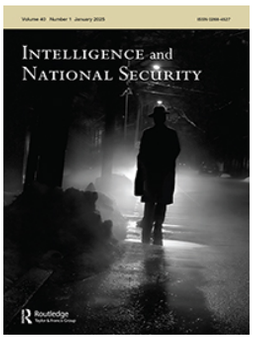


Intelligence and National Security

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/fint20

How do we know if an intelligence analytic product is good?

Kevin Riehle

To cite this article: Kevin Riehle (21 Feb 2025): How do we know if an intelligence analytic product is good?, Intelligence and National Security, DOI: 10.1080/02684527.2025.2468051

To link to this article: https://doi.org/10.1080/02684527.2025.2468051

Published online: 21 Feb 2025.

INTELLIGENCE AND NATIONAL SECURITY
https://doi.org/10.1080/02684527.2025.2468051

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

How do we know if an intelligence analytic product is good?

Kevin Riehle

ABSTRACT

How can an intelligence analysis production organization determine whether analysis is successful? This article explores the three methods that intelligence communities have applied to determine whether analysis is good: Did the analysis meet analytic tradecraft standards? Were the assessments accurate? And did the product make a difference with a decision maker? Unfortunately, none of those evaluation methods is perfect and all three leave questions. It can be just as difficult to determine whether analysis is good as it is to produce intelligence analysis itself. However, all three methods can identify products that approach the ideal.

KEYWORDS

Intelligence analysis; analytic standards; accuracy; evaluation; Hydrogen; intelligence failure

Discussions abound about times when intelligence analysis has been bad. Richard Betts published the seminal work on the topic in 1978, and (Jervis's 2011) book, *Why Intelligence Fails*, has become a standard work on the subject.[1] Researchers have used the 1941 Pearl Harbor attack, lack of warning about the Korean War, the Yom Kippur War, the dissolution of the Soviet Union, terrorist attacks in September 2001, and intelligence leading up to the Iraq War to explain how failures occurred.[2] More recently, the Hamas attack on Israel in October 2023 has renewed these discussions. Often, the blame for failure is placed on analysts. Biases, political influence, insufficient communication between analysts and decision makers, analysis arriving too late, and lack of interagency sharing and coordination can all get in the way of good analysis. Universities even offer courses on intelligence failures.[3]

With all that in the background, the question remains: What is the standard against which an intelligence product is graded? How can an analytical production organization determine whether analysis is successful? What is good analysis, as opposed to bad?

Intelligence analysis has three broad purposes: 1) to make sense out of chaos; 2) to inform of an approaching threat or opportunity; and 3) to persuade to action by providing the consumer with the needed information at the right time to do something about the topic. In being persuasive, an analyst is not telling a consumer what to do but presents information in such a way that the consumer arrives at the same conclusion as the analyst and thereby obtains decision advantage. Brian Holmes, a senior U.S. intelligence officer, calls that "ethical persuasion: 'Ultimately, analysts are guided to produce intelligence in order to ethically persuade an audience, in their case a senior decision-maker, while conforming to the analytic tradecraft standards described in Intelligence Community Directive 203 (ICD 203)'.[4] For that to happen, the analysis must be well organized, cogently argued, and solidly documented.

Intelligence agencies have struggled with questions of analytic success for years. Billions of dollars (or other currencies) have been spent on collecting and analyzing intelligence. Debates in the political realm discuss the money spent on intelligence and the return on investment from that money, some arguing that it is justified by the level of security it provides, others saying that it is more of a burden than an advantage.[5] While other countries' intelligence structures are not as large

CONTACT Kevin Riehle kevin.riehle@brunel.ac.uk

or as expensive as the intelligence community (IC) in the United States, they also try to measure the value of their analytic products.[6] Those debates would benefit from data about successes as well as failures.

In the wake of two commissions that investigated analytic lapses related to the September 2001 terrorist attacks in the United States and the 2003 U.S. invasion of Iraq, the U.S. Congress set in motion a series of actions intended to raise the U.S. IC's level of analytic rigor and prevent similar lapses from reoccurring. President George W. Bush signed the Intelligence Reform and Terrorism Prevention Act (IRTPA) in 2004, which mandated procedural and methodological changes in how the U.S. IC fulfilled its intelligence analytic requirements. That led to the Director of National Intelligence promulgating Intelligence Community Directive (ICD) 200, 'Management, Integration, and Oversight of Intelligence Analysis', which required the IC to 'implement policies and procedures to encourage sound analytic methods and tradecraft throughout the elements of the IC; to ensure that analysis is based upon all sources available; and to ensure that the elements of the IC regularly conduct competitive analysis of analytic products, whether such products are produced by or disseminated to such elements'.[7] Also emerging from IRTPA was ICD 203, 'Analytic Standards', which lists specific standards to which IC analysis should adhere and to which intelligence analysts should be trained.[8]

Such documents require intelligence organizations to evaluate their intelligence analysis to facilitate improvements in tradecraft and practice. Sometimes, evaluations of analytic products are mandated by failures, and other times they are good faith internal attempts to improve. Whichever the motivation, intelligence agencies have applied three types of evaluation criteria to determine whether analysis has succeeded.

(1) Did the analysis meet analytic tradecraft standards? Was it well-written, well-argued, and persuasive? Even if it was wrong occasionally, did it meet the baseline standards?
(2) Were the assessments accurate? Did the forecast come to pass as the analyst said it would?
(3) The ultimate question: Regardless of whether a product met all the tradecraft standards, did it make a difference? If the goal of an intelligence analytic product is to 'ethically' persuade a decision maker to do something, did the decision maker, in fact, act on the intelligence?

Building on the concept that Stephen Marrin introduced in his 2012 article, 'Evaluating the Quality of Intelligence Analysis: By What (Mis)Measure?', this article discusses all three evaluation methods that intelligence communities have tried to implement and the complications that have arisen from them.[9] Unfortunately, none of them is perfect and all three leave questions.

Whether a product is good or bad is not black and white, with gradations in between and some portions of a product having greater impact than others. Intelligence services have used the three evaluation methods to determine whether they are approaching anything close to good. Nevertheless, as with the intelligence analysis products they are intended to evaluate, precise answers can often be difficult to obtain. While academics frequently point out when intelligence analysis has historically been bad, this article argues that, despite efforts to define good analysis, it is just as difficult to determine what that is as it is to produce intelligence analysis itself.

This article will primarily use the U.S. IC as a case example because that is where the author spent a career and because the amount of information available about intelligence analysis in the U.S. system is significantly greater. Some material from other intelligence systems will also help to illustrate points.

Quality assessment process

The U.S. IC has developed mechanisms for looking back at analytic products and grading them based on adherence to analytic tradecraft standards. Robert Gates, the U.S. Director of Central Intelligence from 1991 to 1993, noted in his Senate confirmation testimony that the CIA had created a Product Evaluation Staff as early as the 1980s.[10] Thus, such efforts have been in existence since long before

IRTPA, although that legislation codified them and mandated them across the IC. With the advent of ICD 203, those processes became more formal, with clear criteria for judging products across the whole IC.

ICD 203 describes a set of standards for producing and evaluating analytic products. There are five primary Standards:

- Objectivity
- Independence from political considerations
- Timeliness
- Being based on all available sources of intelligence information
- Implementing and exhibiting the Analytic Tradecraft Standards

The Analytic Tradecraft Standards state that IC products should have the following characteristics:

(1) Properly describes quality and credibility of underlying sources, data, and methodologies
(2) Properly expresses and explains uncertainties associated with major analytic judgments:
(3) Properly distinguishes between underlying intelligence information and analysts' assumptions and judgments
(4) Incorporates analysis of alternatives
(5) Demonstrates customer relevance and addresses implications
(6) Uses clear and logical argumentation
(7) Explains change to or consistency of judgments
(8) Makes accurate judgments or assessments
(9) Incorporates effective visual information where appropriate

As mandated by IRTPA, the Office of the Director of National Intelligence (ODNI) created an organization called the Analytic Integrity and Standards Group (AIS) that conducts ICD 203 reviews of a randomly sampled selection of U.S. IC analytic products annually and submits a report to Congress summarizing its findings. AIS evaluators grade analytic products' adherence to ICD 203 standards on a 0 to 3 scale, with zero being poor and three being excellent. Products are assessed by two graders who, after grading the product separately, discuss their grades together to reach a consensus. The grades are then reviewed by a third independent grader.[11]

AIS grades products against all the analytic tradecraft standards except standard 8: 'Makes accurate judgments and assessments'. That is because of the difficulties in determining accuracy (see discussion of this difficulty below). Based upon its review of all the other standards, AIS can make recommendations about areas where intelligence analysts require more training or reminders.

Each agency in the U.S. IC has an equivalent product evaluation process to grade analytic products in search of areas for improvement. For example, the Defense Intelligence Agency (DIA) has a Product Evaluation Board (PEB) that mirrors the AIS process. The PEB selects products randomly every six months and assigns them to teams of two analysts to review, encouraging all analysts to participate at least once during their careers. Evaluators are chosen from a different division than the one that authored the product, meaning they are evaluating a product they would not have seen previously and about which they may not be subject matter experts. Two analysts review products independently, score each ICD 203 analytic tradecraft standard on a 0 to 3 scale, and then meet to compare their scores. Ideally, the two analysts score the product close to the same. If their scores are different, they discuss them and come to a consensus. They provide a narrative explanation for each score.

The DIA process has been in existence since about 2008 and has accumulated data that shows trends in how well DIA products adhere to analytic tradecraft standards. DIA developed a matrix of characteristics that analytic products should exhibit and that should guide analysts' development of

assessments. It is divided into three levels of characteristics: critical, required, and when appropriate. The critical characteristics include ensuring

- the product has a 'so what'
- the bottom line is clear and upfront
- it gives analysis, not just facts
- it is accurate and timely
- it uses clear, logical argumentation

Required characteristics include anticipating readers' questions, and being free of bias, grammatical errors, and typos. These characteristics align closely with ICD 203 analytic tradecraft standards to assist analysts in writing products that meet the standards. CIA has a similar construct with a few characteristics that are worded slightly differently, but which analysts use for the same purpose.

Analysts at DIA are encouraged to use the matrix from the beginning of the drafting stage. However, it is also used in the post-publication evaluation stage as a set of criteria for determining whether a product follows standards. The scores are reported in aggregate, not individually to analysts, meaning they are not intended to critique an individual but the analytic process overall. They are reported to identify analytic tradecraft areas that DIA needs to emphasize in training and management.

In 2011–2012, this author served as the functional manager for counterintelligence analysis in the Department of Defense (DoD). A product evaluation process was in place when I arrived at that position, but it was not serious and provided little usable data. The DoD counterintelligence analytic community reinvigorated product evaluations by applying DIA's process and started gathering useful data for DoD counterintelligence analysts, including those outside DIA, to use in improving their products. The program was not intended to grade individual analysts but to estimate how close analytic products approached the ideal of good tradecraft and provide recommendations based on aggregated data. Unfortunately, it did not last long after I left the position, and the results had little long-term impact across the DoD counterintelligence community.

The question of analytic tradecraft standards is not universally accepted as a positive. Although their promulgation was mandated in the IRTPA, do analytic standards guarantee that the product is good? They might make accuracy and persuasiveness more likely by reducing the occurrence of critical thinking pitfalls and unclear writing, but how can we be sure? Robert Cardillo, while serving as the DIA Deputy Director for Analysis, stated in 2010 that the PEB, combined with AIS reports, showed a steady improvement of DIA analytic products in relation to ICD 203 standards.[12] But do the grades mean the products are more successful?

Josh Kerbel provided a contrarian point of view, asking, 'What if the tradecraft standards too often become the be-all, end-all of what we call analytic quality?'[13] Rather than writing to a list of standards, he encourages intelligence agencies to increase the creative thinking of modern analysts, who, he says, are faced with a surfeit of information that requires creative approaches to analyze 'complex systems … defined as much, if not more, by the connections between the components as by the components themselves'. He states that creativity is at odds with following a standardized set of rules because 'by one definition, [creativity is] about breaking the rules'.[14]

Van Gelder similarly warned that grading to the standards could lead to a '"box-ticking" approach where a rater – particularly, presumably, one who is tired or bored – checks for superficial signs of adherence rather than "deep quality".'[15] Marcoci, Vercammen, and Burgman also find that evaluations using ICD 203 standards can vary widely, and that, to provide meaningful grades, analysts performing the analysis need to be trained to calibrate their evaluations.[16] Those concerns involve execution of the reviews rather than the validity of the standards themselves. Nevertheless, they raise potential doubts about how well evaluators apply the standards. The DoD counterintelligence analytic community took these concerns seriously, providing standardization training for all analysts

assigned as quality evaluators to calibrate ratings. AIS also offers training for all product evaluators. However, training cannot completely overcome evaluators' different backgrounds and mindsets, especially when they represent various agencies, as in AIS evaluations.

Thorburn, et al., assess that analytic quality ratings based upon ICD 203 standards are not a valid measure of analytic rigor. Using a 'LOTSA' definition of rigor – a combination of logicality, objectivity, thoroughness, stringency, and acuity – they assert that there is no good measure of analytic rigor, including ICD 203 standards, although one should be developed because 'analytic rigor is a cardinal virtue of intelligence'.[17]

Further, Standard 8, accuracy of analysis, is often excluded from product evaluations due to the uncertainty surrounding analytic accuracy. Mark Lowenthal asserts that following analytic standards will not necessarily produce accurate analysis, although they can help analysts avoid pitfalls.[18] That raises the question of whether the grading system misses the essence of what analysis should be doing, providing accurate judgments. Intelligence communities have attempted to address that question as well.

Accuracy in intelligence analysis

Of the analytic tradecraft standards, the most difficult to judge is accuracy. If intelligence analysis is intended to inform, then the information must be accurate. Intelligence products often contain assessments that forecast future developments, giving decision makers 'decision advantage' as they seek to navigate the changing competitive geopolitical environment. But intelligence analysis can only deliver decision advantage if it is accurate – inaccurate information or forecasts actually harms decision makers' ability to respond to a situation, as studies of analytic failures have pointed out. ICD 203 analytic tradecraft Standard 8 says that an analytic product should 'apply expertise and logic to make the most accurate judgments and assessments possible, based on the information available and known information gaps'.[19]

One of the measures that intelligence agencies have applied to their products to determine whether they are successful is to review them in hindsight and assess whether the forecasts they contain were accurate. Accuracy could include forecasting what will or will not happen, when it will happen, what is the likelihood that it will happen, who will be involved, and what the consequences will be. But how does an agency tell if a forecast is accurate?

Several intelligence systems have tried to judge accuracy with varying results. In 1951, the UK Joint Intelligence Committee (JIC) commissioned a project to 'review the conclusions of their main studies on Communist intentions since January, 1947 in order to determine to what extent they have been proved correct by subsequent events and, in cases where their conclusions have been proved wrong, to discover why false conclusions were drawn'.[20] The study reviewed thirty-six JIC assessments from 1947 to 1951, of which it categorized twenty-five as 'proved correct', eight as 'not yet proved incorrect', and three as incorrect. All three of the incorrect assessments involved China. The study concluded that the reason for the inaccurate judgments was that UK intelligence knew less about communist intentions in the Far East than about the Soviet Union, largely because of a lack of SIGINT and because 'we do not yet understand the mind of the Communist Chinese leaders'.[21]

However, the study stated further that the small number of inaccurate assessments told only part of the story. The JIC completely missed the Berlin Blockade in 1948 and the North Korean attack on South Korea in 1950. In a moment of remarkable self-awareness, the JIC evaluators added that sometimes the judgments were drafted 'somewhat equivocally and so allowed ourselves a fairly wide margin of error'.[22] In other words, the assessments were so hedged that they could have been right or wrong, but it was hard to tell.

A larger number of assessments-eight of thirty-six – that were 'not yet proved incorrect' added a further level of complication to the task of judging accuracy: how long do evaluators wait to judge whether the assessment has come to pass? Some products forecast months or years into the future, so when does an evaluator gauge accuracy.[23] In other cases, the timeframe was less explicit. In

March 1950, the JIC assessed the likelihood of a war with the Soviet Union. The assessment initially judged that such a war was unlikely because Soviet leaders believed in the eventual victory of communism over capitalism anyway, so war would not be necessary. However, if Soviet leaders viewed their own military power as sufficient to counter any resistance, or if they received signals that the West was planning an offensive attack, they might order the initiation of war to exploit the opening or protect Soviet interests. The request from the JIC was for an assessment of the probable date when the Soviet Union would consider itself prepared to risk a major war. The assessment, however, while indicating that no war would be likely before 1952, was open ended in timeframe and could have applied to the rest of the Cold War.[24]

In 1955, the U.S. Office of National Estimates (ONE), the organization that produced National Intelligence Estimates (NIE), tried similarly to determine whether the assessments and forecasts in NIEs were correct. Like the JIC several years earlier, the U.S. IC tried to determine why it did not forecast that North Korean forces would attack South Korean forces in 1950, nor that the Chinese would join them when the North Korean forces looked likely to be defeated. That omission was perceived as an intelligence failure, which prompted ONE to look systematically at NIE judgments. As they did after World War II in relation to the Pearl Harbor attack, evaluators in the 1950s sought to determine how to avoid surprise. That same question arose again decades later when the 9/11 Commission tried to determine how to prevent the surprise of another terrorist attack.

To answer that question, the ONE conducted 'validity studies' to determine how often intelligence assessments were accurate. Using a baseball analogy, the IC was trying to determine its 'batting average' in hitting the mark on its assessments. Sherman Kent, the first director of ONE, wrote, 'Few things are asked the estimator more often than "How good is your batting average?" No question could be more legitimate – and none could be harder to answer'.[25] CIA analyst Abbot Smith, Kent's deputy at ONE and later its director, wrote in 1969, 'It would seem reasonable to suppose that one could get a truly objective, statistical verdict on the accuracy of estimates. Go through the papers, tick off the right judgments and the wrong ones, and figure the batting average. I once thought that this could be done, and I tried it, and it proved to be impossible. The reasons are various'.[26]

The CIA studied the key judgments of dozens of NIEs to determine whether the key judgments had come to pass. They allowed some time to pass since the forecast was made, usually 6 to 12 months, because events did not occur the day after the NIE was published.[27] If we follow Kent's baseball analogy, an average of over 300, calculated as a player hitting the ball three out of ten times at bat, would be considered very good. Using another sports analogy, a good quarterback in American football completes 70 per cent of passes. What if the analyst got three out of ten judgments correct? What about 7 out of 10? Could that be considered a good average for an intelligence analyst?[28]

The question validity studies were trying to answer was, were we right and how often? Similar to the 1951 JIC study, validity studies determined three possible conclusions:

- On target: Judgments were or remain valid
- Off the mark: Judgments were flawed or inaccurate
- Cannot tell: Unable to determine validity at this time

What did 'on target' look like? It might mean a clear-cut analysis of events subsequent to the judgment showed that the forecast was accurate. The judgment forecasted a coup, and a coup happened. The judgment forecasted the number of missiles the Soviet Union would produce, and evidence showed that they did, in fact, produce those missiles. According to Kerr and Warner, there were few major Soviet weapons systems that the U.S. IC did not identify and determine their capabilities. That could be considered an analytic success.[29]

Being 'on target' might fall somewhere between all three conclusions, but the general forecast was accurate. It might mean that, although some of the individual judgments were not borne out,

the overall assessment was still accurate – the big picture was right, but individual details might have turned out a little differently. The analysis forecasted a coup and it did occur, but the leaders of the coup were not who the CIA predicted them to be. It might mean that the overall judgment was generally accurate, but related events were still ongoing, and a complete evaluation was not yet possible. It looked like things were headed in the right direction, but as the JIC also assessed, it was too early to tell for sure.

'Off the mark' meant that the product overestimated or underestimated the likelihood that an event would occur or simply wrote something that turned out to be incorrect. Sudden changes in the environment took the course of events in a different, unanticipated direction. Maybe foreign leaders acted in a way that appeared totally contradictory to their own interests and thus was difficult to predict. Some regimes, like North Korea's, might make decisions that do not seem rational. When a judgment was determined to be inaccurate, it was important to determine why it was so, not just to document the error. As the JIC did, analysts asked what went wrong? What did we miss? 'Off the mark' judgments tend to get the most visibility both inside and outside the IC.

Sometimes, validity studies could come to a conclusion about whether an analytic judgment was accurate or not, but often, they could not. Validity studies never reached the point of developing a clear 'batting average'. The vagaries of foreign leaders and the absence of confirmatory information sometimes meant that determining whether a judgment was accurate or inaccurate was impossible. The author of one validity study stated, 'In theory the making of a validity study should be a simple matter – get out the old papers, read them, and note whether the estimates turned out to be true or false. In practice, it is not that simple. Indeed, it is so much more complicated and difficult that it has proved in many respects to be impossible, and this study has turned out quite differently from what its author had hoped it would'.[30]

A lack of visibility into a foreign government's decision making might have made it hard to tell whether the situation had developed in the direction the judgment had forecasted or not. Collecting feedback data can be as difficult as collecting the original data. The long-term nature of a situation might make it impossible to know whether assessments will eventually turn out to be correct. One of the validity studies in the 1950s stated, 'The words "right, correct, accurate", and so on, when applied to our estimates, must still be taken in a provisional sense. Only in a comparatively small number of instances can we be perfectly sure that we were "right".'[31]

As the JIC learned, sometimes forecasts are not written in a clearly confirmable manner. A more recent study by Mandel and Barnes concluded that analysts in the Middle East and Africa Division of the Canadian government's Intelligence Assessment Secretariate (IAS) were remarkably accurate in their forecasts of whether an event would or would not occur. The study employed two subject matter experts not affiliated with the IAS to code the outcomes of the forecasted events and to quantify forecast quality. However, while the analysts forecasted false negatives only 5 per cent of the time and false positives 7 per cent of the time – a remarkable 'batting average' – only 80 per cent of individual forecasts were articulated unambiguously enough to make a clear judgment of whether the forecast was accurate.[32]

Similarly, AIS has attempted to include accuracy in its product evaluations when requested. The 2016 declassified AIS report of Office of Naval Intelligence (ONI) analytic products judged that eighty-seven percent of Intelligence Community future judgments 'tested accurate', while only about fifty-six percent of ONI products met that standard. However, forty-five percent of all judgments were 'unclear' compared to those that attached some sort of condition. AIS assessed accuracy separately from other tradecraft standards in the overall assessment and did not attempt to show a year-on-year trend analysis.[33]

Some have criticized the CIA for not predicting the dissolution of the Soviet Union. They ask, how could the CIA, which focused nearly all its resources and energy on the Soviet Union, not see the end coming.[34] Journalist Uri Friedman ranked it as one of America's ten biggest intelligence failures.[35] CIA analysts claim that criticism is unfair. They say they did

accurately assess the weakening Soviet economy, the increase in societal unrest, and the centrifugal forces pulling the Soviet republics apart. The fact that they did not predict that Mikhail Gorbachev would appear on television on 25 December 1991 and proclaim the end of the Soviet Union should not overshadow the accurate assessments CIA analysts had made leading up to that.[36] They complain that no organization can expect such a level of precision when faced with human decisions.

It is also possible that a judgment occasionally appears incorrect because our own decision maker listened, enacted a policy that changed the situation on the ground, and prevented the forecasted event from occurring, what Betts and Marrin call a self-negating prophecy.[37] This could be said for forecasted terrorist attacks since 2001—many were thwarted before they could happen. Could that be considered a failure or a success? In 2018, soon after assuming the position of Commander of the Estonian Defense Forces, General Martin Herem reportedly told his chief of intelligence, 'Failure is written into your job description'. General Herem cautioned that if the intelligence unit raises an alarm that a war is about to start, the defense forces will listen to that forecast and start preparations and mobilization. Then, if the war does not come, intelligence will be criticized for raising the alarm, even if the preparations and mobilization themselves are what dissuaded the enemy from attacking. But if intelligence fails to raise the alarm and a war starts, that would also be seen as a failure.[38] With the latter scenario, General Herem was articulating exactly what the JIC and ONE were criticized for in relation to the Korean War. Analysis of foreign intelligence threats can fall victim to the same conundrum; an analyst can highlight an intelligence threat, leading a government to increase security and implement counterintelligence measures, which diminishes vulnerability to the threat. Does that mean the analyst's forecast was wrong?

Some things are easier to be definitively accurate about than others. Basic data that involve physical phenomena, such as missile ranges, aircraft operational parameters, etc., are among the easiest to assess. If analysts have quantitative or physical data on which to make a judgment, if scientific accuracy is possible, they have a better chance not only of making the correct judgment but also of being able to look back and tell whether it was correct or not. Other trends can be generally predictable within some range, such as economic and demographic forecasts, and can approach the level of accuracy that is possible with almost scientific certainty. Demographic forecasts have the benefit of quantitative data that can be used to analyze a long-term, slow-moving topic, giving a forecaster time to see trends and develop assessments based on them.

Intelligence questions to which definitive answers are possible are what Gregory Treverton calls 'puzzles' as opposed to 'mysteries'. Puzzles have a clear answer, and that answer can be checked. Treverton further states, 'for the mysteries of intelligence, measures of effectiveness are elusive'.[39] Some parts of an analytic assessment may come to pass precisely as an analyst forecasted, while other parts are overcome by unforeseen or unknowable events. In irregular areas, like political decision making, where the forecast involves individual human decisions as opposed to large, aggregated population trends or quantitative data, it can be difficult not only to make the assessment in the first place, but also to look back and determine whether the assessment was correct. Terrorist attacks fall into this category because they involve a human decision that can change at the last moment, even if an intelligence agency receives warning signs of an impending attack. As the CIA's validity studies in the 1950s looked back at analytic judgments in these types of areas, they found them to be either less likely to be accurate or more often impossible to assess the accuracy.

Thus, post-production reviews intended to determine whether analysis is accurate or not can often be obscured by several factors: it may be too early to tell; some readers' expectations for forecasts may be too high or they may mistakenly believe that intelligence analysis is a form of crystal ball; forecasts might be so hedged that it is difficult to determine whether they were right or wrong; situations on the ground might change based on the assessment itself, which then makes it appear as if the assessment was incorrect; or the problem may be too full of uncertainty to know for sure – a mystery as opposed to a puzzle.

Did the product make a difference?

The third way to assess whether an analytic product is good is to determine whether it made a difference in decision making. This can be the hardest of the three measures to determine. However, if intelligence analysis is supposed to support decision making, it is also the most crucial. An intelligence product may be just one of multiple inputs that a decision maker receives, making it impossible to determine whether a specific product was the reason the decision maker did something. To fully understand whether a product made a difference, we need to crawl into the decision maker's head to determine how much influence a product had on the ultimate decision.

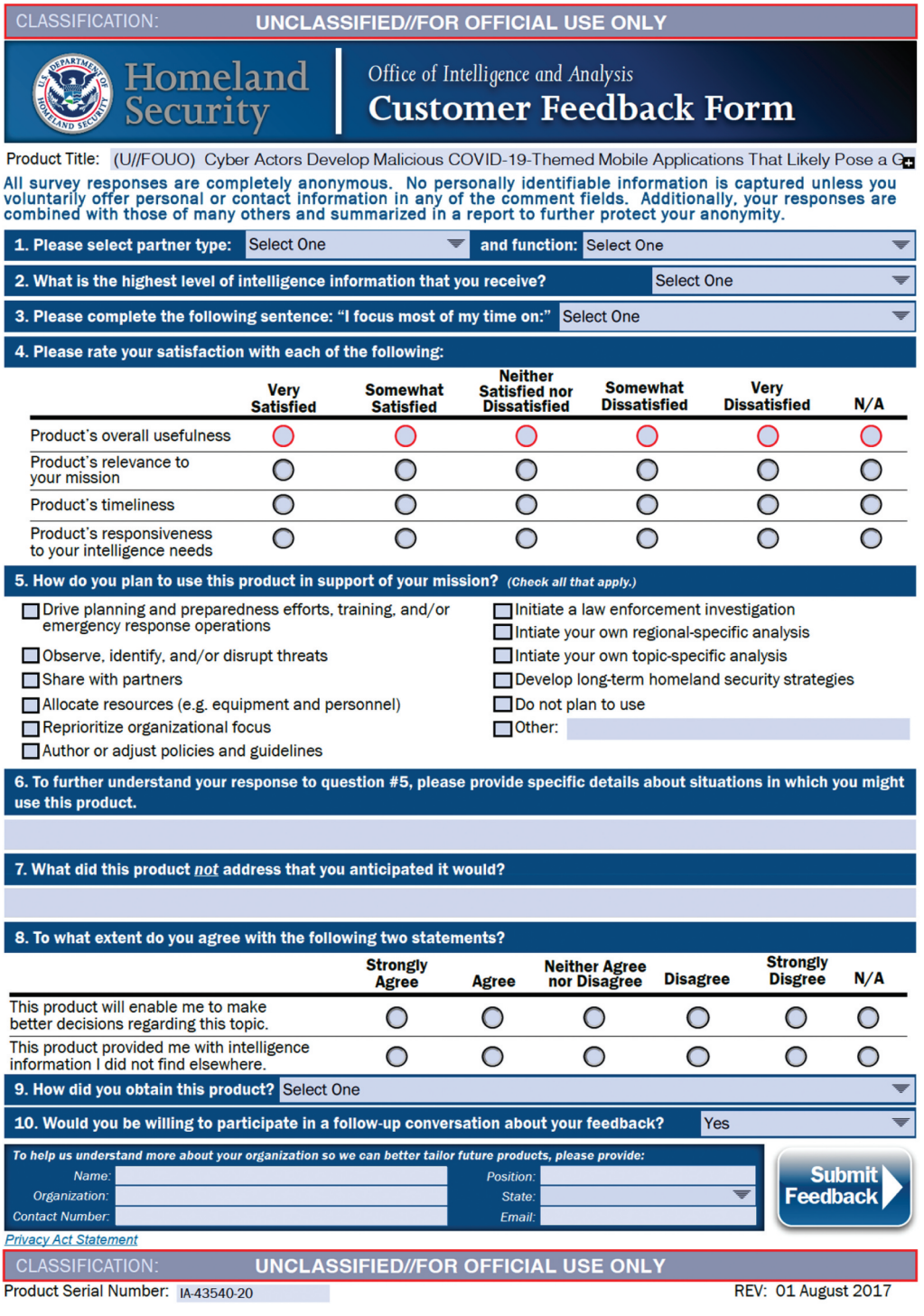
Decision makers are sometimes loath to admit that their own policy has gone awry. Thomas Fingar, a former senior U.S. intelligence community analyst, claimed, 'I learned something a long time ago in this town. There are only two possibilities: policy success and intelligence failure'.[40] An intelligence analyst might lament that if something goes right, the policymaker takes credit for a good policy, but if something goes wrong, the policymaker blames the IC for providing bad intelligence. That can be tempting to do because the analyst behind the intelligence product is usually publicly nameless.

If an intelligence product does successfully persuade a decision maker to act, what is more likely to be seen publicly is the result of the decision, not the intelligence that went into it. John McLaughlin, a retired Deputy Director of Central Intelligence, said, 'Intelligence successes are woven into successful policies in invisible ways'.[41] McLaughlin cited as a key example the 2004 diplomatic effort to persuade Libya to disarm, when Libyan leader Muammar Gaddafi agreed to eliminate his country's weapons of mass destruction program. The agreement was seen as a diplomacy victory, but it, in fact, was preceded by six months of intensive intelligence work.

The CIA began early in its existence to track the impact that its analytic products had on policy. Beginning as early as 1955 and running into the 1960s, the CIA compiled a quarterly list of NIEs that factored into U.S. national security policy. The CIA began compiling these quarterly reports while Sherman Kent was director of the Officer of National Estimates, and at the same time as that organization conducted its first validity study. For example, according to one of the quarterly lists, the July 1958 Special National Intelligence Estimate (SNIE) titled, 'Sino-Soviet and Free World Reactions to US use of Nuclear Weapons in Limited Wars in the Far East' was used in the development of portions of NSC 5810/1, 'Note by the Executive Secretary to the National Security Council on Basic National Security Policy', dated May 1958.[42] The CIA must have briefed the National Security Council while the SNIE was still in draft. The CIA also reported that a July 1958 SNIE, 'Implications of Certain US Earth Satellite Programs', was used in developing NSC 5814/1, 'Preliminary U.S. Policy on Outer Space'.[43] Quarterly reports listed numerous such examples.

It is easier to see impact in some applications of intelligence analysis than in others. Analysts who directly feed law enforcement actions or tactical operations, often with close organizational or literal proximity to the consumer, more often see the results of their direct analytic support. Analysts who are embedded in a policy office may also have that benefit. But even proximity is not a guarantee. This author experienced this on many occasions as an analyst. In one case, I wrote an analytic product that was delivered to a consumer, and a few days later, an operation occurred that was in direct line with what I wrote. I will never know whether my product impacted that action or not because the consumer provided no feedback.

Intelligence agencies constantly try to devise ways to gauge whether a decision maker read a product and whether it was useful and, more importantly, influential. IC analytic products often include feedback forms or surveys asking – practically begging – for some kind of feedback. Forms ask about the usefulness and timeliness of the product, whether it met the customer's requirements, and whether it was used in decision making (See Figure 1). However, as is typical with customer feedback in any industry, only a small number of readers fill out the form and express their views of the product, and those that do tend to be either the most satisfied or the most unsatisfied. Thus, feedback forms can be useful in gauging product value but represent only a small portion of readers.



CLASSIFICATION: UNCLASSIFIED//FOR OFFICIAL USE ONLY

Homeland Security

Office of Intelligence and Analysis

Customer Feedback Form

Product Title: (U//FOUO) Cyber Actors Develop Malicious COVID-19-Themed Mobile Applications That Likely Pose a G

All survey responses are completely anonymous. No personally identifiable information is captured unless you voluntarily offer personal or contact information in any of the comment fields. Additionally, your responses are combined with those of many others and summarized in a report to further protect your anonymity.

1. Please select partner type: Select One and function: Select One

2. What is the highest level of intelligence information that you receive? Select One

3. Please complete the following sentence: "I focus most of my time on:" Select One

4. Please rate your satisfaction with each of the following:

	Very Satisfied	Somewhat Satisfied	Neither Satisfied nor Dissatisfied	Somewhat Dissatisfied	Very Dissatisfied	N/A
Product's overall usefulness	○	○	○	○	○	○
Product's relevance to your mission	○	○	○	○	○	○
Product's timeliness	○	○	○	○	○	○
Product's responsiveness to your intelligence needs	○	○	○	○	○	○

5. How do you plan to use this product in support of your mission? *(Check all that apply.)*

- ☐ Drive planning and preparedness efforts, training, and/or emergency response operations
- ☐ Observe, identify, and/or disrupt threats
- ☐ Share with partners
- ☐ Allocate resources (e.g. equipment and personnel)
- ☐ Reprioritize organizational focus
- ☐ Author or adjust policies and guidelines
- ☐ Initiate a law enforcement investigation
- ☐ Intiate your own regional-specific analysis
- ☐ Intiate your own topic-specific analysis
- ☐ Develop long-term homeland security strategies
- ☐ Do not plan to use
- ☐ Other:

6. To further understand your response to question #5, please provide specific details about situations in which you might use this product.

7. What did this product *not* address that you anticipated it would?

8. To what extent do you agree with the following two statements?

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disgree	N/A
This product will enable me to make better decisions regarding this topic.	○	○	○	○	○	○
This product provided me with intelligence information I did not find elsewhere.	○	○	○	○	○	○

9. How did you obtain this product? Select One

10. Would you be willing to participate in a follow-up conversation about your feedback? Yes

To help us understand more about your organization so we can better tailor future products, please provide:

Name: Position:
Organization: State:
Contact Number: Email:

Submit Feedback

Privacy Act Statement

CLASSIFICATION: UNCLASSIFIED//FOR OFFICIAL USE ONLY

Product Serial Number: IA-43540-20 REV: 01 August 2017

Figure 1. Department of Homeland Security Intelligence and Analysis Customer Feedback Form. Although marked "For Official Use Only", this product is available on the open internet at https://www.gnyha.org/wp-content/uploads/2020/04/U-FOUO-HIA-Malicious-COVID-19-Themed-Mobile-Apps-Likely-Pose-Growing-Threat-to-Third-Party-App-04232020-002.pdf.

Another way intelligence agencies collect customer feedback is by placing a human briefer in front of the customer and observing reactions. A briefer who provides the President's Daily Brief, for example, does not just deliver a product to a decision maker and walk away, but listens to the decision maker's comments and questions and conveys them back to the IC. National Intelligence Officers (NIO) and Defense Intelligence Officers (DIO) meet regularly with senior decision makers, listen to their reactions, and then communicate them to analysts. An early description of NIOs identified their duties to include 'identify[ing] customer needs for national intelligence'.[44] DIOs serve as liaisons with senior elements of the Office of the Secretary of Defense, the Joint Chiefs of Staff, and the Executive Office of the President to both provide intelligence and to listen to their needs. They conduct annual surveys in which they interview decision makers to ask how useful the intelligence they received was. Sometimes, a decision maker can point to a product and say, 'that one really helped me in my decision process'. But often, the decision maker can only speak in general terms because no one intelligence product sticks out as being particularly influential. DIA's Executive Support Office supports defense leaders in the Pentagon and monitors their reactions to the analysis – did they read it, ask follow-up questions, critique it?[45] Decision makers give feedback most often when they are dissatisfied, so analysts will more often know when a piece of intelligence did not matter than when it did.

ODNI's Office of Analytic Integrity and Standards (AIS) also conducts annual surveys of intelligence consumers across U.S. government policy-making agencies, asking questions about product objectivity, timeliness, accuracy, usefulness, and sourcing. Consumers AIS surveyed in 2015 rated U.S. Intelligence Community analytic products as generally positive, with some variation, particularly on the question of sourcing. However, as the summary focused on the Office of Naval Intelligence (ONI), AIS specifically addressed satisfaction with ONI products. Only fifteen percent of respondents reported even having seen ONI products.[46]

When Bill Clinton became president in 1993, he ordered a national performance review to evaluate the effectiveness of all U.S. government functions, including the IC. One of the recommendations for the IC was titled 'Enhance Community Responsiveness to Customers'. Action items under that recommendation included appointing a customer advocate or ombudsman and establishing a process that continuously tracks the needs of the IC customer.[47] As of January 1996, the General Accounting Office (GAO), which Congress tasked with monitoring fulfillment of the recommendations, reported that the Director of Central Intelligence refused to supply documents related to the fulfillment of the IC recommendations citing the CIA's view that intelligence oversight was limited to specific congressional committees.[48] Essentially, the CIA did not recognize the GAO's role in overseeing the CIA, even though the spirit of the recommendation was to enhance the usefulness of IC products for decision makers. Whether that happened has not been publicly revealed.

Even products that meet analytic standards and are accurate might not affect policy. In October 1990, the U.S. National Intelligence Council published an NIE titled 'Yugoslavia Transformed'. It forecasted, with prescient accuracy, the dissolution of Yugoslavia into ethnic-based republics, armed uprisings by Albanians in Kosovo, and consequent severe intercommunal conflict. It predicted that within a year, the Yugoslav federal system would no longer exist and Yugoslavia would dissolve as a state.[49] It also concluded, 'there is little the United States and its European allies can do to preserve Yugoslav unity'. The product was well written. Although it preceded ICD 203, it would have scored well against most standards. Its judgments were accurate – Croatia and Slovenia formally declared independence from Yugoslavia in June 1991, followed soon thereafter by Macedonia and Bosnia and Herzegovina. The latter separations spurred violent ethnic-based reactions, as the NIE forecasted. However, despite both those evaluation measures being remarkably positive, the NIE had little impact on U.S. national security decision making.

In 2015, Gregory Treverton and Renanah Miles sought an answer for why that happened. They concluded that the Yugoslavia analysis did not connect with decision makers' priorities at the time. Decision makers were focused on other pressing issues, like unrest in the Soviet Union. The NIE was published in October 1990, less than a year after the fall of the Berlin Wall and in the midst of other

tectonic changes occurring in Eastern Europe. The NIE also faced an unwelcoming audience in Brent Scowcroft as National Security Advisor and Lawrence Eagleburger as Deputy Secretary of State. Both were 'Yugoslav hands' who felt they were experts in the country and were loath to accept the NIE's conclusions. Treverton and Miles further concluded that the NIE lacked opportunity analysis – in fact, it concluded that there was little the United States could do. Finally, the then-recent reunification of Germany made it more difficult to accept the opposite effect in Yugoslavia. All those factors combined to result in the Yugoslavia NIE falling on deaf ears.[50]

More recently, as Israel looks back to determine how it missed signals that would have warned of an impending Hamas attack, some claim that the problem was not with collection and analysis but with the dissemination of intelligence. Journalist Ben Kaspit asserted that the role of intelligence is not only to send the alert but to make sure it is received. He blamed a recent innovation in the Israeli intelligence community for the lack of proactive delivery of intelligence: a database from which decision makers could pull information rather than the intelligence community pushing intelligence to the appropriate consumer. Intelligence analysis cannot have an impact if it never reaches the decision maker.[51]

The aspiration for intelligence analysis to have policy impact and the ability to measure that impact are complicated by the fact that the readers of analytic products are complex people weighed down with competing priorities and confident that their policy is already heading in the right direction. In some cases, assessments may only resonate with a consumer when they align with already-held views. As Marrin notes, intelligence analysis that contradicts the decision maker's preferred direction, even if it is well written and accurate, can struggle to get the attention the analyst desires.[52] But it cannot hope to penetrate the decision-making cycle if it is not delivered in a persuasive, compelling form.

Conclusion

None of these three evaluation methods – grading against analytic tradecraft standards, assessing the accuracy of judgments, and measuring influence on customers' decision making – is foolproof. Each has downsides that reflect the complex task of reducing uncertainty for decision makers and increasing decision advantage. While there are many publicized examples of instances when products did not adhere to analytic tradecraft standards and thus arrived at incorrect conclusions, when assessments were later proven inaccurate, or when good analysis went unnoticed, there are also positive examples for all three. As the CIA's quarterly reports of NIEs used in NSC policy formulation attest, intelligence analysis has, at times, factored directly into policy. Yet, the Yugoslavia NIE example shows that getting the first two measures right does not necessarily guarantee the third: influence on a decision maker. Assessing whether analysis is good is as difficult a problem as forecasting an adversary's next action. Both involve understanding the mind of a decision maker, whether the adversary's or our own.

What, then, should an intelligence agency do? While none of the three methods is perfect and each can individually lead to false positives or negatives, all provide data that can be used to improve analysis. They can highlight analysis that communicates clear assessments and offers persuasive and articulate exemplars the analysts can emulate. Although a perfect analytic product that demonstrates all tradecraft standards, is presciently accurate, and persuades decision makers to act is hard to find, all three evaluation methods can identify products that approach that ideal. Both success and failures provide valuable lessons, and agencies can use successes identified through evaluation methods to reward existing staff, train new employees, and as steppingstones toward even higher goals.

Notes

1. Betts, "Analysis, War, and Decision"; and Jervis, *Why Intelligence Fails*.

2. See, for example, Kahn, "The Intelligence Failure of Pearl Harbor"; Unsinger, "Three Intelligence Blunders"; Penney, "Intelligence and the 1973 Arab-Israeli War"; and Friedman, "The Ten Biggest"; National Commission on Terrorist Attacks Upon the United States, *9/11 Commission Report*.
3. See, for example, Leiden University; University of Leicester; and University of Buckingham.
4. Holmes, "Pathos, Where Art Thou?".
5. See, for example, Mulligan and Schmitt, "What the Intelligence Community Doesn't Know"; Robinson, "The Viability"; and Judis, "The Case for Abolishing the CIA".
6. See, for example, Leuprecht and McNorton, *Intelligence as Democratic Statecraft*.
7. Office of the Director of National Intelligence, Intelligence Community Directive 200.
8. Office of the Director of National Intelligence, Intelligence Community Directive 203.
9. Marrin, "Evaluating the Quality of Intelligence Analysis".
10. U.S. Senate, Select Committee in Intelligence, *Nomination of Robert M. Gates*.
11. Van Gelder, "The ODNI Rating Scale".
12. Cardillo, "Intelligence Community Reform".
13. Kerbel, "Are the Analytic Tradecraft Standards".
14. Ibid.
15. Van Gelder, "The ODNI Rating Scale".
16. Marcoci, Vercammen, and Burgman, "ODNI as an Analytic Ombudsman".
17. Thorburn, et al, "The IC Rating Scale".
18. Lowenthal, "Towards a Reasonable Standard for Analysis," 307.
19. Office of the Director of National Intelligence, Intelligence Community Directive 203.
20. Joint Intelligence Committee, "Review of Assessment Made Since January 1947".
21. Joint Intelligence Committee, "Review of Assessments of Communist Intentions since January, 1947.
22. Ibid.
23. For example, Joint Intelligence Committee, "The Possibility of War Before 1956" and "Soviet Intentions and Capabilities in 1950–1954".
24. Joint Intelligence Committee, "The Likelihood of War with the Soviet Union".
25. Steury (ed.), "The Law and Custom," 100.
26. Smith, "On the Accuracy of National Intelligence Estimates." Marchio, "'How Good is Your Batting Average?", Marrin, "Evaluating the Quality of Intelligence Analysis," and Tang, "How do we know? All refer to the batting average metaphor.
27. Central Intelligence Agency, "List of 69 Validity Studies"; and Marchio, "'How Good is Your Batting Average?".
28. Lowenthal, "Towards a Reasonable Standard for Analysis," 308–309.
29. Kerr and Warner, "The Track Record of CIA Analysis," 41.
30. Marchio, "How Good is Your Batting Average," 8.
31. Marchio, "How Good is Your Batting Average," 6.
32. Mandel and Barnes, "Accuracy of Forecasts".
33. Office of the Director of National Intelligence, "Analytic and Standards Evaluation Program: Key Results from FY 2015 (ONI and IC)".
34. Arbel and Edelist, *Western Intelligence and the Collapse of the Soviet Union: 1980–1990.*
35. Friedman, "The Ten Biggest".
36. Lundberg, "CIA and the Fall of the Soviet Empire"; and Berkowitz, "U.S. Intelligence Estimates".
37. Betts, "Analysis, War, and Decision"; and Marrin, "Evaluating the Quality of Intelligence Analysis".
38. Roonemaa and Epner, "How Estonia's Military Intelligence".
39. Treverton, "The Soviet Union Was a Puzzle."Ibid.
40. "Official: Iraq War was Both Intel, Policy Failure".
41. Aspen Institute, "Keeping Intelligence Smart".
42. U.S. Intelligence Board, "Quarterly Report"; and Keefer and Mabon (eds.), *Foreign Relations of the United States, 1958–1960*, Document 24.
43. U.S. National Security Council, "Preliminary U.S. Policy on Outer Space".
44. U.S. Intelligence Community Staff, "A Guide to the National Intelligence Community's Production Organizations".
45. Defense Intelligence Agency Public Affairs Office, "DIA and the Presidency"; and Noone, LinkedIn profile.
46. Office of the Director of National Intelligence, "Analytic and Standards Evaluation Program: Key Results from FY 2015 (ONI and IC)".
47. U.S. General Accounting Office, *Management Reform: Implementation*.
48. U.S. General Accounting Office, *Management Reform: Completion Status.*
49. U.S. National Intelligence Council. "Yugoslavia Transformed".
50. Treverton and Miles, "Unheeded Warning of War".
51. Caspit, "A Day After the Massacre".
52. Marrin, "Why strategic intelligence analysis has limited influence".

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Kevin Riehle is a lecturer in intelligence and international security at Brunel University London. He spent over 30 years in the U.S. government as a counterintelligence analyst studying foreign intelligence services, retiring as an associate professor of strategic intelligence at the National Intelligence University in 2021. He subsequently spent two years as an associate professor at the University of Mississippi, Center for Intelligence and Security Studies before coming to Brunel. He received a PhD in War Studies from King's College London, an MS of Strategic Intelligence from the Joint Military Intelligence College, and a BA in Russian and Political Science from Brigham Young University. He has written on a variety of intelligence and counterintelligence topics, focusing on the history of Soviet and Eastern Bloc intelligence services. His next book, *The Russian FSB: A Concise History of the Federal Security Service*, was published in May 2024.

ORCID

Kevin Riehle http://orcid.org/0000-0002-8711-9842

Bibliography

Arbel, D., and R. Edelist *Western Intelligence and the Collapse of the Soviet Union: 1980–1990: Ten Years That Did Not Shake the World*. London: Routledge, 2003.

Aspen Institute. "Keeping Intelligence Smart: Reflections from the Intelligence World." Homeland Security Program, Panel Discussion. April 22, 2016. https://www.aspeninstitute.org/blog-posts/keeping-intelligence-smart-reflections-from-the-intelligence-world/.

Berkowitz, B. D. "U.S. Intelligence Estimates of the Soviet Collapse: Reality and Perception." *International Journal of Intelligence & CounterIntelligence* 21, no. 2 (2008): 237–250. doi:10.1080/08850600701854052.

Betts, R. K. "Analysis, War, and Decision: Why Intelligence Failures are Inevitable." *World Politics* 31, no. 1 (October 1978): 61–89. doi:10.2307/2009967.

Cardillo, R. "Intelligence Community Reform: A Cultural Evolution." *Studies in Intelligence* 65, no. 3 (September 2021): 5–10.

Caspit, B. "A Day After the Massacre, Four Former Commanders of the Unit Entered 8200 headquarters. After a Few Fours, Two of Them Came Out Shocked." *Maariv*. February 23, 2024. https://www.maariv.co.il/journalists/Article-1078519.

Central Intelligence Agency. "List of 69 Validity Studies," Undated but September 1958, CIA FOIA Reading Room, document number CIA-RDP85S00362R000600010001-3.

Commission on the Intelligence Capabilities of the U.S. Regarding Weapons of Mass Destruction, *Unclassified Version of the Report of the Commission on the Intelligence Capabilities of the U.S. Regarding Weapons of Mass Destruction*. March 31, 2005. https://www.govinfo.gov/app/details/GPO-WMD/.

Defense Intelligence Agency Public Affairs Office. "DIA and the Presidency: The President's Intelligence Officer." February 2, 2022. https://www.dia.mil/News-Features/Articles/Article-View/Article/2921536/dia-and-the-presidency-the-presidents-intelligence-officer/.

Friedman, U. "The Ten Biggest American Intelligence Failures." *Foreign Policy*. January 3, 2012. https://foreignpolicy.com/2012/01/03/the-ten-biggest-american-intelligence-failures/.

Gelder, V., and Tim. "The ODNI Rating Scale – Issues Abound." Blog Entry. May 19, 2019. https://timvangelder.com/2019/05/19/the-odni-rating-scale-issues-abound/.

Holmes, B. "Pathos, Where Art Thou? Intelligence Analysis and Ethical Persuasion." *War on the Rocks*. September 17, 2016. https://warontherocks.com/2016/09/pathos-where-art-thou-intelligence-analysis-and-ethical-persuasion/.

Jervis, R. *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*. Ithaca, NY: Cornell University Press, 2011.

Joint Intelligence Committee. "The Possibility of War Before 1956." JIC(47)76, January 23, 1948, TNA, CAB 158/2.

Joint Intelligence Committee. "Soviet Intentions and Capabilities in 1950-1954." JIC(50)92, September 28, 1950a. TNA, CAB 158/11.

Joint Intelligence Committee. "The Likelihood of War with the Soviet Union and the Date by which the Soviet Leaders Might Be Prepared to Risk it." JIC(50)7, March 13, 1950b, TNA, CAB 158/9.

Joint Intelligence Committee. *Review of Assessment Made Since January 1947 by the Joint Intelligence Committee on Communist Intentions*. London, CAB: The National Archives (TNA), Kew, 1951a September, 20.
Joint Intelligence Committee. "Review of Assessments of Communist Intentions Since January, 1947 by the Joint Intelligence Committee," JIC(51)87 (Final), 12 December 1951b, The National Archives (TNA), Kew, London, CAB 158/13, number 87.
Judis, J. "The Case for Abolishing the CIA." Carnegie Endowment for International Peace. December 20, 2005. https://carnegieendowment.org/2005/12/20/case-for-abolishing-cia-pub-17846.
Kahn, D. "The Intelligence Failure of Pearl Harbor." *Foreign Affairs* 70, no. 5 (Winter) (1991): 138–152. doi:10.2307/20045008.
Foreign Relations of the United States, 1958–1960, National Security Policy; Arms Control and Disarmament, edited by Keefer, E. C. and D. W. Mabon. Washington: Government Printing Office, 1996), Document 24, https://history.state.gov/historicaldocuments/frus1958-60v03/d24.
Kerbel, J. *Are the Analytic Tradecraft Standards Hurting as Much as Helping?* National Intelligence University Research Short, 2017 September, 14.
Kerr, R. J., and M. Warner. "The Track Record of CIA Analysis." In Analyzing Intelligence: National Security Practitioners' Perspectives, edited by R. George and J. Bruce, Second ed. 41. Washington, DC: Georgetown University Press, 2014.
Leiden University. https://studiegids.universiteitleiden.nl/en/courses/111014/intelligence-failures.
Leuprecht, C., and H. McNorton *Intelligence as Democratic Statecraft: Accountability and Governance of Civil-Intelligence Relations Across the Five Eyes Security Community—The United States, United Kingdom, Canada, Australia, and New Zealand*. Oxford: Oxford University Press, 2021.
Lowenthal, M. M. "Towards a Reasonable Standard for Analysis: How Right, How Often on which Issues?" *Intelligence and National Security* 23, no. 3 (2008): 303–315. doi:10.1080/02684520802121190.
Lundberg, K. "CIA and the Fall of the Soviet Empire: The Politics of 'Getting it Right,'." Kennedy School of Government Case Program, C16-94-1251.0. 1994. https://lundbergcc.com/wp-content/uploads/2014/12/Fall-of-Sov-Union-final.pdf.
Mandel, D. R., and A. Barnes "Accuracy of Forecasts in Strategic Intelligence." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 30 (June 2014): 10984–10989. doi:10.1073/pnas.1406138111.
Marchio, J. "'How Good is Your Batting average?': Early IC Efforts to Assess the Accuracy of Estimates." *Studies in Intelligence* 60, no. 4 (Extracts December 2016): 3–13.
Marcoci, A., A. Vercammen, and M. Burgman "ODNI as an Analytic Ombudsman: Is Intelligence Community Directive 203 Up to the Task?" *Intelligence and National Security* 34, no. 2 (2019): 205–224. doi:10.1080/02684527.2018.1546265.
Marrin, S. "Evaluating the Quality of Intelligence Analysis: By What (Mis)measure?" *Intelligence and National Security* 27, no. 6 (December 2012): 896–912. doi:10.1080/02684527.2012.699290.
Marrin, S. "Why Strategic Intelligence Analysis Has Limited Influence on American Foreign Policy." *Intelligence and National Security* 32, no. 6 (2017): 725–742. doi:10.1080/02684527.2016.1275139.
Mulligan, K., and A. Schmitt. "What the Intelligence Community Doesn't Know is Hurting the United States." Center for American Progress. September 18, 2020. https://www.americanprogress.org/article/intelligence-community-doesnt-know-hurting-united-states/.
National Commission on Terrorist Attacks Upon the U.S. "*9/11 Commission Report*." July 22, 2004. https://www.9-11commission.gov/report/911Report.pdf.
Noone, J. "LinkedIn Profile." https://www.linkedin.com/in/james-noone-14ab2724/. Noone served as the chief of the Executive Support Office in 2012 and 2013.
Office of the Director of National Intelligence. "Intelligence Community Directive 200, "Management, Integration, and Oversight of Intelligence Analysis,"." January 8, 2007a. https://www.dni.gov/files/documents/ICD/ICD_200.pdf.
Office of the Director of National Intelligence. *Intelligence Community Directive 203, "Analytic Standards,"*. 2007b June, 21. amended on January 2, 2015.
Office of the Director of National Intelligence. "Analytic and Standards Evaluation Program: Key Results from FY 2015 (ONI and IC)." August. 2016, released via FOIA, https://www.dni.gov/files/documents/FOIA/AOPS-2023/DF-2020-00235-ONI_AOPS_Records_2015_2017_2018_2019.pdf.
Official: Iraq War was Both Intel, Policy Failure. NBC News. December 10, 2008. https://www.nbcnews.com/id/wbna28148115.
Penney, M. T. "Intelligence and the 1973 Arab-Israeli War." *Center for the Study of Intelligence* (2012). https://www.cia.gov/readingroom/docs/2012-12-10E.pdf.
Robinson, P. "The Viability of a Canadian Foreign Intelligence Service." *International Journal* 64, no. 3 (Summer) (2009): 703–716. doi:10.1177/002070200906400307.
Roonemaa, H., and E. Epner. "How Estonia's Military Intelligence Secretly Helped Ukraine." *VSquare*. September 15, 2023. https://vsquare.org/how-estonias-military-intelligence-secretly-helped-ukraine/.
Senate, U. S. "Select Committee in Intelligence, *Nomination of Robert M. Gates to Be Director of Central Intelligence*." In 102nd Congress, First Session, Washington, DC: Government Publishing Office, 1991 October 3, https://www.intelligence.senate.gov/sites/default/files/hearings/102799v3.pdf.
Smith, A. E. "On the Accuracy of National Intelligence Estimates." *Studies in Intelligence* 13 , no. 4 (1969): 25–35.

The Law and Custom of the National Intelligence Estimate." In *Sherman Kent and the Board of National Estimates: Collected Essays*, edited by Steury, D. P., 100. Washington, DC: Center for the Study of Intelligence, 1994. https://www.cia.gov/static/sherman-kent-and-the-board-of-national-estimates-collected-essays.pdf.

Tang, J. "How Do We Know? What Intelligence Analysis Can Learn from the Sociology of Science." *Intelligence and National Security* 32, no. 5 (2017): 663–674. doi:10.1080/02684527.2017.1311473.

Thorburn, L., T. van Gelder, C. Shahan Brugh, S. Mi Kim, M. Kolb, and B. Zulauf "The IC Rating Scale as a Measure of Analytic Rigor." *Social Science Research Network* (July 27 2023). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4519926.

Treverton, G. F. "The Soviet Union was a Puzzle. Al Qaeda is a Mystery. Why We Need to Know the Difference." *Smithsonian Magazine*, June, 2007. https://www.smithsonianmag.com/history/risks-and-riddles-154744750/.

Treverton, G. F., and R. Miles. "Unheeded Warning of War: Why Policymakers Ignored the 1990 Yugoslavia Estimate." Center for the Study of Intelligence. October. 2015. https://www.cia.gov/static/Unheeded-Warning-Yugoslavia.pdf.

University of Buckingham. https://www.buckingham.ac.uk/wp-content/uploads/2021/02/GSIN_3Fc.pdf.

University of Leicester. https://le.ac.uk/courses/intelligence-and-security-ma-dl/2023.

Unsinger, P. C. "Three Intelligence Blunders in Korea." *International Journal of Intelligence & CounterIntelligence* 3, no. 4 (1989): 549–561. doi:10.1080/08850608908435120.

U.S. General Accounting Office. *Management Reform: Implementation of the National Performance Review's Recommendations*. December. 1994. https://www.gao.gov/assets/ocg-95-1.pdf. 172–173.

U.S, General Accounting Office. *Management Reform: Completion Status of Agency Actions Under the National Performance Review*. July 12, 1996. https://apps.dtic.mil/sti/tr/pdf/ADA327604.pdf.

U.S. Intelligence Board. "Quarterly Report of National Intelligence Estimates Used in Support of NSC Policy Papers." July–September 1958," Document serial USIB-D-19.2/1, 25 October 1958, CIA FOIA Reading Room, document number CIA-RDP82M00097R000800010010-2.

U.S. Intelligence Community Staff. *A Guide to the National Intelligence Community's Production Organizations and Their Products*. 1975 October, 1, CIA FOIA Reading Room, document number CIA-RDP83M00171R000800050002-3.

U.S. National Intelligence Council. "Yugoslavia Transformed." NIE 15-90. CIA FOIA Reading Room. October. 1990. https://www.cia.gov/readingroom/docs/1990-10-01.pdf.

U.S. National Security Council. "Preliminary U.S. Policy on Outer Space." Nsc 5814/1. August 18, 1958. https://aerospace.org/sites/default/files/policy_archives/NSC-5814-1%20Space%20Policy%20Aug58.pdf.