

*Fernando Toboso Marqués**

Grado de acierto de estimaciones y pronósticos: criterios de evaluación de la metodología y calidad de los análisis

[Recibir BOLETÍN
ELECTRÓNICO](#)

Grado de acierto de estimaciones y pronósticos: criterios de evaluación de la metodología y calidad de los análisis

Resumen:

En el campo de las ciencias sociales, algunos análisis intentan explícitamente ofrecer estimaciones sobre el futuro. Este artículo abordará desde dos metodologías, una de naturaleza predominantemente cualitativa y otra predominantemente cuantitativa, algunos de estos estudios de futuro, explicando en qué se basan y tratando de aclarar los criterios para su evaluación.

Para ello, se basará en la obra de autores norteamericanos de referencia mundial en su campo. Al ser un estudio de alcance limitado, quedarán campos sin estudiar, desde autores españoles a métodos prospectivos, pasando por el análisis y pronóstico de resultados electorales. Por último, se revisarán las normas sobre análisis de la Comunidad de Inteligencia norteamericana y española para averiguar en qué basan su calidad y su posible evaluación.

Abstract:

In the field of Social Sciences, some analysis explicitly attempt to provide assessment about the future. This paper will address two methodologies from a predominantly qualitative and a predominantly quantitative nature, some of these studies of future, explaining what they are based on and trying to clarify the criteria for evaluation.

***NOTA:** Las ideas contenidas en los **Documentos Marco** son de responsabilidad de sus autores, sin que reflejen, necesariamente, el pensamiento del IEEE o del Ministerio de Defensa.

To do this it will be based on the work of American authors world reference in its field. As a study of limited scope, remain unstudied fields from Spanish authors to prospective methods and election results. Finally, the analytic standards of U. S. and Spain Intelligence Community will also be reviewed to find out which base their quality and their possible evaluation.

Palabras clave:

Pronóstico, estimación, calidad del análisis, exactitud, rigor, utilidad, decisores, inteligencia, elecciones, asuntos exteriores, juicio experto, técnicas analíticas estructuradas, algoritmo, teorema de Bayes.

Keywords:

Forecast, estimate, assessment, quality of analysis, accuracy, rigor, utility, decision makers, Intelligence, elections, Foreign Affairs, Expert judgment, Structured Analytic Techniques, algorithm, Bayes` theorem.

Introducción

En el campo de las ciencias sociales, muchos análisis describen las características de una población determinada o realizan un estudio global de un fenómeno desde el pasado hasta el momento presente, resaltando los factores más relevantes, sus regularidades y patrones, asumiendo que dichos factores condicionarán su evolución futura. Otros análisis son más explícitos respecto al futuro, atreviéndose a concretarlo mediante estimaciones o pronósticos. Este campo de las ciencias sociales abarcaría desde estudios de inteligencia a estudios relacionados con asuntos exteriores, pasando por estudios estratégicos, política o seguridad.

El término «análisis» se refiere a su plasmación en un informe o artículo escrito. Los más importantes, como las NIE, (Estimaciones Nacionales de Inteligencia elaboradas por la Comunidad de Inteligencia norteamericana), son el resultado de numerosos y muy diversos análisis realizados con distintas metodologías, lo que complica evaluar su calidad.

«Parece razonable suponer que se podría conseguir un veredicto estadístico verdaderamente objetivo, sobre la exactitud de las estimaciones de Inteligencia. Revisar los documentos, distinguir los juicios correctos de los incorrectos, y calcular el “promedio de bateo”. Pensé que podría hacerse, lo intenté pero resultó imposible»¹.

Desde 1969 analistas norteamericanos, como Smith, se preguntan acerca de la precisión de las estimaciones de inteligencia y ponen esta metáfora sobre el promedio de bateo para concluir acerca de la gran dificultad que conlleva. Aunque con Nate Silver las aplicaciones del promedio de bateo alcanzaron las más altas cotas (ver apartados posteriores), esta metáfora sigue utilizándose ya que, respecto a la exactitud de las estimaciones de inteligencia, no hay tanto consenso².

Este artículo abordará desde dos metodologías, una predominantemente de naturaleza cualitativa y otra predominantemente cuantitativa, algunos de estos estudios de futuro, explicando en qué se basan y tratando de aclarar los criterios para su evaluación. Para

¹ SMITH Abbot E. «*On the accuracy of National Intelligence Estimates*». *Studies in Intelligence* Vol. 13, No. 4 (Fall 1969) pp. 25-35. Disponible en : https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol13no4/html/v13i4a04p_0001.htm Fecha de consulta: 25.05.2016.

² MARRIN Stephen. «*Evaluating the quality of Intelligence Analysis: by What (Mis) measure?*». *Intelligence and National Security*. Volume 27, Issue 6, 2012. Disponible en: <https://es.scribd.com/doc/91721279/Marrin-EvaluatingtheQualityofIntelligenceAnalysis>. Fecha de consulta: 25.05.2016.

ello, se basará en la obra de autores norteamericanos de referencia mundial en su campo. Este trabajo es de gran magnitud y novedoso en nuestra lengua, por lo que no se pretende que sea más que una primera aproximación al tema, quedando aspectos importantes pendientes de otros análisis como se expone en el apartado de «Limitaciones y retos». Tampoco se pretende que los casos expuestos se consideren lo suficientemente representativos como para generalizar unas conclusiones indiscutibles, aunque se detallen varias al final.

Además de abordar la metodología empleada para la elaboración de un análisis, se abordará la calidad del mismo, que sin duda se verá muy influenciada por la metodología usada, pero con criterios propios a la hora de evaluarla. A este respecto, se repasará la visión de la Comunidad de Inteligencia norteamericana y española sobre las normas de calidad de los análisis que elaboran.

Metodología, qué valorar

Si una estimación o pronóstico acierta (se corresponde lo escrito con lo sucedido), se concluirá que la aplicación de la metodología ha sido un éxito, será un factor claro para evaluarla, pero además del acierto, ¿hay otros criterios relevantes? ¿Cuáles serían?

Como mediante el acierto, ya se ha valorado el resultado de la aplicación de la metodología, parece lógico valorar también si el proceso por el que se ha llegado a esas conclusiones se ha llevado a cabo con rigor, e incluso si se van a poder generalizar los resultados obtenidos con esta metodología a otros casos similares o si se podrán repetir con resultados análogos. En definitiva se tendría en cuenta:

- Exactitud o acierto (validez), correspondencia entre la estimación y lo sucedido. Así se evalúa el grado de confianza que merece la información obtenida mediante el proceso de investigación sobre el fenómeno estudiado. No siempre es fácil de valorar ni aun transcurrido el tiempo. Con la exactitud se valora la magnitud/importancia del resultado.
- Rigor en la aplicación del método elegido (validez interna) en el proceso de elaboración. La aplicación de esta metodología exige el seguimiento de unas normas precisas prefijadas que ayudarán a que los resultados obtenidos estén libres de sesgos. El análisis ha sido diseñado y realizado asegurando su coherencia. No se habla de verdadero o falso.

- Posibilidad de generalización de los resultados obtenidos con este método a otro momento, a otra población y un contexto similar (validez externa). Un factor más relativo y quizás menos relevante en el campo estudiado. Las particularidades de cada tema hacen que no existan dos problemas iguales, aunque no se renuncie a una aplicación crítica y no mecánica. En muchas ocasiones con que ese método «funcione» en nuestro estudio no será poco.
- Posibilidad de replicar la investigación por otros analistas obteniendo resultados similares (fiabilidad/consistencia), esto sería la mejor prueba del rigor en el proceso de elaboración, para ello, el análisis debería incluir toda la información necesaria para facilitar dicha repetición. Conceptos como revisión o confirmabilidad adquirirían pleno significado. En la práctica no es tan fácil de realizar. Así en el campo farmacéutico, ámbito casi ideal para la experimentación, los laboratorios Bayer fueron incapaces de replicar casi 2/3 de los hallazgos divulgados en publicaciones especializadas³.

Este artículo se centrará en los dos primeros criterios, y algo en el tercero, al ser más aplicables en estos estudios.

Métodos cualitativos, estimaciones

Se usa la palabra estimación o valoración (conceptos ingleses *estimate* y *assessment*) por parecer más apropiada en estudios de inteligencia, estratégicos, asuntos exteriores, política o seguridad.

En estos campos la mayoría de las veces los datos son cualitativos, con gran incertidumbre, ambigüedad e incluso, en temas de Inteligencia, datos falsos, producto del engaño.

También se encuentran temas con abundantes datos cuantitativos, como las elecciones políticas y las estadísticas relacionadas, susceptibles de analizarse con distintos métodos, entre ellos una «Técnica Analítica Estructurada» de carácter cualitativo: el «Análisis de Hipótesis en Competencia», como se verá.

³ OWENS, Brian «Reliability of 'new drug target' claims called into question». Newsblog Nature. 05.09.2011. Disponible en: http://blogs.nature.com/news/2011/09/reliability_of_new_drug_target.html
Fecha de consulta: 25.05.2016.

Método 1. Juicio experto

Es la manera tradicional en la que se llevan a cabo la mayoría de los análisis estratégicos y políticos. Está ampliamente implantado en todos los ámbitos y países y es de naturaleza cualitativa. Un experto sería aquel que se gana la vida con sus conocimientos en un campo determinado.

El prestigio de un reconocido experto, ganado por sus estudios exitosos en el pasado, avala su análisis en el presente. La experiencia y formación del experto le debe llevar a extraer las conclusiones correctas, ya que el amplio conocimiento de casos históricos similares le debe ayudar a interpretar adecuadamente las evidencias del caso de estudio. No solo es un razonamiento basado en su formación, experiencia e intuición, muchos autores realizan una aproximación crítica⁴ a la realidad, aunque no sigan un procedimiento sistemático que podamos calificar de método o técnica de análisis. Generalmente es un trabajo individual que se desarrolla en la mente del experto y que al final se plasma en un escrito.

Este «no método» es el más antiguo y extendido, y precisamente por ello ha sido más evaluado, exponiéndose más adelante sus limitaciones.

Método 2. Técnicas Analíticas Estructuradas (TAE)

Ante los fallos en el análisis de inteligencia (ataques terroristas del 11S de 2001, y posesión de ADM por Irak en 2002) y conocidas las limitaciones del juicio experto, la Comunidad de Inteligencia (CI, en adelante) de EE.UU., trató de mejorar la forma en que se llevaba a cabo dicho análisis.

En 2004, la IRTPA⁵ encomendaba al director de Inteligencia Nacional la responsabilidad de asegurarse de que «elementos de la Comunidad de Inteligencia efectúen análisis alternativos (comúnmente conocidos como “análisis de equipo rojo”) sobre la información y conclusiones en el análisis de inteligencia».

⁴ PAUL Richard, ELDER Linda. «La mini-guía para el pensamiento crítico. Concepto y herramientas». Fundación para el Pensamiento crítico. 2003. Disponible en: <https://www.criticalthinking.org/resources/PDF/SP-ConceptsandTools.pdf>. Fecha de consulta: 25.05.2016.

⁵ Intelligence Reform and terrorism Prevent Act of 2004. SEC. 1017 (a). Disponible en: <http://www.nctc.gov/docs/irtpa.pdf>. Fecha de consulta: 25.05.2016.

Estos análisis alternativos/«análisis de equipo rojo» se realizaban con una docena de técnicas, que en 2005 empiezan ya a ser conocidos como «Técnicas Analíticas Estructuradas».

En 2009, el Gobierno norteamericano y la CIA⁶. detallan esa docena de técnicas agrupadas en tres funciones «Diagnóstico», «Contrarias» e «Imaginativas».

En 2010, Richards J. Heuer y Randolph H. Pherson, dos exagentes de Inteligencia que tras su retirada del servicio activo se especializaron en la formación de analistas, amplían hasta 50 esas técnicas, en la primera edición de su libro *Structured Analytic Techniques for Intelligence Analysis*⁷, técnicas que aumentan hasta 55 en 2014 en la segunda edición⁸, agrupadas en ocho funciones: «descomposición y visualización», «generación de ideas», «escenarios e indicadores», «generación y prueba de hipótesis», «evaluación de causa y efecto», «análisis de reto», «gestión de conflictos» y «apoyo a la decisión». Estas técnicas además de cumplir con esta «función utilitaria», estarían diseñadas para reducir los fallos en el análisis provocados por sesgos y trampas cognitivas.

Muchas de estas técnicas son «importadas» desde diversos campos (empresarial, financiero, psicología, etc.), donde ya habrían demostrado su utilidad.

Método 3. Otras metodologías y su «jerarquía»

Evidentemente se pueden realizar análisis en estos temas con otros métodos además de los anteriormente mencionados, por ejemplo, estudiar mediante una encuesta los indicios de radicalización yihadista en las prisiones españolas⁹ o el riesgo de

⁶ US Government. CIA. «A tradecraft primer: Structured Analytic Techniques for improving Intelligence Analysis». 2009. <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf>. Fecha de consulta: 25.05.2016.

⁷ HEUER Richards J. Jr y PHERSON Randolph H. «Structured Analytic Techniques for Intelligence Analysis». 2010. Reseña disponible en: <http://www.amazon.com/Structured-Analytic-Techniques-Intelligence-Analysis-ebook/dp/B00B050Y46>. Fecha de consulta: 25.05.2016.

⁸ HEUER Richards J. Jr y PHERSON Randolph H. «Structured Analytic Techniques for Intelligence Analysis». 2014. Reseña disponible en: <http://www.cqpress.com/product/Structured-Analytic-Techniques-2.html> Fecha de consulta 20.01.2016. Fecha de consulta: 25.05.2016.

⁹ TRUJILLO, Humberto M; JORDÁN, Javier; GUTIÉRREZ, José Antonio y GONZÁLEZ-CABRERA, Joaquín. «Indicios sobre la radicalización yihadista en prisiones». 2008. Disponible en: http://www.academia.edu/3617448/Indicios_sobre_la_radicalizaci%C3%B3n_yihadista_en_prisiones Fecha de consulta: 25.05.2016.

radicalización islamista en cinco mezquitas españolas, mediante un estudio de observación participante¹⁰.

A este respecto, es importante destacar que no todas las metodologías de análisis tienen la misma potencia probatoria/credibilidad/estatus, así de mayor a menor, se podrían ordenar en metodología/diseño:

1. Experimental. Se caracteriza por la manipulación activa por parte del investigador de la variable independiente, causa o de tratamiento, el control de las variables extrañas y la asignación aleatoria de los sujetos a los grupos de investigación. Facilita las condiciones óptimas para probar de forma inequívoca hipótesis de causalidad.
2. Cuasi experimental. Asignación no aleatoria de los sujetos a los grupos de investigación. Permite una cierta inferencia de causalidad, aunque es relativamente alta la probabilidad de que existan hipótesis explicativas alternativas. Esta metodología y la anterior son más utilizadas en Ciencias Experimentales: exactas, naturales, de la salud, tecnológicas.
3. No experimental. Encuesta y observación/cualitativa. Ausencia de manipulación de las variables y aleatorización de los grupos. Difícilmente se puede inferir hipótesis causales, normalmente solo permite describir lo estudiado. Metodología más utilizada en ciencias sociales, junto con otras técnicas y herramientas de análisis.

Siendo muy clara esta figura sobre su «jerarquía metodológica»¹¹.

¹⁰ TRUJILLO, Humberto M; LEÓN, Cristóbal; SEVILLA, David y GONZÁLEZ-CABRERA, Joaquín. «Estudio del riesgo de radicalización islamista en cinco mezquitas de una ciudad española». 2009. Disponible en: <http://www.thefreelibrary.com/Riesgo+de+radicalizacion+islamista+en+las+mezquitas+de+una+ciudad...-a0314254310> Fecha de consulta: 25.05.2016.

¹¹ McINTOSH Cameron, LI Jobina «An introduction to economic analysis in crime prevention» Research report 2012-5. National Crime Prevention Centre. Government of Canada. Disponible en: <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/cnmc-nlss/index-en.aspx> Fecha de consulta: 25.05.2016. Traducción (de arriba a abajo): **Ensayos controlados aleatorizados**. Preferiblemente doble ciego. **Diseño cuasi-experimental**. Experimentos sin aleatorización. **Estudios de Observación controlada**. Comparación de resultados entre los participantes que han recibido una intervención y los que no. **Estudios de Observación**. Sin Grupo de Control. **Opinión experta**.

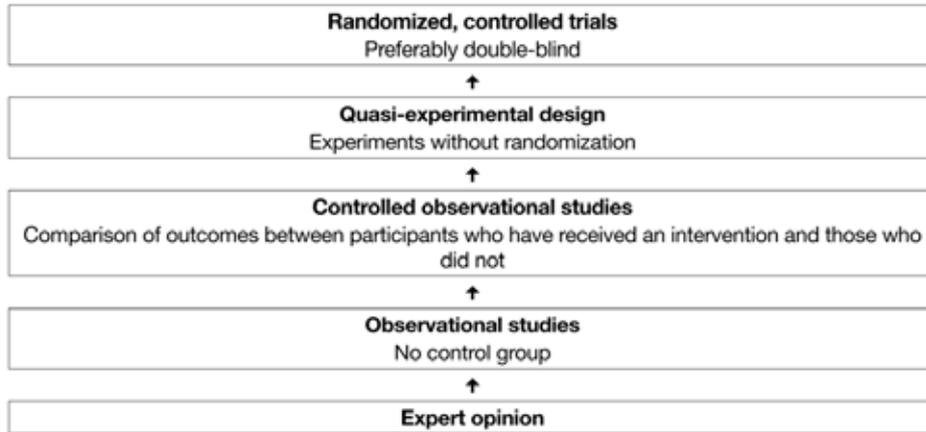


Figura 1. Farrington, 2003; Levin & McEwan, 2001

Un ejemplo en el campo de la medicina ayudará a aclararlo. Un estudio mediante un experimento (asignación aleatoria de los pacientes a los grupos de investigación), «doble ciego» (ni los pacientes ni los investigadores saben quién pertenece al Grupo Experimental o al Grupo de Control) será más creíble y ofrecerá más garantías de calidad que la observación de un Grupo o la opinión de un prestigioso doctor. Se puede afirmar que un estudio tendrá más rigor cuanto más alto se encuentre en esta «pirámide» el método empleado para realizarlo. Su diseño así lo avala.

Un análisis de temas médicos (ensayo clínico aleatorizado), para ser siquiera considerado por la comunidad científica, debe ser elaborado con un método perfectamente regulado y pautas conocidas, siendo el concepto de calidad de la metodología intercambiable con calidad del análisis, ya que lo que se plasma en el informe escrito es el desarrollo y resultado de la propia investigación¹². No podemos afirmar lo mismo en ciencias sociales (estudios de Inteligencia, estratégicos, etc.), donde el formato del informe escrito es mucho más flexible para adaptarse a la audiencia a la que va dirigido y la variedad de métodos para su elaboración es muchísimo mayor, separándose ambos conceptos de calidad (de la metodología y del análisis) y por tanto, deben

¹² CONSORT 2010, lista de comprobación de información sobre ensayo clínico aleatorizado. Disponible en: http://www.consort-statement.org/Media/Default/Downloads/Translations/Spanish_es/Spanish%20CONSORT%20Checklist.pdf Fecha de consulta: 25.05.2016.

establecerse los criterios para evaluarlas. Sobre la calidad del análisis, se dedicarán los últimos apartados de este artículo con los enfoques norteamericano y español.

Evaluación de método 1. Juicio experto

Sin duda la persona que ha realizado la evaluación más completa sobre el juicio experto ha sido Philip Tetlock¹³, este psicólogo norteamericano empezó en 1984 una investigación que se prolongó durante 20 años. En ella entrevistó a 284 expertos, de docenas de países: diplomáticos, abogados, periodistas, profesores, economistas, politólogos, siendo la mitad de ellos doctores universitarios.

Realizó más de 27.450 preguntas en las que los expertos debían comprometerse con estimaciones muy variadas sobre: liderazgo político a corto plazo (ganador de las próximas elecciones) y largo plazo (ganador de las dos próximas elecciones), violencia étnica o cambio de fronteras (separación de Quebec, por ejemplo); evolución de los principales indicadores económicos a 2-5 años; evolución de los gastos de defensa, alianzas militares o adquisición de armas nucleares en 5-10 años y ocho temas especiales desde adopción del euro por la UE, hasta emisiones de CO², pasando por la «nueva economía de Internet» (Microsoft, Amazon, e-Bay, etc.). En conjunto se revisaron los principales conflictos mundiales de ese periodo: caída del muro y descomposición de los países comunistas, apartheid de Sudáfrica, invasión de Kuwait, etc.

Las predicciones fueron, en promedio, solo ligeramente mejores que si hubieran sido realizadas al azar. El que los expertos tuvieran doctorados, mucha experiencia o acceso a información confidencial no constituía un factor diferenciador. Tetlock solo apreció relación entre la fama y el acierto, pero en sentido inverso: a más fama menos acierto, posiblemente ciertos expertos eran muy demandados en televisión por la rotundidad de sus respuestas y no pocas veces, por el espectáculo que proporcionaban.

Tetlock diferenciaba entre los expertos con fuerte convicciones, a menudo con una clara visión teórica con la que interpretaban todo y que nunca cambiaban de opinión,

¹³ TETLOCK Philip. «*Expert Political Judgment: How good is it?*». 2005. Princenton University Press. Disponible en: http://emilkirkegaard.dk/en/wp-content/uploads/Philip_E._Tetlock_Expert_Political_Judgment_HowBooks.org_.pdf. Fecha de consulta: 25.05.2016.

denominándolos «erizos» (de acuerdo al filósofo Isaías Berlín) y los que si cambiaban de opinión ante nuevos datos y circunstancias, más pragmáticos y menos tajantes, a los que denominó «zorros». Para Tetlock, estos matices y no la experiencia, permitían que los zorros fueran mejores pronosticadores tanto en temas que dominaban como en los que no. Pero solo los mejores «zorros» se acercaban al grado de acierto de pronósticos realizados con lo que Tetlock denominaba modelos estadísticos.

Para Tetlock una forma de mejorar el juicio experto individual sería el juicio colectivo de personas bien informadas en la materia, aunque no lleguen al dominio del experto. A este respecto y con el respaldo de la CI norteamericana (IARPA¹⁴) puso en marcha el proyecto «El buen Juicio»¹⁵, donde las respuestas grupales mejoraban un 30% la exactitud del pronóstico realizado por la propia CIA¹⁶.

Evaluación de método 2. Técnicas Analíticas Estructuradas (TAE)

No todas las TAE son iguales, las hay complejas, de carácter cuantitativo (como el «Análisis de redes»), o cualitativo (Delphi), que requieren dominar un software muy específico y la ayuda de especialistas en los mismos y otros expertos y otras mucho más sencillas, con menos requerimientos de formación, tiempo y colaboración (como el análisis DAFO o el diagrama de Venn). Igualmente unas están descritas con un procedimiento de aplicación más detallado que otras.

Los propios Heuer y Pherson han tratado de evaluar sus técnicas¹⁷ mediante la validez aparente de la técnica y mediante experimentos. Si una TAE se diseña específicamente para evitar un sesgo del razonamiento y parece que efectivamente lo

¹⁴ Intelligence Advanced Research Projects Activity is sponsoring the Good Judgment Project. 2015. Disponible en:

<http://www.iarpa.gov/index.php/newsroom/iarpa-in-the-news/2015/439-the-good-judgment-project?highlight=WyJ0aGUlLCJnb29klwianVkJ21lbnQilLCJwcm9qZWNOlwidGhllGdvd2QilLCJ0aGUqZ29vZCBqdWRnbWVudClslmdvb2QganVkJ21lbnQilLCJnb29kIGp1ZGdtZW50IHByb2plY3QilLCJqdWRnbWVudCBwcm9qZWNOlI0>. Fecha de consulta: 25.05.2016.

¹⁵ UNGAR Lyle «The Good Judgement project: a large scale test of different methods of combining expert predictions». AAI Technical Report FS-12-06. 2012. Disponible en: <https://www.aaai.org/ocs/index.php/FSS/FSS12/paper/viewFile/5570/5871>. Fecha de consulta: 25.05.2016.

¹⁶ Revista RFI. La «sabiduría de las multitudes» o cómo la gente predice los fenómenos mundiales mejor que la CIA. 04-04-2014. Disponible en: <http://es.rfi.fr/ciencia/20140404-la-sabiduria-de-las-multitudes-o-como-la-gente-predice-los-fenomenos-mundiales-mejo> Fecha de consulta: 25.05.2016.

¹⁷ HEUER Richards J. Jr. y PHERSON Randolph H. «Técnicas analíticas estructuradas para el análisis de inteligencia». 2015. Editorial Plaza y Valdés. Versión en español de la 1.ª edición, 2010, de este libro. Capítulo 13. Reseñas disponibles en: <http://www.plazayvaldes.es/libro/tecnicas-analiticas-estructuradas-para-el-analisis-de-inteligencia/1493/>. Fecha de consulta: 25.05.2016.

consigue, entonces se puede afirmar que tiene validez aparente, es una suposición lógica, es razonable esperar que el uso de una técnica ayude (como al estudiante le ayuda el uso de ciertas técnicas a afrontar un examen), pero no hay una prueba empírica de ello.

La prueba empírica, el experimento, consistiría en formar dos grupos similares de analistas, asignándolos aleatoriamente a un Grupo Experimental (o de Estudio) y a un Grupo de Control para realizar una estimación sobre un tema. El Grupo de Estudio realizaría la estimación con la técnica de análisis que tratamos de evaluar, mientras que el Grupo de Control no utilizaría ninguna técnica, luego con el tiempo se compararían los resultados obtenidos con la realidad.

Pero Heuer y Pherson concluyen que la experimentación en la práctica es más difícil de lo que parece, dadas las características particulares del análisis de inteligencia, es más, para ellos la evaluación de las TAE no se ha intentado de una forma sistemática. Por ello, proponen en su libro (apartado 13.3) *Una nueva aproximación a la evaluación*, pero el procedimiento es notablemente complicado.

También Robert Folker en 2000¹⁸ y Paul Lehner en 2004¹⁹ lo intentaron antes mediante experimentos con la TAE quizás más representativa: el Análisis de Hipótesis en Competencia (ACH). El experimento de Folker describe un escenario de guerra entre países imaginarios concluyendo que el Grupo Experimental obtiene mejores resultados (con ACH) que el Grupo de Control (sin ACH). Sin embargo, sobre el experimento de Lehner, Heuer y Pherson concluyen que sus resultados no se pueden generalizar puesto que el escenario de prueba era engañoso²⁰, a pesar de que se desarrollaba sobre un hecho real: la explosión en el acorazado USS Iowa en 1989.

Mejor suerte parece que corrieron, también con ACH, en pronósticos electorales Brasfield²¹ en 2009 (experimento sobre la elección del gobernador del estado de

¹⁸ FOLKER, Robert. D., Jr. 2000. «*Intelligence analysis in theater joint intelligence centers: An experiment in applying structured methods*». Occasional Paper No. 7. Washington, DC: Joint Military Intelligence College. Disponible en: <https://fas.org/irp/eprint/folker.pdf> Fecha de consulta: 25.05.2016.

¹⁹ LEHNER, Paul. E., ADELMAN Leonard, CHEIKES Brant A., BROWN Mark J. «*Confirmation bias in complex analyses*». MITRE corporation. Octubre 2004. Disponible en: https://www.mitre.org/sites/default/files/pdf/04_0985.pdf Fecha de consulta: 25.05.2016.

²⁰ HEUER Richards J. Jr. y PHERSON Randolph H. «*Técnicas analíticas estructuradas para el análisis de inteligencia*». 2015. Editorial Plaza y Valdés, p. 330.

²¹ BRASFIELD Drew. «*Forecasting accuracy and cognitive bias in the ACH*». 2009. Disponible en <https://es.scribd.com/doc/35793494/Forecasting-Accuracy-and-Cognitive-Bias-in-the-Analysis-of-Competing-Hypotheses>

Washington) y Wheaton²² en 2014 (experimento sobre la elección del presidente hondureño), confirmando que el Grupo (de Estudio) que usaba dicha TAE mejoraba el acierto en sus pronósticos con respecto al Grupo (de Control) que no la usaba, lo que constituía una evaluación de la exactitud de esa técnica.

Más recientemente, en 2015, el profesor norteamericano Stephen Coulthart²³, realiza una evaluación más completa de las TAE.

Coulthart elabora un estudio sobre la aplicación y eficacia de las TAE en el *Bureau of Intelligence and Research* (INR, Agencia de Inteligencia del Departamento de Estado), este INR es un miembro de la Comunidad de Inteligencia norteamericana con ciertas particularidades: es una agencia de las más antiguas, pero un 25% más joven que la media de la CI (el 73% de sus analistas se incorporaron tras el 11S), pequeña (200 analistas) si la comparamos con la CIA o la NSA (20.000 analistas en toda la CI), no inciden demasiado en la formación de su personal en las TAE (así 1/3 de sus analistas no las utiliza), sus analistas se especializan en un área geográfica de por vida y la presión del tiempo no parece ser agobiante para la elaboración de sus análisis.

El profesor Coulthart para evaluar las TAE se centra en el núcleo de 12, identificadas por el Gobierno norteamericano y la CIA²⁴, por lo que, realizó una búsqueda en Google Scholar de artículos en inglés elaborados con estas 12 TAE. «Solo» encontró resultados de seis de ellas: 838 artículos elaborados con la técnica de *Brainstorming*, 753 con «Análisis de Futuros Alternativos», 31 con «Abogado del Diablo», 21 con «Análisis de Equipo Rojo», 20 con «Análisis de Hipótesis en Competencia», y 17 con «Equipo A/Equipo B». La antigüedad de la técnica se supone que es un factor positivo en relación a su mayor difusión y aplicación.

Las otras seis técnicas son más modernas y quizás por eso no se han localizado artículos, o bien no son públicos, hecho que también puede ocurrir con otros artículos

<http://sourcesandmethods.blogspot.com.es/2010/08/does-analysis-of-competing-hypotheses.html> Fecha de consulta: 25.05.2016.

²² WHEATON Kristan. «Reduce bias in analysis: Why should we care?». 2014 Disponible en: <http://sourcesandmethods.blogspot.com.es/2014/03/reduce-bias-in-analysis-why-should-we.html> Fecha de consulta: 25.05.2016.

²³ COULTHART Stephen J. «Improving the analysis of Foreign Affairs: Evaluating Structured Analytic Techniques». 2015. University of Pittsburg. Disponible en:

http://d-scholarship.pitt.edu/26055/1/CoulthartSJ_ETD2015.pdf Fecha de consulta: 25.05.2016.

²⁴ «A tradecraft primer: Structured Analytic Techniques for improving Intelligence Analysis». *Ibid.*

de las seis mencionadas, en cualquier caso, es el estudio más completo y reciente realizado.

También este mismo profesor norteamericano, mediante la revisión de una selección de esos artículos localizados, llegó a ciertas conclusiones sobre lo que denomina eficacia²⁵ de estas seis TAE, siendo las más creíbles²⁶ las siguientes: «Abogado del Diablo» 72% de Eficacia, «Análisis de Hipótesis en Competencia» 50%, *Brainstorming* 40%. Es decir, de todas las TAE, solo ha podido fundamentar la eficacia de tres de esas técnicas.

En cuanto al rigor en la aplicación de cada TAE en el proceso de elaboración, los estudios de los distintos autores aludidos, apenas mencionan que efectivamente se ha seguido el procedimiento que las caracteriza.

Si en la CI norteamericana, que es donde más extendido está el uso de las TAE, se encuentran estas limitaciones en su evaluación, ¿Qué se podrá decir de la aplicación y eficacia de las TAE en otras Comunidades de Inteligencia o en otros campos? ¿o de su evaluación? A pesar de ello y a la vista de los fallos de inteligencia y los resultados obtenidos con otros métodos, esto no debe suponer un impedimento a su difusión y aplicación.

Métodos cuantitativos, pronósticos

Se usa la palabra pronóstico o predicción (concepto inglés *forecast*) asociada a temas meteorológicos, deportivos y políticos.

Se abordará el tema mediante ejemplos en el campo deportivo (béisbol) y político (elecciones norteamericanas) con un autor que ha ganado su prestigio por el *Acierto* en sus pronósticos en *Ambos* campos: Nate Silver²⁷, joven estadístico norteamericano que saltó a la fama tras su acierto en la elección presidencial norteamericana de 2008.

Silver aconseja pronosticar con actitud de «zorro», esto es: pensar de una forma probabilística y en vez de presentar un resultado único presentar varios resultados posibles. Modificar las predicciones a medida que se conocen nuevos datos, si somos

²⁵ Impacto positivo de la técnica en el rigor y exactitud del estudio. *Ibid*, pp. 118 a 125 y 214. Sin que precise cómo se combinan estos elementos.

²⁶ En función del diseño de investigación, según lo explicado en los artículos revisados. Ver asimismo lo relativo a «jerarquía metodológica», en este mismo artículo.

²⁷ SILVER Nate. «La señal y el ruido». Editorial Península. 2014.

buenos pronosticadores el aumento de información nos debe permitir mejorar nuestras predicciones y por último buscar el consenso para evitar errores de bulto, bien a través de predicciones colectivas bien combinando métodos cualitativos con cuantitativos.

La base del trabajo pronosticador de Silver es doble, por un lado una amplia y más o menos completa base de datos estadística y unas reglas del juego (deportivo y político) cerradas y estrictas.

En el caso del béisbol la base de datos es muy completa remontándose a los inicios del deporte (1929 y ya entonces empezaron las apuestas) con reglas muy precisas que regulan la confrontación entre equipos rivales, principalmente entre un lanzador y un bateador, y donde cada jugador es responsable en gran parte de sus estadísticas, sus papeles individuales son claros y quizás interactúen menos con el resto del equipo que en otros deportes.

En el tema electoral norteamericano, la base de datos no es tan completa como en el béisbol, aunque en la práctica, los posibles ganadores solo son los dos principales partidos políticos, con una situación política caracterizada por la estabilidad y la alternancia.

Tanto los datos como las reglas en ambos campos son públicos y están al alcance de todos aquellos que lo deseen, habiéndose impuesto el joven Silver a otros expertos y estadísticos en resolver la incertidumbre que desde luego existía por muchos datos que se tuviesen.

La clave sería no solo las estadísticas de variables aisladas (al alcance de todos), sino captar la estructura subyacente del tema, las interrelaciones del sistema, saber interpretar y utilizar la señal desechando el sobreabundante ruido, donde el criterio del pronosticador marcaría la diferencia.

Dicho de otra manera, combinar adecuadamente los conocimientos estadísticos cuantitativos con los conocimientos cualitativos de un tema (béisbol, política, etc.), así se identificarían las relaciones relevantes entre los datos. Muchas veces es *a posteriori* cuando estas relaciones cobran sentido y aparecen tan claras que sorprende no haberlas detectado antes, pues ahí estaban.

Para Silver el pecado original de la predicción es priorizar los intereses políticos, económicos o el ego, sobre la veracidad del pronóstico.

Método 1. Algoritmo y Teorema de Bayes

En el béisbol y con los datos estadísticos mencionados, Silver diseñó en 2003 un algoritmo: Pecota, que se puede definir como un sistema de predicción del rendimiento de los jugadores, donde parte de su éxito estriba en compararlos con las estadísticas de otros jugadores similares del pasado, con un acierto más que notable.

Como Silver es también un gran aficionado al baloncesto posteriormente diseñó un algoritmo (Carmelo²⁸) para pronosticar la carrera futura de los jugadores de la NBA, aunque no ha tenido el éxito de su precursor (Pecota).

En el campo político, acertó en 49 de los 50 estados norteamericanos la elección presidencial de 2008, las primeras que ganaría Barack Obama y también en su reelección en 2012.

Respecto al Teorema de Bayes, vio la luz hacia 1763 y desde entonces se aplica en los campos más diversos, desde la rotura de códigos de la máquina «Enigma» en la Segunda Guerra Mundial, hasta en temas de Inteligencia²⁹.

Es un método de probabilidad condicionada, nos indica la probabilidad de ocurrencia de un suceso condicionada por la ocurrencia de otro/s. Podemos ilustrarnos en la red con ejemplos didácticos muy claros, pero en otros casos puede ser difícil identificar cuáles son los sucesos que intervienen y la probabilidad objetiva de los mismos, por lo que al principio es conveniente contar con un guía que nos introduzca.

Silver es un entusiasta de este teorema y podemos suponer que tras cada nueva entrada de datos, lo aplica para refinar el pronóstico anterior y aproximarse sucesivamente al acierto.

Método 2. Juicio experto

Silver no precisa si en su procedimiento, el juicio experto precede a Bayes o como se interrelacionan, aunque parece que en base a los datos estadísticos y las opiniones de los expertos (que tiene muy en cuenta) aplica este teorema para realizar su predicción. Cuando Silver estudia cómo se realiza el pronóstico meteorológico, señala que los

²⁸ SILVER, Nate. Blog «*FiveThirtyEight*». CARMELO NBA Player Projections. Disponible en: <http://fivethirtyeight.com/features/how-were-predicting-nba-player-career/> <http://projects.fivethirtyeight.com/carmelo/pau-gasol/>. Fecha de consulta: 25.05.2016.

²⁹ WEATON Kristan «*Top 5 Intelligence Analysis Methods: Bayesian Analysis (# 5)*». Blog Sources and Methods 2008. Disponible en: http://sourcesandmethods.blogspot.com.es/2008/12/top-5-intelligence-analysis-methods_08.html Fecha de consulta: 25.05.2016.

meteorólogos revisan y corrigen con un lápiz óptico la pantalla de pronósticos facilitada por el ordenador, mejorándola un 10% en temperaturas y un 25% en precipitaciones.

En béisbol también sigue a expertos, aunque los representantes más genuinos de este tipo de juicio (los ojeadores) entrasen en conflicto con las nuevas formas de contratar a los jugadores, como se refleja en la película *Moneyball*³⁰, porque ¿quién acierta más a la hora de contratar nuevos jugadores, los ojeadores expertos o el estadístico con su programa informático?

En temas políticos cita un blog norteamericano que le sirve de referencia.

También menciona la web de apuestas *Intrade*³¹ sobre temas de actualidad política, financiera, climatológica, científica, entretenimiento etc. Comprobando que la media de las predicciones, sobre todo a largo plazo, reduce el error de las predicciones individuales entre un 15% y un 20%.

Evaluación de método 1. Algoritmo y Teorema de Bayes

Para Armstrong³² los principios para evaluar los métodos de pronóstico están basados en los procedimientos científicos generalmente aceptados, detallando una extensa lista, entre otros: Evaluar la fiabilidad y validez de los datos, describir potenciales sesgos de los pronosticadores, comparar pronósticos generados con distintos métodos, evaluar la validez aparente, replicar (repetir experimento en condiciones similares), pronósticos para valorar su fiabilidad, etc.

Tanto Armstrong como Tetlock, son autores de referencia para Heuer, Pherson y Silver, citándoles frecuentemente.

Los resultados cosechados con Pecota en béisbol han sido muy notables, no así en baloncesto con Carmelo.

El Teorema de Bayes es consustancial a la metodología pronosticadora de Silver; sin embargo, es en gran parte desconocido en el ámbito «Inteligencia y Seguridad»

³⁰ Película «*Moneyball*», reseña disponible en <http://www.labutaca.net/peliculas/moneyball/> Fecha de consulta: 25.05.2016.

³¹ «*Intrade*» web de apuestas sobre temas de actualidad. Disponible en: <https://prev.intrade.com/v4/home/> Fecha de consulta: 25.05.2016.

³² ARMSTRONG J. Scott. «*Standards and practices for forecasting*». 2001, pp. 26-34. Disponible en: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1146&context=marketing_papers Fecha de consulta: 25.05.2016.

aunque prometedor si se le pierde el miedo y se recibe formación para aplicarlo, como muestra este estudio³³.

Evaluación de método 2. Juicio experto

En los equipos de béisbol, los ojeadores perviven, si bien con una plantilla algo más reducida que antes y complementados con estadísticos.

Silver realizó su propia evaluación del juicio experto en política, aunque más modesta que el estudio de Tetclok, mediante el seguimiento de los pronósticos políticos realizados por expertos en un programa semanal de televisión. Revisó casi 1.000³⁴, 1/4 de ellos eran muy ambiguos o muy lejanos para tenerlos en cuenta, del resto, los pronósticos «total» o «fundamentalmente erróneos» igualaban a los «total» o «fundamentalmente acertados», era como jugársela a cara o cruz.

Silver también repasa el campo económico donde la mayoría de los economistas no creían que estuviéramos en recesión (1990, 2001 y 2007) hasta que se «determinó» que ya habían empezado³⁵.

Quizás por todo esto, Silver señala que en nuestra sociedad existe una demanda de expertos, pero no de predicciones precisas.

En resumen, al juicio experto se le puede sacar aún más partido complementándolo con otros métodos o mediante el juicio colectivo.

Calidad de los análisis. Enfoque norteamericano

Un poco de historia

Desde principios de los 50 y para identificar las deficiencias (falta de calidad) en sus documentos más importantes: NIE (Estimaciones Nacionales de Inteligencia), la CI recurría a realizar análisis *posmortem*³⁶, donde se revisaban los errores, examinando

³³ MANDEL David R. «*Applied Behavioral Science in Support of Intelligence Analysis*» Defence Research and Development Canada. 2009, pp. 6-16. Disponible en: <http://www.ci.cam.ac.uk/~rja14/shb09/mandel.pdf> Fecha de consulta: 25.05.2016.

³⁴ SILVER Nate. *Ibid.*, pp. 67-68.

³⁵ Información basada en los datos de la Encuesta de Pronosticadores Profesionales, Banco de la Reserva Federal de Filadelfia. Disponible en: <https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/anxious-index/> Fecha de consulta: 25.05.2016.

³⁶ KENT Sherman. «*The making of an NIE*». 1967. Sherman Kent and the Board on National Estimates. Collected essays. Disponible en: <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/making.html> Fecha de consulta: 25.05.2016.

donde existían las mayores lagunas e incertidumbres, así entre 1957 y 1958 se elaboraron 78 informes de este tipo.

En 1969, Smith³⁷, consideraba que en el caso de las estimaciones de inteligencia las dificultades para evaluar su exactitud se agravan dada su naturaleza, así se encuentra: un número ingente de ellas, acceso restringido al estar clasificadas, expresión ambigua y poco clara (aparecen en oraciones subordinadas), condicionadas a ciertos hechos (si ocurre X entonces probablemente ocurrirá Y), o con hipótesis asociadas a distintas probabilidades que dificultan una interpretación clara, pocas veces la interpretación será directa e inequívoca (Si-No). También puede ocurrir que como consecuencia de un adecuado sistema de alerta y una temprana y acertada estimación (más de una desde luego), se hayan tomado medidas y se haya cambiado el futuro que en ellas se preveía, entonces ¿estaban equivocadas?

En 2009, Wheaton³⁸ reivindicaba los aciertos en el análisis, no solo extraer lecciones de lo que no funciona, sino de lo que se había demostrado que funcionaba.

En 2012, Marrin³⁹, además de la exactitud, también consideraba que debía evaluarse de forma retrospectiva otros aspectos de la calidad del análisis: sí impedían la sorpresa y su influencia en la política.

Se trataba de identificar los fallos (relacionados con los sesgos del analista, pensamiento grupal, burocracia) y en consecuencia se elaboraban recomendaciones para solucionarlos. Como difícilmente se erradicaría la sorpresa, o los errores, totalmente, se debería convivir con cierta «tolerancia al desastre», lo que hoy se llamaría resiliencia, aunque en ciertos temas (Armas de Destrucción Masiva) mejor que el error sea mínimo.

Marrin, proponía trasladar el peso de la evaluación al marco del decisor, ya que es «la única persona cuya opinión realmente importa», su percepción sobre la relevancia, influencia y utilidad del análisis serían los elementos a evaluar. Aunque tiene sus inconvenientes, puesto que el decisor puede pensar que la inteligencia es buena en

³⁷ SMITH Abbot E. *Ibid.*

³⁸ WHEATON, Kristan. «*Evaluating intelligence: Answering questions asked and not*» International Journal of Intelligence and Counterintelligence 22/4, 2009, p. 629 Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/08850600903143122?journalCode=ujic20> y <http://sourcesandmethods.blogspot.com.es/2009/02/part-9-final-thoughts-evaluating.html> Fecha de consulta: 25.05.2016.

³⁹ MARRIN Stephen. *Ibid.*

tanto apoye sus propias posiciones e ignorarla en caso contrario, así la Administración Bush habría usado Inteligencia no para informarse en el proceso de toma de decisiones sino para justificar una decisión ya tomada.

Para Marrin, habría que partir de una base compartida: la finalidad del análisis, existiendo distintas visiones; ser exacto, impedir la sorpresa o ser útil, en cualquier caso, concluía que para alcanzar el éxito se contribuía más mejorando la calidad del análisis que tratando de eliminar el fallo.

Queda claro pues, que evaluar de una forma objetiva los análisis de inteligencia, en condiciones reales o simulando un contexto realista, no es tarea fácil.

Pero, por muy razonable que sea hacer constar la dificultad de la evaluación de los análisis y seamos conscientes de ello, dicha evaluación se va a producir inevitable e intuitivamente por parte de los destinatarios (decisores o lectores), tan pronto como transcurra el tiempo que se lo permita, valorando la exactitud de la estimación resultante.

Sobre el resultado recae en la realidad la máxima atención, siendo más difícil de identificar los fallos en el proceso de análisis por lo que estos, previsiblemente, perduraran más en el tiempo.

Pero a pesar de las dificultades, e inconvenientes no es un tema que podamos ignorar, evaluar la exactitud de las estimaciones es posible y deseable⁴⁰.

Visión actual

Para el profesor Coulthart la calidad de un análisis de inteligencia tendría dos componentes: rigor suficiente y exactitud⁴¹. Un análisis es riguroso cuando lo es en toda su profundidad, el rigor está relacionado con la coherencia y seguimiento de un conjunto de normas, que según investigadores de la Universidad de Ohio (Zelik, Patterson y Woods⁴²) serían estas 8 dimensiones: «Exploración de hipótesis»,

⁴⁰ FRIEDMAN Jeffrey A, ZECKHAUSER Richard «*Why assessing estimative accuracy is feasible and desirable*». Intelligence and National security 2014. Disponible en: <http://hks.harvard.edu/fs/rzeckhau/Assessing%20Estimative%20Accuracy.pdf> Fecha de consulta: 25.05.2016.

⁴¹ COULTHART Stephen J. *Ibid.*, p. 14.

⁴² ZELIK Daniel, PATTERSON Emily S. y WOODS David D. «*Understanding rigor in information analysis*». June 2007. Proceedings of the eight International NDM Conference. Disponible en: https://www.researchgate.net/publication/228809190_Understanding_rigor_in_information_analysis Fecha de consulta: 25.05.2016.

«búsqueda de información», «validación de la información», «análisis del sesgo de las fuentes», «análisis de sensibilidad», «colaboración de especialistas», «síntesis de información» y «explicación crítica», con un seguimiento de estos criterios más «suficiente» que metodoso. Para ellos, el rigor en análisis de información, o el rigor analítico, refleja una valoración de la calidad del proceso antes que del producto (resultado) del análisis.

Sobre el rigor, el término «suficiente» depende del contexto y de la envergadura del análisis (complejidad, disponibilidad de datos y tiempo) ya que un proyecto que implique un desarrollo de varios meses requerirá más rigor que un proyecto a desarrollar en un día.

Para Coulthart, el análisis es exacto cuando existe un alto grado de correspondencia entre el juicio analítico y lo que ha pasado en el mundo real, se podría decir, con el transcurso del tiempo, que el análisis ha acertado.

Por su parte, el director Nacional de Inteligencia estableció mediante una Directiva, en junio de 2007⁴³, unas Normas de elaboración y evaluación de sus productos analíticos que, a pesar de lo expuesto anteriormente, apenas cambió en su actualización de enero de 2015⁴⁴.

Articula la responsabilidad de los analistas de inteligencia «para procurar la excelencia, integridad y rigor». En lo que es su interpretación de la calidad, los análisis de la Comunidad de Inteligencia norteamericana deben seguir las normas siguientes:

- a. Objetivos.
- b. Independientes de consideraciones políticas.
- c. A tiempo.
- d. Basados en todas las fuentes de Inteligencia.
- e. Cumplir con las normas de análisis siguientes:
 1. Describir apropiadamente la calidad y credibilidad de las principales fuentes, datos y metodologías.

⁴³ Office of the Director of National Intelligence. «*Analytic Standards. Intelligence Community Directive 203*». 21.06.2007. Disponible en: <https://www.hsdl.org/?view&did=744163> Fecha de consulta: 25.05.2016.

⁴⁴ Office of the Director of National Intelligence. «*Analytic Standards. Intelligence Community Directive 203*». 02.01.2015. Disponible en: <http://www.dni.gov/index.php/intelligence-community/ic-policies-reports/intelligence-community-directives> Fecha de consulta: 25.05.2016.

2. Expresar apropiadamente y explicar la incertidumbre asociada con los principales juicios analíticos. Probabilidad de ocurrencia de un suceso (expresada en 7 términos y porcentajes: remoto: 1-5%, altamente improbable: 5-20%, etc.) y la confianza del analista en su juicio (términos y porcentajes numéricos más precisos que en la Directiva de 2007).
3. Distinguir apropiadamente entre la principal información de inteligencia y las suposiciones y juicios del analista.
4. Incorporar hipótesis alternativas.
5. Demostrar relevancia para el cliente y señalar las implicaciones del análisis.
6. Usar argumentación clara y lógica.
7. Explicar si existe cambio o continuidad, con juicios anteriores.
8. Realzar juicios y evaluaciones precisas.
9. Incorporar información visual efectiva cuando proceda (norma no contemplada en la Directiva de 2007).

Es la visión oficial del rigor analítico por parte de la CI norteamericana, mientras que los 8 atributos del rigor de los investigadores de la Universidad de Ohio, antes expuestos, representan el punto de vista académico del tema. Siendo benévolo podemos observar coincidencias sobre fuentes (diversas, creíbles, objetivas) y exposición de resultados (diversas hipótesis, argumentación objetiva y clara), pero la Directiva, incorpora al proceso de elaboración un componente de utilidad: las necesidades del destinatario, y las concretas: a tiempo, relevancia e implicaciones, cambio o continuidad. También incorpora la metodología de análisis, primero, al hablar de objetividad, señala que los analistas deben emplear técnicas de razonamiento para mitigar los sesgos y posteriormente menciona: «Describir apropiadamente la calidad y credibilidad de las principales fuentes, datos y metodologías. Los productos analíticos deben identificar las principales fuentes y las metodologías en las que están basados los juicios» (la Directiva de 2007 no incluía metodologías, solo fuentes), por lo que habría que recordar la jerarquía metodológica y que en función de la metodología empleada el rigor del análisis variará. Sin embargo, los investigadores de Ohio no consideran la metodología como un atributo del rigor.

Se observa como las normas a) y b) (objetivos e independientes de consideraciones políticas) son redundantes, lo que lleva a pensar que han tenido serios problemas de este tipo en el pasado y una de sus prioridades es evitarlos. Así lo manifiesta el profesor Mitchell⁴⁵ sobre el intento partidista de manipulación de análisis de inteligencia en 2003 sobre Irak.

Estas normas son utilizadas también para evaluar sus análisis, apreciándose que algunas de ellas son más fáciles de verificar que otras. Que un análisis llegue a tiempo requiere de criterios de comprobación más imparciales que, por ejemplo, la comprobación de las normas a) y b) citadas, más difíciles de comprobar, de hecho el Congreso norteamericano lo ha investigado en numerosas ocasiones sin encontrar evidencias de politización⁴⁶.

Así mismo, se supone que estas comprobaciones difícilmente podrán ser realizadas por alguien extraño a la propia Comunidad de Inteligencia, sino ¿cómo valorar si el análisis está basado en todas las fuentes de inteligencia?, precisamente la protección de las fuentes de inteligencia es un asunto especialmente sensible. En cualquier caso, estas normas indican claramente que los análisis deben sujetarse a ellas, en un afán de transparencia y de facilitar la revisión del proceso para su mejora.

Sobre la exactitud, la Directiva de 2007 señala que hay que realizar juicios y valoraciones precisos, aunque la exactitud a veces es difícil de establecer y solo puede ser evaluada retrospectivamente si la información necesaria es obtenida y se encuentra disponible. La Directiva de 2015, también repite que hay que realizar juicios y valoraciones precisos, de forma que el cliente (destinatario) tenga claro el mensaje enviado por el analista. Esa claridad de significado permitirá valorar la exactitud cuándo toda la información necesaria esté disponible.

Aunque el cumplimiento de estas normas persiga el rigor y la exactitud del producto, no los asegura.

⁴⁵ MITCHELL, Gordon R. «*Team B intelligence coups*». Quarterly Journal of Speech. Volume 92, Issue2, 2006, pp. 144-173. Disponible en: <http://www.pitt.edu/~gordonm/JPubs/TeamBqjs.pdf>

⁴⁶ ROSENBAACH Eric. «*National Intelligence Estimates. Memo in report Confrontation or Collaboration?*» Congress on the Intelligence community. 2009. Disponible en: http://belfercenter.ksg.harvard.edu/publication/19150/national_intelligence_estimates.html Fecha de consulta: 25.05.2016.

Por último, esta Directiva menciona las responsabilidades de distintos directivos de la CI para aplicar estas Normas y desarrollar un amplio programa de evaluación de los análisis, lecciones aprendidas e identificación de las mejores prácticas.

Por ello, es de suponer que como antes señalaban las NIE a las que realizar un análisis *posmortem*, ahora también señalen que NIE requieren un seguimiento en el tiempo (retroalimentación) para disponer de información y comprobar su exactitud, cuantas más se evalúen más posibilidad de extraer conclusiones válidas.

Así lo ha hecho un investigador, Mandel, en Canadá, con este estudio⁴⁷ publicado en 2009, donde no parece que el seguimiento y valoración precisa de la exactitud de las estimaciones constituya un problema. En él, revisó 649 predicciones elaboradas entre marzo de 2005 y octubre de 2006 por la División de Oriente Medio y África del *Intelligence Assessment Secretariat (IAS)*. El 89% de ellas fue posible codificarlas en términos claros de resultados (no ambiguos). La valoración de la exactitud se realizó utilizando criterios objetivos: calibración (grado en que la probabilidad predicha coincide con la observada) y discriminación (grado en el que se distingue entre casos en los que ocurre el evento y los que no). Las predicciones fueron calificadas de «buenas a excelentes».

Posteriormente, este mismo investigador, Mandel⁴⁸, extendió su estudio desde marzo de 2005 a diciembre de 2011, cubriendo un periodo aproximado de seis años y revisando 3.881 pronósticos de inteligencia estratégica. Desechando los juicios condicionados, imprecisos y otros, valoró la ocurrencia o no de 1.514 pronósticos (39% del total) teniendo en cuenta y definiendo los criterios de: experiencia del analista y la dificultad, importancia y marco temporal de resolución del pronóstico⁴⁹.

⁴⁷ MANDEL David R. «*Applied Behavioral Science in Support of Intelligence Analysis*». Defence Research and Development Canada. 2009, pp. 17 y sucesivas. Disponible en: <http://www.ci.cam.ac.uk/~rja14/shb09/mandel.pdf> Fecha de consulta: 25.05.2016.

⁴⁸ MANDEL David R., BARNES Alan. «*Accuracy of forecasts in strategic intelligence*». PNAS 2014. Disponible en: <http://www.umass.edu/preferen/You%20Must%20Read%20This/MandelBarnesPNAS2014.pdf> Fecha de consulta: 25.05.2016.

⁴⁹ Tabla Excel con datos del estudio de MANDEL y BARNES de 2014. Dataset S1. Disponible en: www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406138111/-/DCSupplemental/pnas.1406138111.sd01.xls y «*Supporting information*» Disponible en: <http://www.pnas.org/content/suppl/2014/07/11/1406138111.DCSupplemental/pnas.201406138SI.pdf> Fecha de consulta: 25.05.2016.

Calidad de los análisis. Enfoque español

A falta de normas de acceso público de la Comunidad de Inteligencia española, que señalen los requisitos de calidad (rigor) del proceso analítico, se puede usar como referencia un punto de vista muy interesante y antes señalado: los destinatarios de los informes de Inteligencia, según el estudio del profesor Díaz⁵⁰ de 2006, que se anticipa a las conclusiones de Marrin de 2012, con la fuerza que otorga haber recogido las opiniones de 56 políticos y altos cargos y 23 embajadores españoles.

Así, los políticos y altos cargos entrevistados valoran de mayor a menor importancia: brevedad de los informes, que lleguen a tiempo y su relevancia; algo que los productores de inteligencia suelen ordenar de forma inversa, según dicho profesor.

La valoración de los embajadores consultados es la siguiente: que lleguen a tiempo, sean completos y fiables, estén listos para ser empleados, sean relevantes, estén ajustados a las necesidades de cada consumidor, fáciles de comprender y aplicar su contenido.

En resumen, el profesor Díaz destaca en el apartado «Formato y calidad de la Inteligencia»: A tiempo, preparada para actuar, precisión, objetividad, utilización de los informes y relevancia.

Este enfoque permite reflexionar sobre la finalidad principal de los análisis en el ámbito de Inteligencia y Seguridad que no es otra que facilitar la toma de decisiones anticipatorias; por tanto, debemos considerar el criterio «utilidad» en la calidad de los análisis. Dicha «utilidad» estaría en función de si el análisis ayudó al destinatario del mismo a tomar una decisión, con independencia de que esta con el tiempo se demostrara acertada o errónea, ya que el criterio «exactitud» lo valoramos aparte y si lo tuviéramos en cuenta, estaríamos considerándolo por duplicado.

Los requisitos para valorar esa utilidad señalados por el profesor Díaz, coinciden en varios puntos (objetividad, a tiempo, relevancia) y se adelantan a las normas analíticas de la CI norteamericana, reafirmando que las autoridades de Inteligencia

⁵⁰ DÍAZ FERNÁNDEZ Antonio M. «*El papel de la comunidad de inteligencia en la toma de decisiones de la política exterior y seguridad de España*». Observatorio de política exterior de España. Fundación Alternativas. Documento de Trabajo 3/2006. Disponible en:

http://www.fundacionalternativas.org/public/storage/opex_documentos_archivos/xmlimport-aPrKiN.pdf

Fecha de consulta: 25.05.2016.

estadounidenses, al contemplar esas normas, pensaban no solo en el proceso de elaboración del análisis sino en su entrega y utilidad para el decisor.

En cuanto a la precisión mencionada por el profesor Díaz, se asimilaría a la exactitud. Otros factores aludidos (preparada para actuar, utilización de los informes), podrían ser tenidos en cuenta por la CI española a la hora de determinar sus normas analíticas.

Más adelante el profesor Díaz⁵¹ señala: «Uno de los principales problemas detectados es la falta de evaluación del producto de inteligencia. Nos encontramos con un sistema en su mayor parte unidireccional y donde se consume lo que llega, pero no se evalúa de forma sistemática ni su calidad, ni su utilidad, ni su eficacia».

Limitaciones y retos

Como se ha podido observar, las referencias principales están relacionadas con autores anglosajones, ya que además de precursores, ofrecen grandes facilidades para la divulgación y discusión de sus trabajos, queda por tanto, realizar un estudio sobre los autores españoles.

Tampoco se ha tratado un campo muy relevante en estudios de futuro: la prospectiva, que puede ser objeto de otro estudio para averiguar sus resultados. Igualmente no se ha profundizado en el análisis y pronóstico de resultados electorales, al exceder los objetivos de este artículo.

Por otra parte, y como una aplicación de los conceptos teóricos aquí tratados, recientemente se ha publicado un artículo con sugerencias prácticas para mejorar la calidad de un análisis, que se concretan en una lista de comprobación de 15 puntos.⁵²

Como el tema expuesto es extenso y todavía hay mucho que decir, tan pronto como vea la luz este análisis, se abrirá un debate en LinkedIn en el grupo «Analistas de Inteligencia».

⁵¹ DÍAZ FERNÁNDEZ Antonio M. *Ibíd.*, p. 49.

⁵² TOBOSO MARQUÉS, Fernando. «Sugerencias prácticas para mejorar la calidad de un análisis de inteligencia. Lista de comprobación de 15 puntos». 9 de mayo de 2016. Análisis GESI (Grupo de Estudios en Seguridad Internacional) 11/2016. Disponible en: (<http://www.seguridadinternacional.es/?q=es/content/sugerencias-pr%C3%A1cticas-para-mejorar-la-calidad-de-un-an%C3%A1lisis-de-inteligencia-lista-de>). Fecha de consulta: 25.05.2016.

Conclusiones

1. Exactitud de estimaciones. Evaluación del juicio experto y de las Técnicas Analíticas Estructuradas

El Juicio Experto ha sido evaluado al menos parcialmente en el mundo anglosajón y distintos ámbitos, con los resultados sobre su exactitud y limitaciones expuestas. Si previamente logramos identificar a los mejores «zorros», sigue siendo muy valioso e imprescindible, en muchos aspectos de estudio de un tema: divulgación, reflexión, actualización y también estimación del futuro.

Las TAE todavía están en fase de difusión e implantación, máxime en nuestro país, por lo que todavía no han sido suficientemente evaluadas y comprobadas su exactitud y posibilidades.

2. Exactitud de pronósticos. Evaluación de la metodología de N. Silver

Silver falló en las elecciones británicas de 2010⁵³. Si las comparamos con las elecciones norteamericanas, el que hubiera un tercer partido en liza, una base estadística posiblemente menos completa, unas normas electorales distintas, un conocimiento personal más limitado, y seguramente sin juicios expertos de referencia, explicarían su «falta de éxito».

Tampoco en baloncesto ha cosechado el éxito que en béisbol, con lo que la generalización de los resultados obtenidos con sus métodos es limitada pero ¿quién no firmaría con que su método «funcione» en un deporte o en las elecciones de un país como EE.UU.?

El mérito en la exactitud de los pronósticos realizados por Silver reside en la combinación de métodos: algoritmo, Teorema de Bayes, juicio experto y..., su genio personal.

3. Evaluación de la calidad de los análisis

A igualdad de metodología empleada (según lo visto sobre jerarquía metodológica) los criterios que influyen en la calidad de los análisis y que serían los aspectos evaluables de los mismos, podrían ser:

⁵³ ULFELDER Jay «[Why the World Can't Have a Nate Silver](http://foreignpolicy.com/2012/11/08/why-the-world-cant-have-a-nate-silver/)» *Foreign Policy*, 08.11.2012. Disponible en: <http://foreignpolicy.com/2012/11/08/why-the-world-cant-have-a-nate-silver/> Fecha de consulta: 25.05.2016.

- Exactitud o acierto (validez), correspondencia entre la estimación y lo sucedido. No siempre fácil de valorar ni aun transcurrido el tiempo, según lo visto. Aunque intuitivamente realicemos esta valoración, conviene contar con asesoramiento estadístico y juzgar objetivamente sus componentes: calibración y discriminación, previamente se han podido «codificar» las estimaciones en términos claros, desechando las ambiguas. No es imposible, otros ya lo han hecho.
- Rigor y utilidad (validez interna), de acuerdo al cumplimiento de unas normas prefijadas. El análisis ha sido diseñado y realizado asegurando su coherencia, no solo en el proceso de elaboración sino atendiendo a su finalidad: ayudar a su destinatario en la toma de decisiones.

La prioridad o ponderación entre ellos puede ser objeto de estudio y puede depender del ámbito de estudio/caso concreto, plazo de estimación, dificultad e importancia o las exigencias del destinatario (tolerancia al error).

Corresponde al organismo productor del análisis establecer los criterios de calidad a cumplir por el mismo y que servirán en su elaboración (rigor y utilidad) y evaluación.

A esta última, colaborarán con dichos organismos, los destinatarios de los análisis, pues «el valor de la inteligencia es aquel que le otorgue el consumidor, ese es el verdadero y único indicador»⁵⁴.

Una vez establecidos los criterios de evaluación y su ponderación, se estará más cerca de realizar una evaluación adecuada de la calidad de los análisis y realizarla sistemáticamente, pues solo así se averiguará cómo, cuándo, en qué y quienes realmente realizan análisis de calidad.

Una de las máximas dificultades en dicha evaluación reside en la definición del cuerpo teórico que lo sustente, aquí los especialistas del mundo académico pueden contribuir en gran manera. La voluntad de todos para conseguirlo y aplicarlo hará el resto.

Conclusiones finales

- El formato escrito de un análisis de ciencias sociales (estudios de Inteligencia, estratégicos, etc.) no va a coincidir, ni indicar, la metodología utilizada en su elaboración. Los más complejos pueden además incluir diversos análisis realizados

⁵⁴ DÍAZ FERNÁNDEZ Antonio M. *Ibid.*, p. 61.

con distintas metodologías, por lo que se deben identificar los criterios para la evaluación tanto del análisis como de la metodología.

- Un componente esencial de la calidad del análisis es la metodología con la que se ha elaborado.
- No solo se debe tener en cuenta el resultado (exactitud) sino también el proceso (rigor).
- Las normas exigibles para la elaboración de un análisis (rigor) serán los factores sobre los que posteriormente será evaluado, además de la exactitud.

*Fernando Toboso Marqués**
Profesor Metodología de Análisis
CISDE