20 March 1981

MEMORANDUM FOR:

Deputy Director, PAO

FROM:

PAO

SUBJECT: Background Information on Production and Evaluation
             Elements in NFAC

1. <u>Senior Review Panel (SRP)</u>: The SRP was constituted in December
1978 and presently consists of four members: General Bruce Palmer,
Ambassador William Leonhard, Herbert Rothenberg, and Ambassador Lincoln
Gordon. The SRP has no staff save for [   ] an experienced   STAT
and senior CIA officer who functions (alone) as its Secretariat. The
topics which the SRP addresses may be either self-generated or done at the
request of the DCI or his Deputy for National Foreign Assessment (D/NFA).
The SRP will make known to the D/NFA the analytic and clerical assistance
it will require for any given study, and the appropriate persons will be
seconded to the SRP for the duration of the study. Generally, these
people will come from the National Foreign Assessment Center (NFAC) of
CIA. If, however, assistance is needed from other elements of the
Community (DIA, or State/INR, for example), the D/NFA will request the
assistance of such organizations.

2. <u>National Foreign Assessment Center (NFAC)</u>:

     a. <u>Collection Requirements and Evaluation Staff (CRES)</u>: It
should be clearly understood that this staff evaluates <u>only</u>
collection systems on NFAC's behalf (through its Special Projects
Group) and has no responsibility for evaluating NFAC or other
Community production.

     b. <u>Planning, Management and Evaluations Staff (PMES)</u>: The
staff consists of the Budget and Fiscal Group, an Administrative
Group, and a Production Planning and Review Group. The first two
deal primarily with resource allocation for NFAC, not for the
Community. The last attempts to ensure that individual offices
within NFAC are given the analytic resources they require but has
taken no strong position in guiding and planning production. This is
left to the NFAC senior management and office directors. This group
does not concern itself with Community intelligence production
planning.

SUBJECT: Background Information on Production and Evaluation Elements in NFAC

3. <u>National Intelligence Council (NIC)</u>: Neither the National Intelligence Officers (NIOs) individually, nor considered collectively as the NIC, have any responsibility for evaluating Intelligence Community production. They do, of course, exercise "quality control" over intelligence estimates, IIMs, or other papers produced under their aegis, but have no charter responsibility for, nor have they attempted to initiate any, Community evaluative role. Although, on occasion, individual NIOs have made some attempt to coordinate, albeit very informally, production among Community intelligence producers, such efforts have been spasmodic and ineffective. The NIOs, also, with some exceptions, have done little to fulfill their role of maintaining close links with the policymaking Community in order to ascertain the degree to which their products are responsive to policy requirements.

4. <u>Special Assistant for Community Interests</u>: This function was created in the spring of 1980 and is filled by [   ] He has no   STAT
staff and although he is attempting to coordinate such projects as the Community User Survey through the Interagency Working Group on Intelligence Production, he has not undertaken any real initiatives in coordinating production; the User Survey will be perhaps the most important step in attempting to "evaluate" the responsiveness of Community intelligence production (in its broadest sense) to the needs of policymakers. The Working Group, chaired by [   ] is composed of   STAT
representatives of State, NSA, and DIA with [   ] as the RMS (now IC   STAT
Staff) representative.
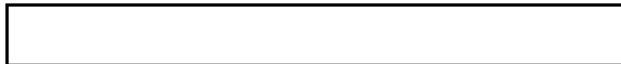
STAT

STAT

cc: [   ]
PAO Subject
PAO Chrono
D/PAO

CONFIDENTIAL

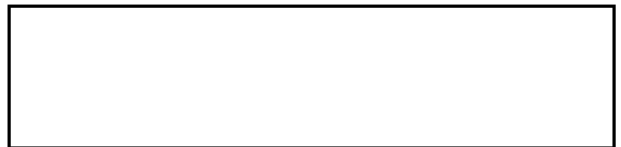Seminar on Evaluation of the Intelligence Product

STAT

Center for the Study of Intelligence
Central Intelligence Agency
Washington, D.C. 20505

21 April 1980

TR 80-10008   STAT

CONFIDENTIAL

CONFIDENTIAL

Criticism should not be querulous and wasting, all knife and root-puller, but guiding, instructive, inspiring, a south wind, not an east wind.

--Emerson, Journals

SEMINAR REPORT: EVALUATING THE INTELLIGENCE PRODUCT

Evaluating finished intelligence is not a new subject for discussion among intelligence managers. Three major reports by intelligence officers and outside groups written within the past three years have characterized the absence of an evaluative mechanism as a serious deficiency in the intelligence structure. The studies disagreed on where the responsibility for this evaluation should be assigned, but there was agreement that some method of evaluation was needed to judge the utility and quality of analysis, to assess the impact of intelligence on consumers, and to provide lessons for intelligence managers to guide future production.

On April 21, 1980, a group of 30 people--intelligence professionals and Congressional staff officers--met under the auspices of the Center for the Study of Intelligence to share their views and to exchange ideas on the evaluation of intelligence products. The agenda included discussion of three major aspects of the issue:

1. Is there a continuing need for a systematic evaluation of finished intelligence products? If so, who should be responsible for it?

2. How should judgments about utility and quality be made? Can the development of an adversary relationship between the intelligence producers and judges be avoided?

3. How can the lessons learned from an evaluation process be translated into guidelines for future production?

A Need to Institutionalize Intelligence Evaluation

In a December 1978 meeting held to discuss the evaluation of human-source intelligence collection, a group of intelligence collection managers concluded that it was imperative to establish a systematic means for evaluating intelligence collection methods and results. This experience suggests that production elements might

CONFIDENTIAL

also have to devise an evaluative mechanism for their products. One seminar participant commented that the "choice" apparently made in the past by producers has been simply to ignore the whole issue. Only recently, he felt, has the Intelligence Community begun seriously to examine the feasibility of such a mechanism.

The necessity for and the desirability of developing some procedure for evaluation was acknowledged by the participants. Questions arose, however, in discussing what should be evaluated and by whom it should be done. To clarify thinking, it was suggested that the issue be visualized in terms of a three-sided figure:

--First, one must consider carefully what is to be evaluated. There is in reality not one but a whole spectrum of "whats" ranging from the analytic process itself to the publication of a piece of finished intelligence. Whether referring to the product or the process, it is important to distinguish the type of intelligence being considered--national or departmental, current or long-range.

--Second, the criteria used to judge the product must be clearly defined. If one is using direct feedback from the consumer to judge product responsiveness to consumer needs, then one must be aware of the cognitive problems associated with the users' perceptions of their requirements. If one is using objective norms of sound analysis, then such norms must be identified, codified and agreed upon by the production judges.

--Third, it must be clear who is to perform the product evaluation, whether that be the agency producing the product, an outside group, the DCI staff or CIA's Senior Review Panel.

While there may be different perspectives associated with each of these three elements, it was felt that a combination of the three would provide a more comprehensive way of thinking about the problems related to evaluation.

One participant felt that it was necessary to review the criticism leveled at the Intelligence Community in regard to product evaluation. He suggested that there are three possible responses to such criticism:

--We can attempt to placate the critics by pointing out that analysis is a creative process and that we are doing as well as can be expected by bureaucratic standards.

--We can take the criticism to heart and strive to develop a mechanism for assessing performance, if only by providing a means of customer feedback.

--We can develop some method for providing the top management
with a data base needed to judge how well it is doing and for
helping it to use the resources of the community more effectively.

There was some debate as to the cost-effectiveness of these
responses. One discussant felt that whatever evaluative system
is developed will use up considerable manpower over time. We
should try, therefore, to satisfy as many of these objectives as
we can simultaneously. Another felt that while these objectives
were not mutually exclusive, the mechanisms needed to deal with
them may be quite different. Nevertheless, against the whole
cost of the community's effort he argued that the cost of additional
staff to provide product evaluation would be trivial. Finally, a
third participant questioned the pertinence of this debate on cost.
We must take a closer look at how manpower would be used before
dealing with the cost problem in more detail.

Internal Versus External Evaluation Mechanism

Having agreed generally that a formal evaluation mechanism
for the intelligence product was needed, the discussion narrowed to
the question: To whom should this responsibility be given? Earlier
studies have usually avoided specific recommendations in this regard.

This prompted a participant from DIA to cite his own experiences
in describing one workable internal evaluative format. He acknow-
ledged his distrust of any outside evaluation of intelligence
estimates, especially when undertaken after a period of time had
elapsed. Such an external evaluating group may not see or be aware
of all the factors or data used at the time the estimate was
written. In his opinion, it would be more useful to have a National
Intelligence Officer bring together the original author(s) and any
others who had taken part in the discussion and final approval of the
draft. They could then go back and see where and why they were right
or wrong as a group and review how the final collective analysis was
made. An internal process such as this would not only be useful and
stimulating but would also provide valuable insights for updating
the existing estimates.

There is at least one precedent for an external as well as an
internal evaluative mechanism. In the mid-1970s the Intelligence
Community Staff (ICS) had as one of its functions the current
evaluation of interagency products, although this later developed
into a postmortem process before falling into disuse. When the
Director, NFAC, insisted in 1977 that the evaluative function
belonged within NFAC, the ICS dropped its efforts to revive the
mechanism. This recollection prompted one discussant to raise the
question of the role and charter of the NFAC Senior Review Panel
to judge intelligence products. One member of the panel present

CONFIDENTIAL

at the seminar commented that it was still largely in an experimental stage; an assessment of its performance and effects of its judgments are not yet complete. He stressed, however, that the Senior Review Panel was not a replacement for the ICS's community-wide evaluation system.

Perhaps the most pertinent question to be considered in any discussion of an internal evaluation mechanism is whether or not the people who had been actively involved in the production of the product are capable of judging it with sufficient objectivity. Several participants thought that producers could indeed be accurate judges. DIA has developed what it believes to be a useful format for product evaluation as an internal process. The DoD policy relating to product evaluation obligates producers to carry out systematic evaluations of their own products. A summarization of the results is used as an input to future management and budget programs. While perhaps less objective than an academic effort performed in the abstract, the DIA system nevertheless offers a means of improving the intelligence product within the bounds of DoD resources.

Like artists, intelligence analysts are highly sensitive to any criticism of their work. Reviewers must respect this while persisting in their efforts to render constructive judgments. The consensus of those present was that the product of an independent review would not be as helpful to intelligence managers as one resulting from an internal evaluation. The ideal solution would be a combination of the two--a mechanism which reflected a knowledge and understanding of the intelligence system in combination with a perceived ability to judge the utility of its product without bias. Some participants felt that a group of experienced insiders-- if they could develop the detached independence associated with an external group--might be accepted as professional evaluators. Thus, we could have our cake and eat it too.

Several participants pointed out that a form of evaluation automatically takes place in many areas of production as supervisors of analysts critique the paper while it is being produced. Some felt that the best evaluation process is one that occurs before publication of the product. Formulating the terms of reference and participating in the review and coordination process should provide an on-going evaluation of the quality of analysis. If one waits for a year or more before doing an evaluation, the intelligence judgments may be seen as incorrect in the light of subsequent information, even though the analysis may have been as accurate as possible when it was written.

There exists a general sense that an evaluative process which is not imposed from above is essentially useless; there must be a

CONFIDENTIAL

4

CONFIDENTIAL

commitment from the top in order to have any evaluation mechanism
function effectively. If the evaluation is performed by an inside
group and accepted by production managers, then its results may
prove to be influential on future intelligence products. If the
evaluation is done by an external group which exerts no real
authority or direction on the production elements, then it will be
easier for intelligence managers to ignore the evaluative
recommendations.

The principal problem is not the potential adversary relation-
ship between intelligence producers and evaluators who judge the
products, but rather involves bureaucratic managerial resistance..
Any outside evaluation procedure will be viewed as an intrusion
on the prerogatives of intelligence managers, who want to keep
control of their product. One participant remarked that the
interviews he conducted on this subject encountered little resis-
tance until he reached the higher levels. One intelligence official
told him in effect, "I don't want my people to read about their
work; I want them to work." Another senior officer in the CIA
was quoted as saying that we ought not to spend so much time
worrying about where we have been, but where we are going.

Consumer Judgment

Assuming that some type of evaluative mechanism can be estab-
lished, to what degree should it rely on consumer reaction? Several
participants felt that policy consumers--the ultimate recipients of
intelligence information--were the ones most able to judge the
usefulness of intelligence products. The people in the best position
to judge the quality and utility of an IIM on naval readiness, for
example, would be the fleet commanders and the program officers and
implementers. Another discussant, however, argued by analogy that
the quality of medical care cannot be judged fairly by the patients
receiving it, nor can consumers judge intelligence products without
falling into self-serving distortions.

Since intelligence is ultimately the servant of the policy maker,
consumers ought to be consulted in any evaluation effort. If one
begins tinkering with an intelligence evaluation procedure, consumers
should be brought into the process in a more relevant way than the
rather ad hoc systems of the past. This is not meant to imply,
however, that consumer reaction should provide the sole evaluative
input.

There are times when excessive reliance on consumer reaction to
the usefulness of intelligence may prove to be misleading. In
interviewing consumers during the process of individual postmortems,

CONFIDENTIAL

CONFIDENTIAL

one intelligence officer uncovered responses such as "I didn't read it" or "I read it but I didn't believe it" when asking about intelligence product utility. Another common reaction to intelligence analysis--probably resulting from a simultaneous reading of common mail--was "I already know that." The merging of intelligence into policy formulation at an early stage also makes consumer responsiveness more difficult to weigh. What some policy makers see may not be intelligence per se, but some "middle" product in which intelligence inputs have been buried.

Judging Utility

One participant sensed a certain amount of "hand-wringing" in this seminar in regard to the assessment of consumer needs and the ability of consumers to articulate those needs, especially in the area of such "soft" measures as quality, timeliness and accuracy. ·From his experiences with consumer surveys and questionnaires, he found that intelligence can and does get useful feedback from its consumers. But surveys must ask specific, detailed questions about particular products--not a vague "How do you like...?" which may elicit such broad responses as "I like it, it's free."

The experience of DIA has also shown that an internal intelligence evaluation mechanism may usefully apply commercial marketing techniques such as bartering or threats of product elimination. For example, consumers were not at all reluctant to articulate their needs when they were informed that only three of five particular services would remain operable and were asked to list their preferences. Although DIA has gone far in the development and use of such management techniques, others in the community are not aware of these efforts.

There was some discussion on whether it was possible to make judgments on intelligence without looking at the larger policy framework. Some participants thought that it was possible to get adequate feedback on whether intelligence was used, how it was used and under what circumstances it was used. Others thought that such replies tended to be too simplistic. Consumer response to the question "Did you use intelligence" has little to do with the quality of the intelligence product, for there are many reasons why they may not have used it. How the utility of intelligence is judged depends strongly on which consumers are being asked. This is a problem of consumer judgment--one must be aware of the dangers in asking the wrong consumer whether or not a particular report was worthwhile. There must be some effort at a more systematic surveying of reactions across the entire known community of consumers.

CONFIDENTIAL

6

CONFIDENTIAL

Much time and effort is expended in asking consumers about the utility of the intelligence product. But what is really meant by "useful," and useful to whom? How can product utility be effectively measured in terms of its degree of influence on policy action or its relative impact on the perceptions of policy makers? Is an evaluation exercise designed only to find out how well we are or are not doing?

It was suggested that there are in fact two important aspects of this problem which must be clarified. First, the quality of intelligence and its relevance or utility are not uniquely related to resources, although allocation of resources is important to quality. One can make an effective argument that a reduction in resources could make an overall improvement in the intelligence product. Increasing available resources also increases the tendency to depend on someone else to do the necessary work. Working groups whose relationship to themselves and the issues is uncertain will result in products of limited quality. The resource-quality relationship must be addressed as a separate issue.

Second, one must be more specific when examining the consumers of interagency products. Members of CIA represent only one agency opinion, and trying to evaluate interagency products from an exclusively NFAC perspective results in a narrow, limited view of the problem and its solutions. The objectives and principal concerns of DIA, the State Department and the individual service branches may be quite different. These underlying interagency differences must be taken into account in evaluating the utility of the finished intelligence product.

Evaluating Quality

On the issue of quality improvement, several questions were raised. How do we know when we have succeeded in improving the intelligence product? What does "better" intelligence mean? Is there some intrinsic attribute divorced from the question of how well the consumer likes the product which can be used as a guideline? How should an evaluative process make quality judgments about intelligence?

At a recent symposium, Professor Richard Pipes of Harvard remarked that he would have given one intelligence estimate an "F" if it had been turned in to him by one of his students because of its lack of "academic rigor." Is the equating of intelligence products with academic research, however, really a fair way to judge intelligence? Perhaps the difficulty lies in part in the fact that quality and utility judgments at times may be at variance with each other.

CONFIDENTIAL

7

CONFIDENTIAL

Presently, there exists no generally accepted objective criteria to judge intelligence. Subjective judgments of quality such as "good" or "bad" will inevitably be contested by someone. If we can define approximate standards for questioning the utility of an intelligence product, we can also set questions concerning quality:

--What are the assumptions being made?

--Is the analysis reasonable? Rigorous?

--Is the product timely and well written?

The process of evaluating "quality" is itself rather subjective in nature, analogous to judging a wine-tasting contest or a figure skating event. For evaluating intelligence products, the lack of objective guidelines is a hindrance to any group judgment. This need not, however, invalidate the notion of developing a systematized judgment mechanism. Evaluations made under such circumstances are a common kind of experience. In the academic community, for example, the evaluation of prospective students and of student papers does not always follow objective criteria, but involves personal judgments. Yet there is no wide margin of disagreement about what constitutes a satisfactory university paper.

There was some discussion on the applicability of a right/ wrong dichotomy to act as a criterion for judging intelligence products. One discussant felt that most consumers had built-in appraisal or evaluation systems which work faster than any which could be set up. For example, a person who uses intelligence will let the producer know if he thinks the analysis is wrong. The problem with this is that even if an internal evaluative unit is set up and evaluates a product as "good", a user may shake his head in resignation and say the evaluation is as "wrong" as the product.

Other participants felt that "wrongness" was not a proper standard for judgment. Can situations ever be depicted in such black-and-white terms? For example, in 1973, the Intelligence Community said that the Arabs would definitely not attack Israel. While the community judgment was wrong, the more significant question should be why was it wrong. Such issues should not be dealt with simplistically. One can be right for the wrong reasons and, particularly when deception is involved, draw a sound but incorrect conclusion from the data available. A right/wrong dichotomy, it was judged, is an inadequate framework for evaluating intelligence components and their methods and products.

CONFIDENTIAL

CONFIDENTIAL

One participant concluded the discussion on quality judgments by suggesting that what was needed was a reasonable set of attributes against which each product could be judged. Evaluators could then be more confident of the results and intelligence managers more comfortable with using this feedback. An acceptable evaluation process may have one or all of the following benefits:

--justify budget and resource allocations;

--stimulate closer communication between intelligence producer and consumer;

--improve product quality.

Postmortem Judgments

Although most of those present at the seminar felt that a serious new effort at evaluating intelligence products should be made, many believed that any process considered should be selective. If we should try to evaluate production from a day-to-day perspective, we will become bogged down by sheer volume. It was generally agreed that any evaluation process should be particular in scope, dealing with specific issues or topics concerning Intelligence Community services or failings. One must then be careful, however, not to generalize too much from these specific cases.

For all the objectives about the unfairness implicit in a postmortem's perfect hindsight, the passage of time does bring useful new perspectives. Postmortems can produce fair judgments when intelligence does everything right and still comes up with the wrong answer; they can also show when intelligence has "lucked out" for all the wrong reasons. There are lessons to be learned in either case.

While the postmortem function of the Intelligence Community Staff was potentially beneficial when the lessons learned were internalized by the production elements, the process tended to stimulate defensive attitudes rather than critical or creative thoughts. Some seminar participants were bothered by the automatic assumption of postmortem writers that something went wrong with the intelligence system.

Another criticism was the lack of a postmortem process to address intelligence successes. It would seem reasonable to think that there are lessons to be learned from successful intelligence estimates.

CONFIDENTIAL

9

CONFIDENTIAL

Evaluation of Collection

Since intelligence collectors have already established a methodology for evaluating their services, the seminar discussion turned its attention to the experiences of collectors in order to see if any parallels could be drawn for production. According to one intelligence collector present, there is a continuum along which one must operate. At one extreme is quality measurement which involves looking at the product against certain accepted criteria. At the other extreme is a measurement of cost effectiveness which involves relating results to initial objectives and actual resource input. The collection evaluation process is trying to address itself to a wide range across this spectrum. Now the collection community is trying to apply the lessons it has learned from this evaluation process, both in terms of substance as well as collection management.

In the opinion of one collector, any evaluation system ought to be broad enough to touch both extremes of this continuum. Another discussant, playing the role of a devil's advocate, saw potential problems in the system and observed that some collectors may be reluctant to report interesting but unevaluatable information from the field out of fear that the system will give them a "poor grade."

A Critique and a Conclusion

One participant summed up his opinions on the feasibility of an evaluative mechanism and the application of its findings as follows:

First, he sensed that some of the other seminar members were quite negative about using consumer reactions to intelligence products as a basis for judging quality or utility. He disagreed that consumers lacked the ability to articulate these concerns. Where else would intelligence find out about the effectiveness of its products if not from its users?

Second, people around the community have said that establishing an evaluative system would be too demanding of resources. In his view, less than 1% of existing resources would be required. Is this too much to ask for quality control? If the problem is too tough to address all at once, we ought to work at it incrementally. It seemed to him that not much was going on as far as production evaluation was concerned. This made him wonder if intelligence managers really wanted to establish an evaluation system.

Nevertheless, the general feeling among the participants was that there is a whole series of things we can do to make judgments concerning intelligence products. We can look at the nature of

CONFIDENTIAL

CONFIDENTIAL

the product by itself, the rigor of its analysis, its assumptions and resulting statements; we can look at the atmosphere or environment in which it was produced; we can examine consumer reaction to the final product. It was also generally agreed that looking at the stream of production rather than its individual products is more likely to prove useful. The major unresolved issue remains: who should perform the evaluation of the product?

To establish an evaluation system, the process must be started with people who are in a position to apply the lessons learned. Any evaluative mechanism must be designed so that analysts and managers can see something beneficial in it for them. Commitment to evaluation from the top levels is important, but simply imposing it from above will not work. Perhaps a good starting point would be to survey how intelligence managers look at product evaluation.

In the end, most managers are in the same position. If they cannot demonstrate that they are producing something of value, they will lose in the competition for production resources. Managers must show that they are spending their resources effectively. Results from an accepted evaluation program would provide inputs for budget considerations.

CONFIDENTIAL

11