente apoyada por alguna descripción del mecanismo psicológico o fisiológico por el que se podría esperar que ocurriera algún cambio, y que eventualmente estuviera apoyado por evidencia antes de ser aceptado. Con excepción de los resultados atípicos que deben tomarse con precaución (American Polygraph Association, 2011), la evidencia disponible no apoya la noción de que los formatos de preguntas de prueba que incluyen preguntas de asunto-externo superan a los formatos sin preguntas de asunto-externo.

## Sumario

La posición de la Comisión de Investigación de la APA es que el criterio de validez y la generalización de los conocimientos pertenecientes a los formatos de prueba de todo tipo, no depende del nombre del formato de preguntas, ni tampoco de la incorporación de hipótesis falsas dentro de la estructura de prueba. El criterio de validez y la generalización depende, en parte, de la validez de los constructos subyacentes que determinan la producción, registro y cuantificación matemática o estadística de los datos registrados. El constructo central subyacente para los formatos de prueba con preguntas comparativas de cualquier tipo, incluyendo el AFMGQT, el Federal You-Fase y los formatos Federal Zone Comparison Test, es que la fuerza de respuesta a los diferentes tipos de estímulos de prueba variará con niveles estadísticamente significativos en función del estado de culpabilidad o inocencia del examinado con respecto a un asunto conductual que es el objetivo, y que la saliencia diferencial y la magnitud de la respuesta pueda ser cuantificada para predecir el estado de criterio del examinado dentro de un rango conocido de precisión que es significativamente mayor que el azar.

Se ha demostrado que los exámenes de evento específico con dos y tres preguntas relevantes proporcionan una precisión de criterio con tasas superiores al azar (American Polygraph Association, 2011), pero en este momento, la evidencia sugiere que todos estos formatos funcionan a pesar de, y no gracias a, la hipótesis falsa y no aprobada del super-dampening y de las preguntas de asunto-externo. Aunque estas preguntas intentan evaluar la posible función de un asunto externo como la causa de los resultados inconclusos de los exámenes calificados numéricamente, la interpretación de las respuestas a las preguntas de asunto-externo parecen ser una preocupación fenomenológica para la que no hay reglas publicadas ni evidencia que den soporte para la eficacia del uso clínico de estas preguntas. En resumen, el uso de preguntas de asunto-externo sigue siendo una sugerencia hipotética para la que no existe evidencia actual y que tampoco es probable que emerja. Si uno fuera a sostener que la pura inclusión de las preguntas de asunto-externo incrementan de alguna manera la validez, aun cuando la validez de estas preguntas fuera cuestionable,<sup>10</sup> entonces los requerimientos éticos para la práctica basada en la evidencia dictaría que esto también debería exigir evidencia replicable antes de proceder a requerir la inclusión de estas preguntas dentro del formato de preguntas.

<sup>10</sup> Notamos que la lógica no es clara acerca del porqué se esperaría que la inclusión de una hipótesis-falsa incrementará la precisión de criterio.

El cálculo del grado de incertidumbre o del nivel de significancia estadística de las calificaciones sumadas en los exámenes, está en función de dos factores principales: 1) si es que las preguntas objetivo de prueba se interpretan bajo el supuesto de un criterio de varianza independiente (es decir, el que el comportamiento pasado que afecta el criterio de culpabilidad vs. inocencia en cualquiera de las preguntas, se cree que no tienen ningún efecto sobre del criterio de culpabilidad vs. inocencia de otras preguntas objetivo), o bajo el supuesto de criterio de varianza no independiente (es decir, que se espera que el comportamiento pasado que afecta el criterio de culpabilidad vs. de inocencia de cualquiera de las preguntas podrá afectar potencialmente el criterio de culpabilidad vs. inocencia de otras preguntas objetivo), y 2) el número de preguntas objetivo de estímulo, quienes junto con el número de repeticiones de la secuencia de preguntas de prueba, afectarán el número total de observaciones estímulo-respuesta que componen el conjunto de datos de prueba, que pueden afectar la aplicabilidad de los datos normativos disponibles para las calificaciones sumadas de prueba,<sup>11</sup>

La conclusión de la comisión de investigación es que el uso propuesto del AFMGQT se ajusta suficientemente a los mismos principios validados de los formatos de preguntas de prueba en los formatos Federal You-Phase y Federal Zone Comparison Test, y que la generalización de los datos normativos disponibles parece razonable. Del mismo modo, considerando la falta de evidencia, descripción o indicadores de que las preguntas de asunto-externo juegan un papel en la formulación de los resultados de las pruebas, la generalización del conocimiento actualmente disponible acerca de la precisión de la prueba, también parece ser una sugerencia razonable. Aunque la presencia de asuntos externos podría estar relacionada con resultados de las pruebas, no existe evidencia que apoye la hipótesis de que las respuestas a las preguntas de asunto externo tienen por sí mismas efecto alguno en los resultados numéricos, la exactitud de la prueba, o la generalización de los datos normativos disponibles actualmente utilizados para calcular el nivel de significación estadística de un resultado de prueba.

Notamos que el uso propuesto del AFMGQT no añade una hipótesis nueva para el modelo de evaluación, sino que elimina las preguntas relacionadas con una hipótesis que parece ser falsa. Aunque no hay una investigación experimental directa con respecto al efecto de la inclusión o exclusión de las preguntas de asunto-externo en la secuencia de preguntas de prueba, una comparación de evidencia publicada acerca de las tasas de precisión de criterio de diferentes formatos de preguntas de prueba, mostraron que la inclusión o exclusión de preguntas de asunto-externo dieron como resultado distribuciones de puntuación numérica para los que los rangos normales se superponen sustancialmente (American Polygraph Association, 2011). De hecho, se ha demostrado que los formatos de preguntas de prueba que no utilizan o interpretan las preguntas de asunto-externo, producen niveles de precisión de criterio iguales o superiores a la precisión de criterio de los formatos que incluyen el uso tradicional de las preguntas asunto-externo.

Son dignas de mención varias observaciones precautorias, aunque parecerán obvias una vez que se articulen. En primer lugar, no sirve a la profesión el imponer el uso de una hipótesis falsa o restringir el uso de formatos de prueba que no incorporan una hipótesis falsa.

<sup>11</sup> Los modelos de transformación numérica y sumaria que no incluyen la sumatoria de datos de prueba (es decir, que confían en promediar, estandarizar o realizar cualquier otra función estadística de sus calificaciones) podrían verse menos afectados por el número de preguntas de prueba o número de repeticiones.

Además, no sirve a la profesión el hacer que los examinadores sean vulnerables a la crítica en torno a una hipótesis falsa y a requisitos de procedimiento que no generan diferencia alguna. Tampoco sirve la profesión o a la comunidad el perpetuar reglas que no son aditivos de la eficacia del modelo de prueba poligráfica. Tampoco sirve a la profesión ni a la comunidad el crear o imponer estándares que sirvan a personalidades individuales, al tiempo que se reniega de la obligación de formular estándares de práctica basados en evidencia con significado. No incrementa nuestro conocimiento o inteligencia el dirigir nuestra atención y recursos hacia cuestiones que no contribuyen realmente con la efectividad de la prueba. Tampoco mejora la posición o la estatura de la profesión poligráfica en el ámbito más amplio de la ciencia, de las pruebas y del análisis forense el permanecer anclado a conceptos y terminología que no se soporta con evidencia, o son inconsistentes con la ciencia y que no contribuyen para incrementar la precisión en la discriminación entre el engaño y la veracidad. El rigor científico dicta que descartemos ideas y prácticas que demuestran ser hipótesis falsas, y fomentemos prácticas que se ajusten a la evidencia científica.

### Recomendaciones

En consideración de la disponibilidad de los datos normativos generalizables y del conocimiento generalizable de las características de precisión de la prueba, el uso propuesto del AFMGQT debe ser aprobado sin la necesidad de protocolos de ética adicionales, relativos a la notificación a los examinados ni a los profesionales de referencia del uso de métodos experimentales para los que no se encuentran disponibles datos normativos y de evidencia publicada acerca de su validez y efectividad. Nuestra conclusión es que no sirve a la profesión o a la comunidad el tratar de mantener una posición dogmática que es inconsistente con la evidencia científica, del rol o la necesidad de preguntas de asunto-externo. De cualquier manera, notamos que existe una confusión potencial en torno a la interpretación de los resultados de prueba obtenidos mediante el uso del formato AFMGQT. Los datos de prueba obtenidos mediante el uso de este formato, tradicionalmente se interpretan bajo el supuesto de varianza de criterio independiente, ya que el uso propuesto implica el supuesto de preguntas de prueba para las que se asume una varianza de criterio de no independencia. Por esta razón, se sugiere que los examinadores apliquen los estándares de práctica en los que aclaren en su informe de prueba los supuestos de evaluación con respecto a la varianza de criterio independiente o no independiente. Sugerimos además, que el lenguaje del informe debe reportar a los profesionales de referencia las bases de datos normativos que describen los resultados probabilísticos que subyacen a las conclusiones categóricas con respecto al engaño o veracidad. Ofrecemos ayuda de los miembros del comité de investigación de la APA en la formulación del lenguaje del informe estándar y capacitación para lograr estos objetivos sugeridos para la notificación y presentación de informes.

## Referencias

- American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40(4), 196-305.
- Backster, C. (1962). Methods of strengthening our polygraph technique. *Police*, 6, 61-68.
- Backster, C. (2001). A response to Krapohl & Ryan's "belated look at symptomatic questions." *Polygraph*, 30, 213-215.
- Capps, M. H., Knill, B. L., & Evans, R.K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22, 285-298.
- Department of Defense (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retrieved from <http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf> on 3-31-2007. Reprinted in *Polygraph*, 40(1), 2-66.
- Honts, C., Amato, S., & Gordon, A. (2004). Effects of outside issues on the comparison question test. *Journal of General Psychology*, 131(1), 53-74.
- Honts, C. R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae *United States v Scheffer*. Available from the author.
- Kircher, J. C., Horowitz, S. W. & Raskin, D.C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*, 12, 79-90.
- Krapohl, D. J., & Ryan, A.H. (2001). A belated look at symptomatic questions. *Polygraph*, 30, 206-212.
- Krapohl, D. J. & Ryan, A.H. (2001). Final comment on the belated look at symptomatic questions. *Polygraph*, 30, 218-219.
- Matte, J. A. (2001). Reply to rejoinder by Donald J Krapohl and Andrew H Ryan. *Polygraph*, 30, 220-222.
- Matte, J. A. (2001). Comments on Krapohl & Ryan criticism of Capps, Knill & Evans research on symptomatic questions. *Polygraph*, 30, 216-217.
- Nelson, R. & Handler, M. (2013). A brief history of scientific reviews of polygraph accuracy research. *APA Magazine*, 46(6), 22-28.
- Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B. & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph*, 40, 67-78.
- Raskin, D. C., & Honts, C. R. (2002). The Comparison Question Test. In M. Kleiner (Ed.), *Handbook of Polygraph Testing*. Academic Press.