THE RESEARCH PERSPECTIVES PANEL

I. INTRODUCTION: EVALUATING THE EVALUTORS

JAMES G. ABERT, Vice President for
Research and Development,
National Center for Resource Recovery

MR. GRANDY:

At this point in our program, we are starting the second phase which will continue until tomorrow morning. This is the discussion of current research experience from the perspectives of the researchers.

We are pleased to welcome Dr. James Abert who will given an introduction to this part of our program. Jim is currently Vice President for Research and Development at the National Center for Resource Recovery. He is a Mechanical Engineer with a doctorate in Economics. He was Deputy Assistant Secretary for Evaluation and Program Monitoring at HEW for two years, between 1969 and 1971.

In between the time he left HEW and today, he has studied and published widely on a variety of topics. In addition to his Government service and his industrial work, he has been involved in a number of policy study committees for the National Science Foundation, for the National Academy of Engineering and for the National Academy of Sciences.

It's a pleasure to welcome him. He is going to talk on evaluating evaluators; hopefully, I think, in terms of Jim Stockdill's warning, only one level deep.

MR. ABERT:

Thank you very much. Some of you may know that the National Center for Resource Recovery is concerned with refuse recycling. I

have been told on several occasions (more often, the further I get from Washington, D. C.), that a couple of years of HEW is really good preparation for a career in garbage.

Both in Government and in the research community, it should be clear by now that the term evaluation lacks precise definition. Among producers and users, there is wide variation as to what constitutes evaluation and how it differs, if it does, from (among others) field experimentation, demonstration-research projects and, to choose another term, action-research programs.

Exact definition, however, is probably of little importance. Regardless of the exact meaning of the term, it appears that evaluation has become somewhat of a fad, if not yet an entirely proven, integral part of the management process.

I think it is important to ask at this time if the resources devoted to evaluation are a valued activity in the constant search to improve the efficiency with which public sector funds are spent. This is not to suggest that a definitive answer to the question can be given.

I would say that throughout the Government, the foundations for evaluation laid some years ago have grown into a full-fledged evaluation emphasis. I have chosen the word emphasis carefully, and it is to stress that the development and institutionalization of an evaluation program is an evolutionary process. It is not done overnight. Indeed, it is not done in a year or two. How long depends on the interest and determination of those responsible for its direction and the support given to its growth. To graft it to a hostile bureaucracy requires both toughness and tender loving care. It does not "take" easily.

The stakes are high, not only because of the employment generated within the Government but its effect on that segment of the private sector which responds to the RFP's to do the evaluations. I have heard evaluation characterized as Federal aid to contractors, and to some extent that is true.

However, the real stakes are where the big bucks are. Evaluation can become, perhaps inevitably is, a political device which can be used to promote support for an advocate's program or reduce enthusiasm for an opponent's proposals.

Evaluation is important in other areas as well. It provides the financial incentive for academicians to train their intellectual ordnance on the target of improving the management of public funds. Some may argue that they often fail to find the target. Perhaps they fire with biased sights, or perhaps the target itself is poorly understood by those in the user community whose articulation of what mark was to be hit is often only clear to them after the fact of the evaluation.

Finally, evaluation provides the wherewithal to expand the general knowledge base in areas where the more traditional data collection services have not ventured. At the least, the social sciences should have seen and should continue to see more dissertations in what might be called the "grand design."

Putting this aside as a spillover benefit, and presumably it is a benefit, the basic question concerning the valuation of evaluation is "Do evaluation outlays produce greater efficiency in program output than the costs of the evaluation efforts?"

In general, short of saturation, more information is better than less. Yet there are costs involved. Are the likely improvements in program targeting and management worth the cost of data collection and analysis? Evaluation can cost more than it is likely to save, although the definition of "save" is a problem here.

There is a more subjective side as well, indeed even emotional--so emotional that evaluation can become counterproductive. This is particularly true when one begins to evaluate in earnest, where only lip service has been paid to this function in the past.

As you know, the setting of program objectives and the choosing of evaluation criteria are in themselves a very emotional undertaking. Program managers generally are not anxious to do it. In fact, trust, confidence, honor and many of the more noble aspects of life seem to be strongly challenged by evaluation.

The tools for estimating the worth of policy-related information are primitive at best. Much of the information obtained simply helps the program manager to understand his program better. To relate this information in some casual way to program improvement and then to further measure the value of this improvement appears to be beyond today's practice.

Partially for these reasons, the usual chronicles of evaluations accomplishments--a successful study or two offered as evidence of the achievements of the evaluation program--often leave the listener with a feeling of "Well, maybe the expenditures on evaluation have been worthwhile; and again, maybe they haven't."

Has progress been made? At the outset, it seemed that many felt an evaluation program should appear fullblown. Of course, this has not happened. The formulation and implementation of a viable program is a step-by-step process. One builds on what one has accomplished in the prior period. One does not grasp for options that have low probability of being achieved. The need to set reasonable sights and to plan for evolutionary growth with many mid-course changes does not seem to have been appreciated fully either at the outset, or now. Also, it is usually not present when observers of evaluation programs, no matter how objective they may claim they are, attempt to evaluate evaluation.

There is still much to be done. The key to the future growth and acceptance of evaluation is the development of recognized approaches to the conduct of evaluation, in particular to establishing the reliability of the judgments made by field staff.

Of course, it is necessary initially to obtain and to maintain high-level support. Because evaluation's image is that of uncovering or demonstrating the negative, it is generally only grudgingly and reluctantly accepted by those on the receiving end. While the degree of support of the evaluation activity can be reflected in a variety of ways, the position of the evaluation office in the organizational structure will be a principal indicator. Clearly, if such units are directly linked to principal decision- and policy-makers, the possibilities of influence will be noted throughout the organization.

Along the same line, evaluation, in my view, should be legislatively mandated and treated as a program in its own right including a mandated budget.

In addition, thought must be given to evaluation in the structuring of operating programs such that more of them can, in fact, be evaluated.

Finally, program evaluation must hew to a nominative approach that forces judgments as to the accomplishments, or lack of accomplishments if such is the case, of the program being evaluated. In general, the research paper suitable for publication and useful for promotion does not fit the bill.

Time will bring with it a greater appreciation for the real-world context of evaluation. This must be so, or the evaluation parallel will be that of the formal discipline-focused research program, lodged far down in the agency, far from the policy arena.

Looking to another facet of the evaluation picture, it is too soon to tell if the political process has been sufficiently sensitized to allow evaluation to continue with its evolutionary growth. There are still many hurdles to be overcome, not the least of which is institutionalizing evaluation requirements, procedures and dissemination to the extent that past lessons are not relearned by each succeeding change in department and agency management.

Only time will tell whether evaluation lives up to the reasonable expectations of its advocates or turns out to be a relatively short-lived but expensive experiment. Thank you very much.

MR. GRANDY:

Thank you, Jim. In your remarks I think there are some reinforcement of comments of some other speakers this morning. Also

something of a challenge to this audience to help facilitate the institutionalization process so that evaluation becomes solidly enmeshed in the fabric of public program management.

II. STUDENT ATTRITION AT THE FIVE FEDERAL SERVICE ACADEMIES: AN IN-DEPTH AUDIT

CHARLES W. THOMPSON, Assistant Director,
Federal Personnel and Compensation Division, and

JOHN K. HARPER, Acting Director,
Systems Analysis Group
Federal Personnel and Compensation Division
General Accounting Office

MR. GRANDY:

Next on our program will be the presentation of our first research paper. This will be done by two gentlemen from the General Accounting Office, Charles Thompson who has long been a staff member at GAO, and his colleague, John Harper. Both are in the Federal Personnel and Compensation Division. Their report, as you see from our agenda, deals with Student Attrition at the Service Academies. The document is available, displayed with other literature out in our anteroom.

I think Mr. Thompson is going to speak first, and will then turn the discussion over to Mr. Harper.

MR. THOMPSON:

I'd like to spend a few minutes discussing the problem of attrition that we faced, our general approach to addressing the problem, a few of the more significant findings, our recommendations and some of my perceptions as to the factors which may have influenced their utilization.

The military academies exist primarily for one purpose--to develop career military officers.

Even though the academies account for only about 10 percent of the initial grade officers acquired by the military services, academy graduates are nonetheless considered among the more highly desirable officers.

To the extent that large numbers of students who would make good career military officers leave the academies before graduation, the effectiveness of the academies' program becomes questionable.

In recent years, attrition at the academies has been high, and it has been increasing, and these increasing trends, particularly at the Air Force Academy, prompted Senators Birch Bayh and William Proxmire, as well as other members of Congress, to request a GAO study of the problem.

Figure 2, below, will give you a better sense of what their concern was.

For four of the five academy classes which graduated in either 1974 or 1975, attrition reached near-term record levels. For example:

- The Air Force Academy graduating class of 1975 had a 46 percent attrition rate, the highest in its history;
- The Military Academy reached an 11-year high of 40 percent attrition;
- The Naval Academy, a 12-year high of 39 percent attrition;
- The Merchant Marine Academy, an 11-year high of 48 percent attrition; and
- The Coast Guard Academy had 46 percent attrition.

In light of these statistics, there were serious questions being raised as to whether the academies were adequately accomplishing their mission. When we add the additional consideration of costs--over $100,000 per graduate--the issue becomes not only one of program

FEDERAL SERVICE ACADEMIES
ATTRITION RATE BY GRADUATING CLASS YEAR

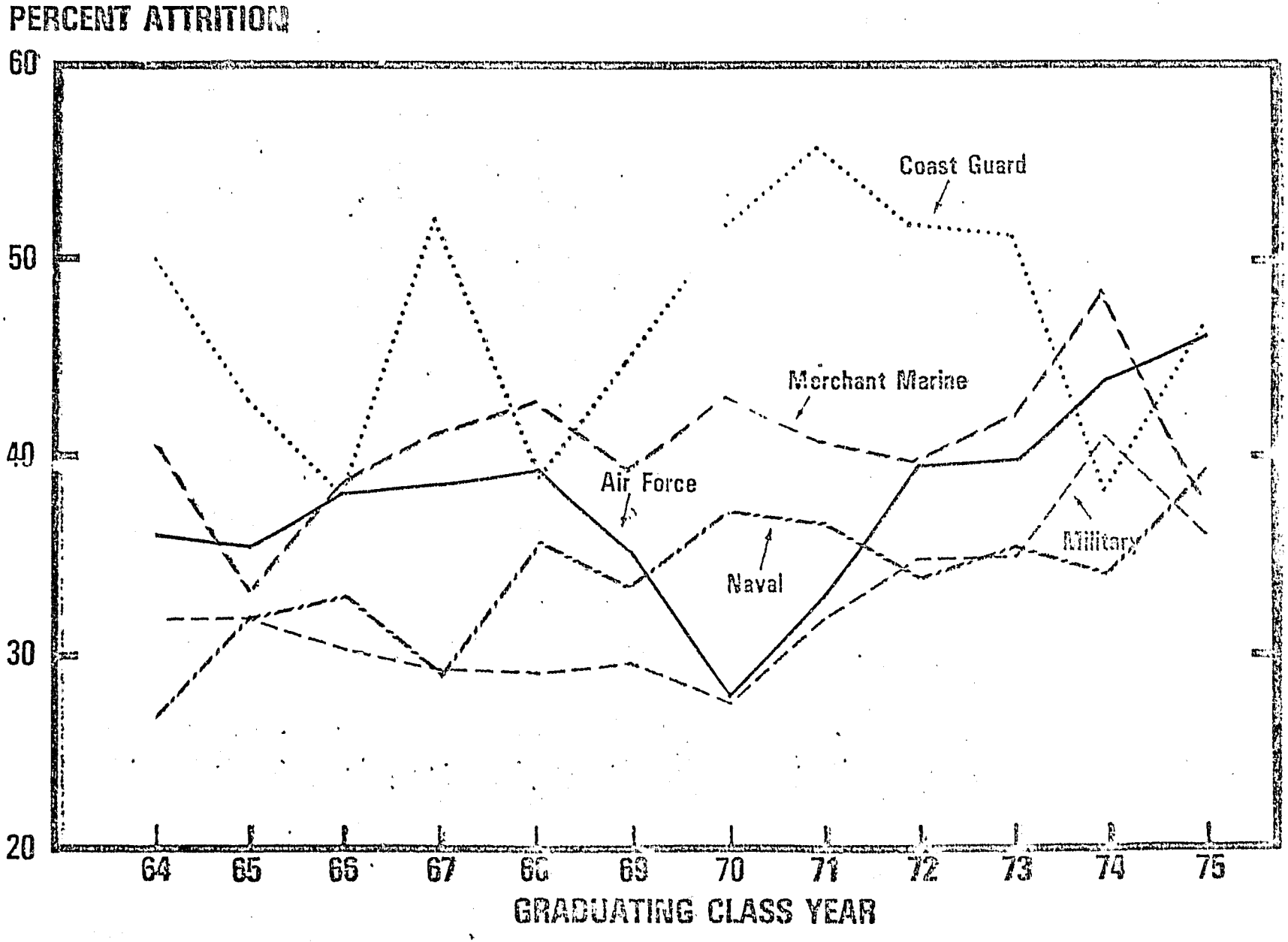

FIGURE 2

effectiveness but also one of program affordability. If attrition could be reduced, the academies could provide more graduates at a more affordable cost. Without reduced attrition, alternative sources of officer acquisition become more plausible and more attractive.

With this in mind, our study objectives became two-fold--first identifying those factors which contribute to the high attrition rates, and, second, proposing program alternatives which we believed would permit the academies to reduce their attrition rates without degrading the quality of the graduates.

We recognized at the outset that some attrition is inevitable and desirable since selection of only those who would make good career officers is unrealistic. Attrition, therefore, serves as a desirable screening device for those students who do not measure up to the standards considered essential to the military profession. Yet, our data suggests that, in addition to weeding out those whom the academies felt were undesirable, they were also losing many potentially good career officers. In fact, one academy superintendent estimated that 20 percent of voluntary dropouts were potentially good career officers.

We felt, therefore, that if we could identify those major factors contributing to the student attrition and recommend changes to them without decreasing the quality of the output, we would be making a contribution to improving the effectiveness of the academies' program at a more affordable cost.

Let me very briefly review for you our approach to the attrition issue, for it is the acceptance of this approach and the steps that were taken to increase acceptance which determined, at least in part, the acceptance of our results and the extent of implementation of our recommendations.

Figure 3 below, shows a rather simplified version of the model we adopted to identify the factors contributing to attrition.

Conceptually, we viewed attrition as resulting from the interaction of three distinct influences: (1) the characteristics that students bring with them to the academy, such as abilities, commitment and expectations, (2) the effect of the academy environment on the students, such as the quality of the academic and military programs, and (3) the external environment which affects students while they are at the academy, such as national economic conditions in general.

Through a rather extensive review of the existing research on attrition, as well as through discussions with current and former academy officials and students, we identified those factors within each of these three areas that could potentially contribute to attrition. These factors, then, formed the basis for our data collection efforts.

Our primary data collection source was a questionnaire we developed and administered to over 20,000 current and former academy students. In addition, we obtained extensive data from academy records and from an annual survey of incoming academy students administered by the American Council on Education. In total we collected or obtained over 500 specific items of information on each student which we hypothesized were related to attrition.

Because of apparent differences in the academies' environments and in the students who go there, we decided to perform separate analyses of each academy.

Further, within each academy, separate analyses were made for each of the three timeframes--the first summer preceding the fourth

ATTRITION MODEL

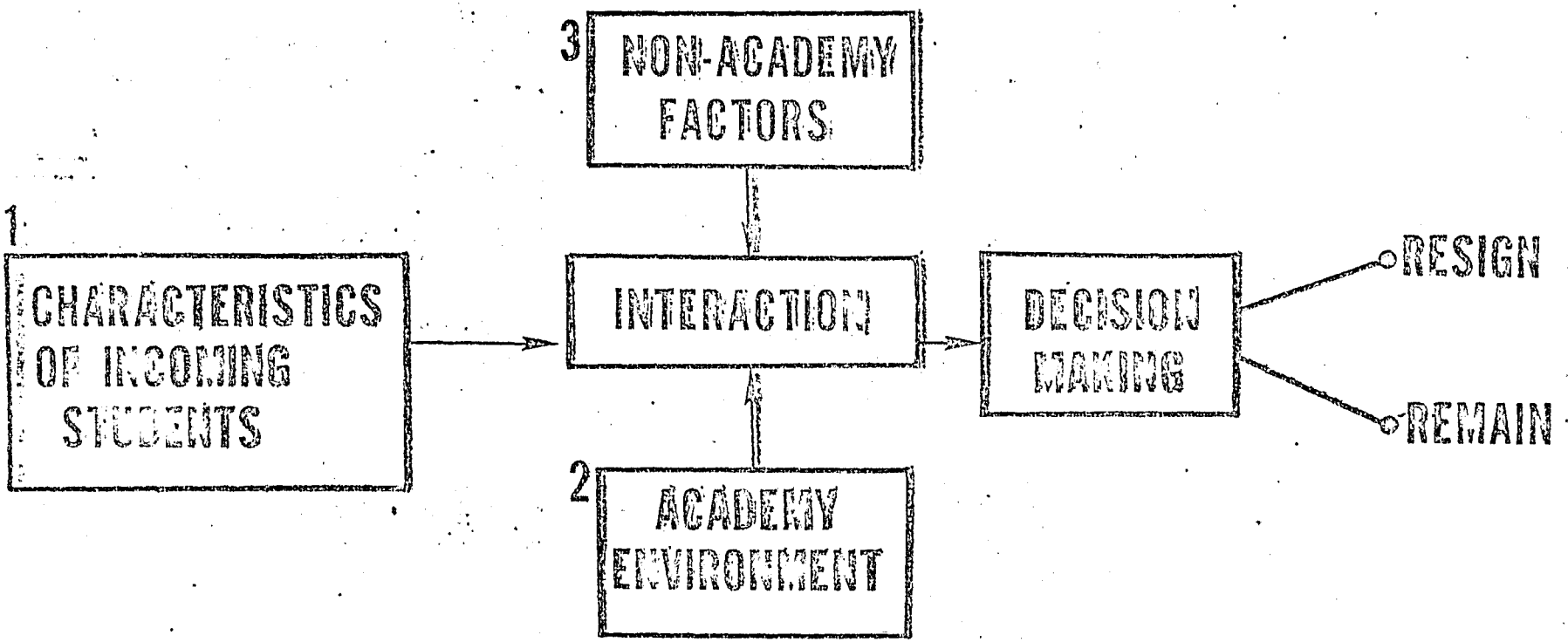

FIGURE 3

class academic year (chis is normally the first two months that a student is at the academy); the fourth class or freshman year; and the third class or sophomore year. These three timeframes were chosen because about 85 percent of all attrition occurs during these first two years and because our prior research indicated there to be different reasons for attrition depending on the timeframe.

Let me now briefly discuss some of our findings to give you a perspective of what they were like and the recommendations we made from them.

In general our recommendations tended to fall into three categories: First, major changes to academy policies, practices or tradition. These tend to be rather high-risk changes, in that if they were implemented and later proved to be wrong, they could have a major detrimental impact on the academies' mission. Second, relatively minor changes to policies or practices. They tend to be low-risk changes. And third, recommendations for further research or redirection of research.

Let me illustrate by discussing a few specific findings: we found that one of the most important factors related to attrition during the students' first few months at the academy is their initial level of commitment at the time they entered. Those students who have lower levels of commitment have significantly greater probability of dropping out.

Our measure of student commitment was made up of a number of questions which the students answered when they entered the academy. These concerned the chances they would transfer to another college before graduating, drop out of college temporarily or permanently,

change their career choice, or get married while in college. Each of these actions almost always requires the student to leave the academy.

Those who dropped out saw their chances of doing each of these things to be significantly greater than those who stayed. Figure 4 below illustrates this point. It shows the responses of first summer dropouts and current students about the chances they would transfer to another college before graduation.

At the Air Force, Military, and Coast Guard Academies, 35, 31 and 46 percent respectively of first summer dropouts stated at the time they entered that there was a "Very Good Chance" they would transfer to another college. Whereas only 2, 4 and 6 percent respectively of current students made this response.

This initial level of commitment is extremely important. There have been leaders at the academies who adopted a philosophy that if a student doesn't want to be at the academy, then the academy doesn't want him. And their programs, especially during the first summer, were designed to test, and I want to emphasize the word test, a student's commitment. However, our study suggests that this philosophy may have driven some good students out.

It's my view that the academies failed to adequately recognize that low commitment is typical of individuals at this age. For example, the next figure gives an indication of this low commitment as it relates to the academies (see Figure 5 below).

For the total class which entered in, for example, 1973--this would be the far right bar on each chart--between 43 and 58 percent of students stated that there was some, or a very good, chance they

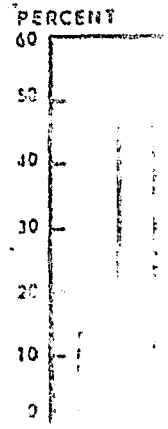

FIGURE 4

PERCENT OF ENTERING STUDENTS WHO BELIEVED THERE WAS "SOME" OR "A VERY GOOD" CHANCE THEY WOULD CHANGE THEIR CAREER CHOICE

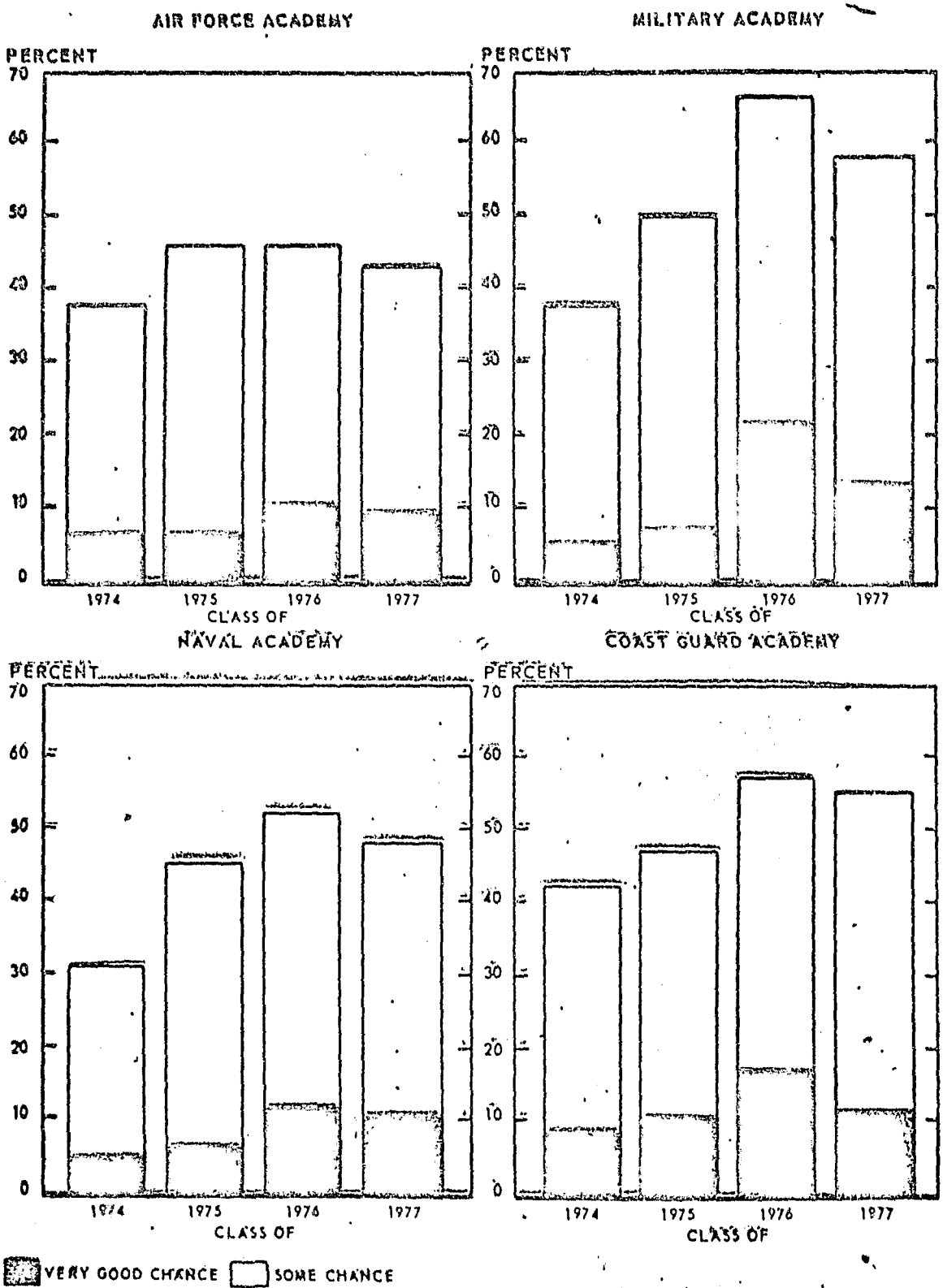

FIGURE 5

would change their career choice. The point I am trying to make is that it is better for the academies to make the assumption that students are not highly committed and design their programs to develop commitment, rather than merely test it.

Generally, the academies agreed with our finding on the importance of commitment to retention; and at least the Air Force Academy instituted an intensive reexamination of their first summer program in an effort to make it more commitment-developing rather than commitment-testing.

However, the responses from all academies were very mixed on the extent to which they agreed to institute specific changes to practices which appeared to be more commitment-testing than commitment-developing. For example, we found that the requirement to memorize and recite trivia, such as sports scores and titles of movies, the heavy emphasis on drills and ceremonies, and the heavy emphasis on creating stress were directly related to attrition.

The more minor of these changes were readily accepted. For example, the Military Academy reduced the level of drills and ceremonies by 35 percent, with further reductions planned. The need to reduce the memorization and recitation of trivia was also generally accepted.

On the other hand, a more major change, that is, the need to review and possibly modify the extent of stress in the environment, was not accepted. In fact it was strongly defended as necessary.

I'm not suggesting that the academies should have accepted all of our findings and made immediate changes. The point I'm trying to make is that the degree to which a finding is accepted and acted

upon is, to some degree, a function of the potential risk-level of the change. And the greater the risk, the more the decision-maker will require additional supportive data before a change is made, particularly if the change is contrary to his predisposition.

Therefore, while the results of some of our findings were not immediately acted upon, they did provide an additional source of information, which, when combined with other supportive studies to follow, will hopefully result in a critical mass and cause a change.

We should not always be disappointed when high-risk type recommendations are not acted upon. We don't necessarily have to live with the consequences.

In closing, let me summarize what I perceive to be some of the factors which influenced the use of our evaluation results.

First, use is, at least in part, a function of the extent to which the decision-maker has confidence that the results are valid; and this again to some degree is a function of the soundness of the approach and the clear, understandable link between the approach, the results and the conclusions and recommendations. We can increase acceptability and use by involving the decision-maker, or subordinates whose opinion he respects, in the process from beginning to end. Recommendations for change, particularly major change, should not come as a surprise at the end.

Second, if we can involve other outsiders of the group doing the study in the study process--again ones whom the decision-maker respects--we provide an important secondary group to whom the decision-maker can look for confirmation of the conclusions and recommendations.

Finally, don't expect that all recommendations will be acted upon. The higher the risk, the less chance that change will take place from the results of one study, no matter how sound. Also, the greater the decision risk, the greater the need to bring the decision-maker and others along with the study.

It is my personal view that researchers or program evaluators can and need to have more interaction and communication with the ultimate decision-maker. If we are to maximize the chances of results implementation, we need to build a greater sense of trust between the decision-maker and the evaluator--trust in his methodology, trust in the validity of his conclusions and the soundness of his recommendations, and perhaps, most important, trust in the evaluator himself. Thank you.

I'd like to turn the discussion over to John Harper, who will further discuss some of our findings and some recommendations.

MR. HARPER:

I will talk about two factors which seem to have affected the extent to which findings from our study could have been and, indeed, actually were used as a basis for policymaking. Let me stress that this is a personal view. Others might well have seen different factors as crucial in determining the extent of utilization.

The first factor was the context in which the study was done. I would like to talk about that context in terms of power relations among the principal actors in the study (see Figure 6 below). I want to do that because it's my feeling that those relations: (1) made the study possible, (2) partially determined how the study was done, and (3) affected the extent to which it was utilized.

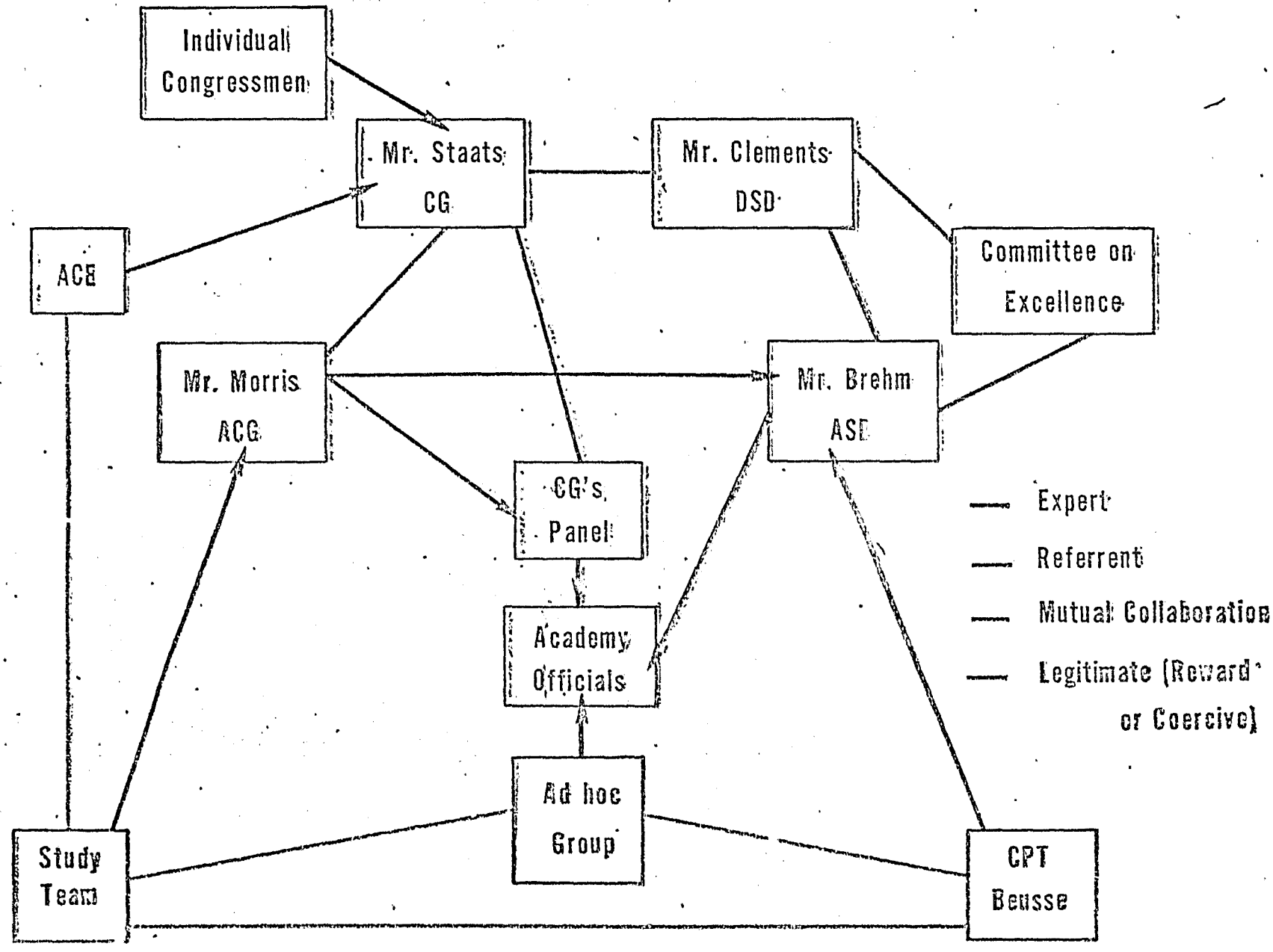
POWER BASES
Individual Congressmen
Mr. Staats CG
Mr. Clements DSD
ACE
Committee on Excellence
Mr. Morris ACG
Mr. Brehm ASD
CG's Panel
Academy Officials
Ad hoc Group
Study Team
CPT Beusse
Expert
Referrent
Mutual Collaboration
Legitimate (Reward or Coercive)

FIGURE 6

There are at least four types of influence, or power, one person can have over another. The first type is expert power which exists when the division of labor in an organization produces groups or individuals with specialized knowledge or expertise necessary to accomplish the organization's primary mission.

The second type is referrent power and is based on the extent to which one person identifies with, or is attracted to, another person because the other person behaves or believes like the influenced person.

The third type of power flows from formal organizational relationships which permits someone to dispense sanctions and rewards based on shared norms.

The last type of power is mutual or collaborative power where the direction of influence alternates between actors.

With these brief definitions in mind let me give a personal view of the principal actors involved in our study and the types of influence they exerted over one another. These remarks are limited to the military academies (Army, Navy and Air Force) because they were the opinion leaders for the other academies in this study.

As Chuck mentioned earlier, the study initially requested by several members of Congress was of factors related to attrition at the Air Force Academy. Attrition had risen dramatically there, and the Superintendent had made a number of hard statements about the institution's lack of concern over it. Several of those who requested the study were perceived by some in the Department of Defense as holding unfavorable attitudes toward the military.

For various reasons, Mr. Staats, the Comptroller General who has headed GAO since 1966 and who was Deputy Director of BOB for many years before that, decided that the study should not have a limited focus but should extend to all of the service academies.

About the time this decision was made, Mr. Staats was in touch with the then Secretary of Defense, Mr. Schlesinger, and Mr. Clements, the Deputy Secretary of Defense. Both of these individuals had been concerned with civilian oversight of DOD's education programs. This concern led to creation of the Committee on Excellence in Education, composed of Mr. Clements and Mr. Brehm, the Assistant Secretary for Manpower, as well as the Service Secretaries. The academies were a principal item on the Committee's agenda.

Mr. Staats was also in touch with a number of members of Congress who expressed reservations about the benefits to be gained from this study. Senior officials at several academies also expressed reservations about the study.

Recognizing the sensitivity of the issue being addressed, Mr. Staats and Mr. Morris, the Assistant Comptroller General, decided to bring together an outside panel to consult with GAO on the study. Mr. Clements suggested a number of former, high-ranking academy officials as candidates for the panel. To add balance, Mr. Staats solicited names of civilian academic administrators from the President of the American Council on Education. The panel which was established consisted of Chancellors of the Universities of Texas, Illinois, and Pittsburgh, and the President of Tuskegee Institute; Vice Presidents of Harvard, MIT, Stanford, Michigan, and Tulane; and former

Superintendents of each of the academies. The other members of the panel--there were 17 in total--were no less important or illustrious.

A number of the civilian administrators had held high-level positions at the academies. For instance, the Chancellor of the University of Pittsburgh had been Chairman of the Social Science Department at West Point and Chairman of the Economics Department at the Air Force Academy.

We met formally with this panel on five occasions over a two-year period and met informally with individual members several times during that period.

Prior to the first panel meeting, the project team presented a proposal for this study to Mr. Morris, who had served as an Assistant Secretary of Defense on two separate occasions. Mr. Morris liked what he heard and communicated that feeling to Mr. Brehm.

The proposal was also enthusiastically received by key panel members at that first meeting. The panel gave the study a certain kind of legitimacy. It also forced us, as researchers, to keep our feet on the ground and it served as a vital communication link to senior academy officials. Meetings of the panel were held at each of the military academies and their senior officials participated in the meetings. This opportunity for them to express reservations about the study to such an illustrious group and to have those reservations moderated--when combined with the informal conversations which occurred between senior academy officials and some of the panel members, I believe, helped to overcome the resistance mentioned earlier.

Mr. Staats' and Mr. Morris' interactions with senior DOD officials added another kind of legitimacy which became important in overcoming particularly strong resistance by academy officials at certain other points of the study.

At the more mundane level, the study team was strongly influenced by the work of the Office of Research at the American Council on Education. The design of the study imitated the "input-output" model which had characterized ACE's research on college impact since 1968.

The research team was fortunate to have as a liaison in DOD a Captain with academic experience in organizational behavior and work experience in survey research. His expert influence was helpful at upper levels in DOD, and his collaboration was helpful with an ad hoc group we had formed to provide us with technical assistance. This group was variously composed of mathematicians, psychologists, and management scientists from the academies; computer scientists and researchers from the military personnel labs; and manpower program managers from the service headquarters.

We met formally a number of times with this group and informally with some of its members. The circumstances surrounding our second meeting give some example of the types of influence at work on this level of the project.

Prior to that meeting, the GAO study team had developed a pool of questionnaire items, and had discussed the study design and hypotheses behind each item with its own field teams. Those field teams had returned to the academies with the questions typed--and not very neatly--one to a page, to discuss them with senior academy officials and ad hoc group members. The teams had been instructed to emphasize

that the questions made up a first draft item pool, and that we were primarily concerned with whether hypothesized causes of attrition had been adequately sampled.

The scales for some of the questions were not balanced, and a number of items were clearly biased against the academies. During the time the field teams were discussing the item pool, we corrected many of these deficiencies. But we made a mistake. We typed the corrected questions one after another in survey questionnaire format. We added response boxes which had not been there before for each question and, in short, developed a fairly professional-looking questionnaire--even for a draft.

The ad hoc group began arriving at the Pentagon from all over the country at 8:00 A.M. on a Monday morning for what was to be a one-day strategy session among themselves before meeting with the GAO study team. Many of them had been given strong marching orders when they left the academies. Well, when they were given the new draft of the questionnaire with its completed-looking appearance, it met with strong resistance. By 9:30 that morning, the meeting with the GAO team had been cancelled; and talk was that the ad hoc group had been dissolved, the academies would not let their students participate in the study, and any study we might be able to do would not be considered legitimate by the academies.

Needless to say, there was a great deal of sideways and upward communication. Chuck and I communicated with Mr. Morris. Captain Buesse communicated with Mr. Brehm. Mr. Morris and Mr. Brehm communicated. And Captain Buesse and I spent several days discussing hypotheses and response scales.

I don't know whether it was reason, or power, or something else, which prevailed; but the ad hoc group did meet about two weeks later. In a hectic five-day meeting, the study team established its professional credibility and convinced the program managers that we weren't trying to give the academies a bad name.

As time went on, the ad hoc group supported the study more and more to academy officials, although they were never blind to its technical problems.

I have spent a good deal of time talking about dynamics at work during the study because they changed by the time it was done. They changed because the actors changed. By the time the reports were issued, Mr. Morris and Captain Buesse and many senior academy officials had left. I can't say how these changes affected utilization, but I do feel they were important in developing something which could be utilized. I also believe that the various advisory groups served as a vital, independent communication link between us and the academies; and further, where our methodology and findings were accepted by them and communicated to the academies, the probability of implementing those findings was increased.

The second major factor affecting the utilization was the intractable nature of some of the technical problems in inferring causality, interpreting factor scores, assuming a certain model, and nonresponses.

We had no control over treatments, and random assignment was out of the question. For that matter, we did not know enough about critical variables to design an experiment. Moreover, we did not have the time to conduct a panel analysis which would help us infer the direction of dynamic relations. Finally, the limited number of

academies precluded drawing meaningful conclusions about objective organizational characteristics such as ACE had done in assessing college environment impacts.

We were left, for the most part, with a post-hoc, correlational study based on self-reports of academy experiences and evaluation. In short, a weak foundation upon which to base recommendations for change.

We collected a great deal of information on each student and dropout. At the prodding of the ad hoc group, we performed a series of factor analyses on the data.

For those of you who have done factor analyses before, let me say we learned something. The computer-generated factor scores were occasionally uninterpretable when one compared item validities with factor validities. Some of the factors were accounting for negative variance. And finally, the structure of some of the factors made it difficult to develop recommendations.

We assumed the general linear model throughout; and, perhaps as a result, the size of our correlations was not overly impressive.

Finally, while the rate of questionnaire return by dropouts was high (73 percent), it was not perfect. ACE conducted analyses of the non-respondent characteristics and could not conclude that they differed from the characteristics of those who did respond. By the same token, we could not conclude that the two groups were necessarily the same in terms of academy experience and evaluations.

We recognized all along that these limitations existed, and we candidly stated in our final report "that a correlational study (as

ours was) does not establish clear cause-and-effect relationships and that surveying student perceptions after the fact presents special problems of data interpretation. Alternative interpretations exist."

We tested the validity of those other interpretations as best we could, but admitted to not being able to recognize or test them all. Therefore, we went on to say "Because alternative interpretations are always possible from survey data of the type we collected, our conclusions and recommendations have been stated cautiously."

Despite these limitations, we felt we learned some things from our study. Probably the principal reason is that there was a research base on which we could build.

Several of the academies had been doing attrition-related research for years. We collected all of the studies that could be identified and focused on 84 of them for detailed analysis and synthesis.

These studies left us with two impressions. First, very few of them had to do with the environment at the academies--far and away, the majority had to do with the relationship between characteristics at entry and attrition. And second, perhaps only one of the environment studies could be considered to possess what Stanley and Campbell refer to as "internal validity"--the sine qua non of scientific research.

We found the studies useful, nonetheless, because they explored dimensions of student characteristics that we did not explore. Some of what first appeared to be anomalous responses in our questionnaire--i.e., dropouts responding the way we hypothesized current students would, and vice versa--became interpretable only when we considered the implications of those entry dimensions the academies had explored.

Let me give you an example. At West Point, we found that those who stayed were less certain than those who left about their responsibilities and about what officers or upperclassmen thought of their performance. Similarly, those who stayed reported being bothered by having too little authority and responsibility. Also, their view of leadership was to have upperclassmen encouraging them to give their best effort and maintain high standards of performance. Dropouts, on the other hand, had a view of leadership as support from classmates.

These findings became interpretable (because we were asking questions about the environment) only when previous academy research on personality characteristics was viewed in light of the intensely competitive environment of the academies. That research indicates dropouts are largely non-competitive and are not achievement-oriented. They appear to have higher needs for affiliation and affection. Those who stay are concerned about achieving in terms of a standard of excellence, and are more independent in their interpersonal relations. Clearly, role ambiguity and not feeling enough responsibility would be bothersome to such people.

After arriving at this interpretation, we suggested that West Point might want to reexamine the amount of stress and competition in its environment. The Academy and DOD didn't like that suggestion. They pointed out that the stress and competition simulated what graduates would face on the battlefield where they would be responsible for the lives of others. We allowed as how this argument had appeal, but questioned what it meant with respect to other officer acquisition programs where students do not experience the same level of stress and competition 24 hours a day, 7 days a week, for 4 years. I don't think that any effect the reasoning out of the implications of the argument

might have had can be separated from the effects which the recent cheating scandal and Congressional interest in academy competition might have had.

As Chuck mentioned, where there was a predisposition by senior academy officials to accept our findings, they were in fact acted upon. The amount of drills and ceremonies and rote-memory of trivia were reduced, and some extracurricular activities were instituted.

However, where we challenged deeply-ingrained attitudes about the academies, there was strong resistance to our findings. The competitive environment was one such area. Another was the finding that dropouts did not perceive the educational program as having the high quality which the current students did.

The possibility of longitudinal research was precluded by the steps we took to insure confidentiality. Such steps drive up the cost of this type of research because you can't amortize the cost of design and data collection through repeated measurements. Nonetheless, we believe our study does add to the academies' fund of knowledge. But more importantly, we are an agency of the Congress; and ultimately our work should feed into their decision-making. In this case, it did. The Senate Committee on Appropriations used information from our study as one justification for recommending closure of the academy prep schools. The Committee also expressed an intent to critically review the academies' actions regarding our recommendations, and specifically with respect to the competition in the environment--a finding with which they agreed. Thank you.

MR. GRANDY:

Thank you very much, gentlemen.

III. DISCUSSION (SPEAKERS AND SYMPOSIUM PARTICIPANTS)

MR. GRANDY:

Let's take a few minutes at this point and see if you have some questions of these speakers about their work or their presentation.

PARTICIPANT:

I'm Joel Garner with LEAA. I have a few questions, one of which directly relates to your study. First, who do you consider was your primary audience and who in effect set the objectives for the study? Clearly you work for the Congress and two specific Senators asked for this study. But it seems from your discussion that you worked primarily and directly with the service academies and with the Department of Defense, and that they, in the operation of the study, became the primary users or individuals who set the objectives of the study or of the evaluation. The second question is one that other people have raised but never answered, and the question is, should Congress in either legislation or requests to GAO be required to set specific evaluable objectives? If we can't expect agency administrators to do this, maybe we should expect Congress to do that and only do evaluations when Congress is very explicit about its objectives.

MR. THOMPSON:

I don't think there is any question our primary focus was to the Congress. We work for them, and we report to them. But we also recognized that we have other potential points at which we can have the results of our work implemented. If we can work with the academies throughout the job, still maintaining our independence, and get them to appreciate what we are trying to do, the objectiveness of what we are trying to do, they are much more likely to act on their own

to implement some of our recommendations. I think at least to some degree, that was the case. It was clearly the requests of Senators Proxmire and Bayh which set the scope for our work.

MR. HARPER:

What we tend to find is that the process of doing our work oftentimes achieves whatever objectives we might have hoped to achieve by producing a report because, at the point in time when the agency gets the opportunity to comment on the drafts of our report, my experience is that the agency likes to say, "we have already instituted action to correct whatever it is you have found."

MR. GARNER:

Did Senator Proxmire and Senator Bayh use your report in any way? Did they do anything with it?

MR. HARPER:

Let me just mention that we wound up ultimately on this job with something in the neighborhood of 12 to 16 requests from Congress, letters from Congressmen asking us to do work in this area. One of the spinoffs of this job was to look at the training programs in terms of the amount of harrassment that was going on. We issued a separate report on that.

MR. THOMPSON:

Frequently on a request from someone like Senator Proxmire, the request comes not from him as an individual, but as a member or Chairman of a subcommittee. At the completion of our review, the Comptroller General testified before the House Subcommittee on Legislation and National Security, Committee on Government Operations. It's really that Subcommittee that has the power to act.

One other point, I think, that is important is that very frequently we find that change takes place before the study is finished. Just the pressure of the Congressional inquiry, just the pressure of GAO looking at the problem is enough to get the academies or to get the agency to begin seriously thinking about it and begin to make changes. What we found was that shortly after we got in, attrition began to come down. Whether it was as a result of our work or not, I don't know. It's nice to think so.

PARTICIPANT:

Paul Hammond, University of Pittsburgh. I am struck that the description you have offered us is of a study commissioned by Congress in which you have not characterized--you have not described a persistent set of contacts with your Congressional mentors and in which, if I were a suspicious-minded Senator, I might have concluded you sold out to the enemy and made your deals before you got to me. I am familiar with this kind of working with people who are the subject of studies. Twelve years at the Rand Corporation makes one familiar with that process. But particularly where one is working across two constitutional branches of Government, I wonder if you could not say something more about how much your Congressional clients concerned themselves with whether you are not, for example, putting out premature signals as to which direction the academies should be moving in terms of reform; or conversely, do you want to simply tell us that in a situation like that, the answers are so obvious that the moment concern is expressed, everyone knows which direction to move in?

MR. HARPER:

You have asked a number of questions. I'll take the easiest one and leave the rest of them to Chuck. We in fact did have contact throughout this job with staff people from the various Congressional offices. I participated in meetings with Senator Bayh's and

Senator Proxmire's staff people. Also, as tends to happen in any organization, you get people who come to have specialized functions. On this job, we had someone who was our point of contact with the Congressional people. Our concern was with getting a study done that was as right as it could be from a scientific or methodological point of view because I think both he and I, as well as other people in the organization, saw that the study had great danger for us. There are a group of people up on the Hill who would have liked nothing better than to really challenge whether the service academies ought to exist. There is another group of people that would think that is the worst kind of heresy you could perpetrate. The only way we could win on that one was having something that we could at least defend from a methodological point of view. That was our concern.

MR. THOMPSON:

In terms of setting the direction for change, I think just the fact the Congress was questioning the high and increasing attrition pretty well laid the pressure for where the academy should be directing their attention. I don't think also that our being involved in the academies and with DOD throughout the study had any bias, so to speak, on the results. It was our feeling that it was the best way to go to try to get maximum utilization of what we had to offer when we got through.

PARTICIPANT:

John McGruder from the Department of Transportation. I am curious about whether you looked into historical periods, because it would seem to me that in the mid '70's, with Vietnam over, that you might have found somewhat the same kind of situation that existed after World War II, or after World War I, when it really was not unusual to have a higher attrition in the academies. At least it

would be my expectation that that would be the case. To what extent was this a reason or did you look at it at all? You didn't mention that.

Secondly, your study was obviously a tremendous effort and yet it seems to me as if the principal finding that you came up with was that those who aren't very interested in the program to begin with are much more likely to drop out. And I guess I wasn't too surprised by that finding. Can you help me out a little bit because I am not really amazed by that?

MR. THOMPSON:

Let me deal with the last one first, and John can take the first one last. The commitment one is just the example I used. There are three pages of conclusions and recommendations in the report which deal with all aspects of the academy. In terms of the commitment conclusion, it may seem obvious to you and it does to me; yet, given the program at the academies and the way in which academy people talked about those programs indicated to us that it wasn't as obvious to them.

MR. HARPER:

You see their assumption had been that when the students walked in the door, they were committed. The question was, let's find those who are the most committed and keep them. What we were suggesting was a major policy kind of change in the sense that the emphasis in the program wasn't to be just testing, but was also to be motivating. That is, let's not make that assumption that they are committed. Let's try to develop that commitment in them. It's an expensive process to lose these people.

Let me deal with the first question. The point was made very strongly to us by the academies that we should look at historical data. There turned out to be only two academies that had that kind of data, and they were West Point and the Naval Academy. Their attrition after World War I and World War II was quite high. It was low during the depression and that sort of thing. Our major argument for not using it was that there simply wasn't sufficient data to go back and say that there were enough conditions in those periods in time which were similar to the conditions here with only one variable changing--that is, end of war, not end of war. Also if I recall correctly, the West Point and Naval Academy figures didn't jibe. They were conflicting. The Naval Academy's figures showed a steady attrition rate after the wars--steady in terms of what it had been during the war and before the war. West Point was very anxious for us to use those figures, and the Naval Academy wasn't; so we didn't figure they were relevant because of the kinds of arguments that I mentioned.

PARTICIPANT:

I'm Tom Richardson, Department of Commerce. How in fact do you know that the people that are coming out of the system are really the best officers? I would gather that the people who rate them as good mean, they are responsible, they do what they are told, et cetera, et cetera. In fact, given the national interest, perhaps the ones that are dropping out are the best candidates and the ones that are staying in are not. It seems to me that that is an important question to look at, too.

MR. THOMPSON:

No question about it. The problem is we only had two years to do the study. We had to cut off a small piece. The effectiveness of the academies' programs is not only a function, we recognize this,

of the people that go through but also the performance of the graduates. We just didn't have the time or resources to look into that.

MR. RICHARDSON:

In fact you don't know about that?

MR. HARPER:

We present data--what you have in xerox is the main report. There are three appendices that were attached to that. These are printed on both sides of the page so there is a lot of reading matter here. The Appendix B, where we synthesize the academies' studies, presents a lot of information about what the dropouts are like as compared to those who stayed. I suspect you can go through there and form some judgement as to whether they are losing the kinds of people that they should keep. It's a process of socialization. To some extent, at least our review of the studies indicates those they lost at the very beginning are the ones they produce at the end. In other words, they lost leaders right at the beginning because they are not going to socialize to that kind of a system. They are already leaders. Then they go through and mold the others into leaders.

MR. THOMPSON:

I think another important point that came out of our review of the studies was that, as John mentioned earlier, most of the studies we looked at were directed towards trying to control the attrition through controlling selection. There was a very strong reluctance to examine the environment. Our concern there was that over a long period of time, if you began to use the graduates as the criteria by which you selected new students, and then create a cycle, you begin to narrow the diversity of the people that come in to the point where eventually you have one type of person coming out. We weren't convinced this was in the best public interest.

MR. ABERT:

You mentioned that you looked at the characteristics of the students when they came in. I was looking at Barron's 1976 the other day, and I was surprised to see essentially how competitive the military academies are in terms of their ability to attract or at least accept students with the highest academic standards. I would say that across the board in the military academies you are talking about the 97th percentile on Student Aptitude Tests with the Naval Academy a little higher than that. What do you find about retention correlated against high, low, medium SAT scores?

MR. HARPER:

In some academies and in some time frames, there is a positive correlation between measures of academic ability and attrition. When I say positive, what I am talking about are the coefficients that are very small and significant because the Ns are so large. We wound up concluding that while there is some relationship, even those who drop out are of very high ability.

MR. THOMPSON:

Just to give you an example of what we are talking about in terms of the quality of the incoming graduates, in the Air Force Academy for the SAT math score, the average was around 660. The national average was around 460. That is pretty consistent, except for the Merchant Marine Academy, in terms of quality.

PARTICIPANT:

Jim Robinson, Department of Labor. I thought the most interesting thing in your study was something you glossed over very quickly which is that one morning, the academics announced to you that there would be no study or that they wouldn't cooperate with it. Isn't perhaps the most interesting finding in your whole study that there is a basic

question of who is in charge of the military and that the military could sit there and say, "That's it. We are not going to cooperate with your study. Good-bye. Your project is over."

I mean perhaps that is off the point of the evaluation study proper, but certainly the whole idea of your study is that Congress had some right to go in there and ask the military some questions; yet at least some of the military people questioned your authority to even be there and decided it was up to them whether they wanted to cooperate, rather than the other way around under the Constitution.

MR. THOMPSON:

That happens a lot. It's not just DOD. We get questioned a lot about why we are there and we get a lot of flack. I think in the end, we normally get in, so we are used to it.

PARTICIPANT:

Walter Bergman, IRS. One thing I am really interested in, in addition to what you have brought up (obviously, I don't know what is in all those volumes). Did you in any way impact on the selection system? Did you have feedback into the selection process itself?

MR. HARPER:

We discussed in our report the question of whether you could adequately present a picture of the academy to someone who hadn't been there. We weren't talking about changing the selection procedures so much as we were talking about giving adequate information to people about what the academy was like so they could select themselves out before they were nominated or appointed. Our feeling was that there needed to be more of that because the research is quite good in that area to indicate this is a good way to go about bringing people into the organization. But we also talked about perhaps not being able

to do that adequately with the kind of life you are talking about at the academy so what they need to do is identify early in that first summer the people whose commitment is wavering. What we saw in the area of traditionally collected selection variables were such weak relationships that I don't think we would have wanted to make any recommendation anyway. For the most part because they have been selecting top level people, there is no variance. There is very little variance on these measures. They have already been preselected, so you don't find the kind of correlations you normally do. It is hard to develop tests to measure commitment prior to entry.

MR. THOMPSON:

I think also we felt that the academies were doing more than enough research on selection, and in fact our recommendation in the report is that they provide a little more balance in their research and start examining their environment a little bit more in terms of its contribution to attrition.

MR. GRANDY:

Thank you very much, Chuck and John. We'll adjourn our program at this point, pick it up tomorrow morning with the other research papers.

BANQUET ADDRESS NOVEMBER 17, 1976

A CONGRESSIONAL VIEW OF PROGRAM EVALUATION

DONALD ELISBURG, Staff Director and General Counsel
U.S. Committee on Labor and Public Welfare

MR. GRANDY:

As you know, our guest speaker tonight is Donald Elisburg. Mr. Elisburg had a distinguished career in the Department of Labor for quite a few years prior to assuming his position with the Committee on Labor and Public Welfare in the Senate where he is currently the General Counsel. He has his juris doctorate from the University of Chicago and served in the Department of Labor in a variety of positions from 1963 up until 1970 when he joined the Senate staff. He has, I think, a keen appreciation and knowledge of evaluation problems. From our discussion during dinner, I know he has some interesting views. His comments concerning the perspective of the Congress on evaluation is likely to be very helpful to us and thought-provoking. We do appreciate his being able to take time from a busy schedule to address us. Mr. Elisburg.

MR. ELISBURG:

Thank you. As many of you may realize, speakers from the Congress almost always begin with a certain amount of disclaimer. Despite the four or five thousand professional staff people who work for the Congress and the Senate, there are only 535 elected representatives; and when you are employed by a committee, you are responsible to the Chairman (in this case, Senator Williams of New Jersey). I always remark that he is free to disavow anything I have to say on any subject. I'll give you the best views I can, but they are my own and I hope you don't take them as necessarily attributable to the elected officials.

In thinking about the process of evaluation, and the study which was presented to you this afternoon, it occurred to me at dinner that the Congress is really engaged in a tremendous amount of evaluation through its arm of the General Accounting Office, but we don't normally think about it in terms of evaluation. We think about it as having GAO do a study, or GAO do an investigation, or something breaks loose in the newspaper and you say, "Oh, boy, we better do something about it," so they send the GAO in to take a look. That really is a rather extensive investigative arm and, therefore, to some degree, an evaluative arm of Congress. Lest anybody think this is not a significant kind of career that those of you in the business have embarked on, this afternoon there came across my desk one of those documents that you never look at. But because it was a very nice package and had a little short note clipped to it, I decided I would at least take a look at it before I put it in my outbox. The title of this nice book is, "Recurring Reports to the Congress: a Directory. 1976 Congressional Resource Book Series."

The note says that this comes from the GAO and a copy is enclosed for your use. The third volume, Federal Program Evaluations, will be distributed in December. This particular document is, I guess, a list of various kinds of reports that come to the Congress for the umteen thousand statutes that exist. The point is that this source book and Volume I which came out about six weeks ago, listed the various kinds of data collection processes that various agencies use. It was a red book, compiling lists of things that the Government is doing and that you are all doing, either as members of the Federal establishment or engaged in some relation to it. My point is that this kind of compilation never existed before. People in Congress probably have no idea of what they have fostered over the years, and perhaps after a couple years of putting this book out they will be sorry they ever got into it.

But I think it's clear we are in some new arenas of doing business with each other. Consequently I really appreciate the invitation to address this symposium.

There are always innumerable conferences going on in Washington. When you think about it, though, I dare say few, if any, are as important in substance as this one which examines how Federal agencies utilize program evaluation. The subject requires intense consideration if the Executive Branch is going to maximize its administrative responsibilities in implementing programs fostered and enacted by the Congress.

Personally I am pleased that this part of the symposium in process includes the Congressional view of evaluation as well. Hopefully the participation of those of us who are connected in some way with the Congress will contribute to the success of your very timely and essential meeting.

The Congress differs markedly from the administrative agencies with which many of you are associated. It is a body responsive to the wishes of multiple, sometimes conflicting, sometimes shifting constituencies. Almost all Governments have an Executive Branch. Our nation is one of the few in the world which entrusts its lawmaking to an independent, periodically elected representative body. And that body functions in a milieu which, by its very nature, is heavily politicized.

By politicized, I mean that the Congress listens carefully and continuously to its broad array of constituencies. I think that this listening is the Members' first and most basic source of evaluation: it is a very finely tuned antenna. Sometimes the listening is carried out scientifically through the use of sample surveys. Sometimes it is carried out intuitively, as when rumblings and

grumblings, wishes and preferences are brought to awareness by the delegations to the offices of the Senators, Representatives or Committees. The unrelenting deluge of mail and the representatives of special interest groups bringing their clients' wishes and complaints to Congressional attention are two additional sources of evaluation playing a role in the assessment of policies and programs.

This is probably the method of evaluation prescribed or implied in the Constitution. It is the means by which our system of Government has worked for two hundred years. It is not a perfect system as everyone knows, but it has been, on the whole, a successful device to balance overwhelming societal concerns with individual liberty and rights.

In spite of cyclical praise or scorn, the Congress has maintained, as its primary means of evaluation, the legislative judgment for which it is accountable to the electorate. This basic fact alone conditions the way in which more systematic, scientific program evaluation is viewed by the Congress.

While growing recognition of professionally-based, expertly-conducted program evaluation has been evident in recent years in the Congress, legislators and their staffs view this important secondary evaluation supplement within the political framework Constitutionally required of them.

Systematic impact assessment of Government policies and programs has been accorded increasing acceptance by Congress. However, the products of such assessments are looked upon as tools with which to shape the essential substance of programs attracting a following or an opposition among the constituencies having an interest in them, including that amorphous electorate whose opinions are often made known only on Election Day.

It cannot be stressed too strenuously that scientific program evaluation is itself evaluated by the Congress in terms of its utility to promote the effectiveness and precision of legislative judgments in a political milieu.

In recent years, program evaluation has been made a requirement of many policies and programs enacted into law. In some cases, this requirement has taken the form of directives to Cabinet Officials to set aside and allocate a fixed proportion of funds to evaluate a selected program. In other cases, impact statements have been required. Impact statements can utilize program evaluation together with other research devices designed to provide assessments of net impact. Both forms of Congressionally-mandated evaluation have the same purpose: to delegate to the Executive Branch a duty to determine what if anything happened as a consequence of the policies or programs tagged for special review. This may be another form of cop-out, perhaps by the Congress, but it is a way of putting the burden on the Executive Branch to do the work.

More recently, Congress has asked directly whether or not our policies and programs are cost-effective--whether we as the public are getting "the right authorized impact of the legislatively appropriated dollar," and whether the nation's economic interests and social well-being have in fact been promoted, especially by human resource efforts.

Much of this budget-related interest emanates from the provisions and procedures of the Congressional Budget and Impoundments Control Act of 1974. Many Senators and Representatives favor the Act, in part because it provides a budgetary window into the inner workings of programs. That window is made possible by fiscal analysis and fiscal priority-setting. I would venture to say that many Senators

and Representatives are not totally pleased with the Budget Act because the scientifically-developed new procedures--that require Authorizing Committees to look at specific dollar amounts and that require an overall picture of what is being appropriated and what isn't--have raised very serious questions, particularly in the social arena, like, How do you keep from being shortchanged because you don't come up with the right numbers on the computer?

More recently, Congress has begun to consider in the formulation of the Sunset Bill, steps which would institutionalize program evaluation and review at the heart of Congressional decision-making. In the Senate, for example, the Government Spending and Economy Act of 1976--that was the Sunset Bill--proposed that programs be terminated on a mandatory basis every five years and reauthorized only after a close-scrutiny program review. Fortunately, the bill was not acted upon, but the idea has attracted a following in Congress. While legislation of this nature has many features, the degree of dependence on program review techniques would be tremendous. Were Sunset Legislation enacted in some form, recognition of program evaluation as a secondary means of Congressional decision-making would have attained an enhanced status. I think it would be nicknamed the Evaluators' Full Employment Act. It would also have been accorded grave responsibility as an instrument of public trust.

But even if such developments were to occur, would the public trust indeed be well-placed? What would the Congress be buying?

Many experienced Congressmen and their staffs are concerned that the Congress will become dependent upon a program evaluation establishment--valuable in concept, but unproven in product. Opponents of the systematic use of program evaluation point out that such research

is an art form of marginal reliability and that reliance upon such an art form is in itself more in the nature of folk medicine than of science.

The issues cited as the source of such suspicion are commonplace:

- that the assumed posture of objectivity among program evaluators often masks subtle but important biases and hidden agendas;
- that the questions set for discovery, if published at all for client consideration, have predetermined answers;
- that the procedures utilized frequently neglect the most important variables often included in initial designs and later dropped because of difficulty in research management or unexpected costs;
- that there persists an inability or unwillingness to merge the contours of various impact evaluation studies so that common patterns of findings can be codified and differences in findings highlighted;
- that interpretations of findings are cast in terms far in excess of their value and far overstated to listening audiences; and
- that the conduct and packaging of evaluative research supports first the publication interests of the investigators and too often relegates the needs of clients and sponsors to second place.

Whether or not these assertions can be supported substantially, the doubts exist. Program evaluation experts will point out that the Congress has its own peculiarities, biases and statements which lose support when subjected to rigorous analysis. But the Congress bears the accountability of the electoral process in setting forth its assertions into law, overseeing its handiwork, and supporting its decisions from the Federal Treasury. Obviously, the task before us is to look beyond the concerns (while keeping them in mind) in order to explore some principles which would enhance the utilization of program evaluation by the Congress. The task requiring attention

is to develop program evaluation standards and approaches which will notably assist the Congress in its accountability for public policy.

If program evaluation is to become truly useful to the Congress, those conducting research as the agents of the elected officials should consider three principles in adapting their works to the needs of the Federal Legislature.

The first principle, and perhaps the most difficult to achieve, is that program evaluation must be preceded by policy analysis and mission analysis. Policy analysis in turn calls for a rigorous study of the substance of the policies giving rise to programs. Policy analysis calls for the consideration of the goals enunciated during the formulation of policies and programs. Policy analysis requires attention to drift and shift between policy as legislatively mandated, and policy as executively implemented. Policy analysis requires careful attention to the process, the actors, the subtle differences which result in a policy product.

Mission analysis is the description and explanation of whether a program adheres to the objectives set forth in the policy. The fundamental question of concern to the Congress is whether a program carries out the mission established for it in the policy. It is to that issue that constituent concerns are addressed as well.

I would stress here that policy and mission analysis require as much research skill and time as any other element demanding the attention of the program evaluator. Policy analysis requires case study techniques; selective use of surveys; employment of content analysis of documents; and utilization of journalistic and investigative techniques which employ accepted standards of corroboration. It also means you have to be able to write clearly.

The foregoing implies, of course, that the researcher can generate the trust necessary to conduct an adequate policy analysis as a preliminary step to informing the Congress about the impact of programs. But that trust is essential inasmuch as the questions central to program evaluation are likely to be derived from policy and mission analysis.

I would also add here as an important factor that most members of Congress and their staff have been trained with legal concepts and investigative techniques. It is not surprising then that they frequently regard the standard of evidence utilized by many program evaluators as inadequate. When one reads through program evaluation reports and is struck by the large number of tables pronouncing this test as statistically significant and that test as unassailable evidence of a particular program impact and one reads on further to find that conclusions have been drawn entirely from aggregating such statistical inferences of proof, it is not surprising that the clear and convincing evidence standards, or the preponderance of the weight of the evidence standard used in legal thinking seems, by contrast, far more reliable.

In short, persons connected with the legislative process are not likely to be convinced that large numbers of associations of variables prove a point. Common sense requires complex situations to be judged with all available evidence--both the context of the situation and the specific variables artifically isolated for examination--before conclusions can be made. That is scientific jargon for saying that you must do a careful job. Program evaluation and policy analysis, in particular, will be judged by the Congress according to a standard of evidence not usually advanced in the program evaluation with which many of you may be involved.

A second principle to assist the adaptation of program evaluation to legislative activities is that the evaluator must understand where the Congress will find evaluation useful. While evaluation studies may be useful in the formulation of bills, program evaluation is most relevant to Congressional oversight. Congressional oversight is a shorthand term we all use for what we do, with a broad license to do anything, after a statute has been passed. Congressional oversight is the means by which Congress accounts for the policies and programs it authorizes and appropriates. The common techniques utilized in Congressional oversight include investigations; hearings; site visits; audits; analyses of special and recurring reports required by statutes; meetings and meetings and meetings to consider the impact of appropriations of funds for program support; procedures to consider formulation of the Federal budget under the provisions of the Budget Act, and so on and so on. Obviously program evaluation could have a strong role to play in some of these activities, a lesser role in others. The important point is that an understanding of the conduct of oversight is itself important. Familiarization with the techniques utilized, procedures employed and the settings for oversight activity cannot be substituted.

Finally, I would suggest that attention be given to the way in which program evaluation studies are interpreted, presented and packaged. Congress, I am sure, is acutely aware that various constituencies in a political milieu may be activated in favor of or opposed to a program by the expert character of an evaluation report. I might also add that the Congress or individual Senators or Congressmen may well be influenced by whether you can relate the five years of your evaluation study in the 15 minutes that you have at a public hearing, that is, how well you can do it, how well you can synthesize and set forth, while you are sitting there on a TV camera, the essentials of what you have been trying to do.

One is also acutely aware that public hopes ride high on programs finally forged from Congressional actions. Hard-won advances, particularly in the human resources field, may suffer permanent, unwarranted damage if the evaluation and interpretations are unjustifiably sweeping, if packaging is conducive to sensationalism in the public media and if presentation does not relate to the concepts or procedures conventionally employed by the Congress. Hopefully, this foregoing litany will provide some basis for your discussions tomorrow as the business of the symposium proceeds. I think the Legislative Branch has an important stake in program evaluation as it goes about making and shaping the public policy with which we are all going to live. The prospects for Congressional utilization of program evaluation are very great. In our own Labor and Public Welfare Committee, we have begun for the first time in its history to institutionalize some of the evaluation ideas; and that primarily means to appoint relatively permanent staff to think about it. That is a big step. It's a big step in a fairly tight-budgeted operation, where you have relatively small numbers of people, to assign someone to start thinking about the evaluation of programs and something resembling a systematized oversight.

The evaluation research and the people who conduct it--that is, all of you--may very well become a very important augmentation to the fundamental framework of legislative decision-making. You may not all welcome the prospect, but I think it's more than just around the corner. It's true because of the Budget Act and many of the other possibilities, the Sunset Act, for example, that have evolved around the Congress, and the fact that the Congress is now dealing with a budget of some $400 million and some odd a year, really a billion a year. Evaluation is really a kind of program technique that is not new, but it is

something that is going to increase its respectability; and consequently, I think it is going to be an important adjunct to the legislative process. Thank you.

PARTICIPANT:

I liked your speech very much. Congress is my first interest, and I was pulled into criminal justice for want of a job. We have been talking about the need to specify objectives, and Congress often passes acts like the Crime Control Act that says, "Reduce crime and improve efficiency." The Act itself has to specify the objectives of that Act, and how do you measure that? Is it reasonable to expect that Congress might specify objectives very clearly--that the objective of an act could be very specifically stated in the act itself?

I give as an example the Speedy Trial Act of 1974, where Congress not only specified the objectives, and showed how to figure out whether a speedy trial is achieved, but wrote the evaluation design into the Act itself. It has been done at least in that one case. The point is that agency administrators never specify objectives. Can Congress do it?

MR. ELISBURG:

I understand the point. The legislative process does not lend itself to regulation writing. By and large when the Congress has gotten into writing in detail the specifications of how it wants something carried out, it either gets into trouble or the events of time pass it by; and you have to relegislate. With respect to the question of being able to spell out the policy, however, almost every major piece of legislation has a findings and purpose section which can go on ad infinitum trying to spell it out. I would recommend to anyone dealing with a serious legislative enactment, for example, a major program, that you look not just at the words in the statute, but

that you look at the legislative history and the committee reports, I think you will find from these sources a more detailed program lay-out of what it is the Congress intended and wants done with these programs.

PARTICIPANT:

I have a two-part question. The first part deals with your statement concerning policy evaluation. I'd say that all during the day, there has been some question about what kind of focus we should have in our evaluation. Some suggestions were that evaluation would be better if it were narrowed to doing what our boss wants to see. Others, more expansive, wanted to do what everybody wants to see. What would your views be on that?

My second point has to do with the acceptance by the Congressional Representatives of this information. Friends of mine in the Congressional Budget Office say that, in fact, they feel that their activities are viewed by many of the Representatives as constraining. The fact that they come up with facts means the decisions that the Representatives can make are somewhat weakened in certain lights. I would see evaluation providing the same kind of data which would be equally constraining. How do you feel about that?

MR. ELISBURG:

As to the first part, I would view the question of how an evaluation should be done from the standpoint of whether I was the boss or everybody else. I think the problem is really of defining the policy. You have to really take the time to understand in a legislative context what it is that the Congress had in mind, what the objectives were, and how those objectives have been met? What was it that the Congress was trying to set forth? Otherwise you might just as well be evaluating apples when Congress is talking about oranges.

The second part as, to whether anybody is going to feel constrained, is really a question of growth and development of the institution with which you are dealing. Fifteen years ago, the Senate Labor and Public Welfare Committee had a dozen employees and very few statutes within its jurisdiction. The Senate Labor and Public Welfare Committee now has in excess of 125 employees, which is not necessarily large in terms of a Government agency, but it is responsible for reviewing programs which represent in excess of $40 billion a year. There are literally hundreds of them. When you are talking about legislatures which have to deal with that kind of fantastic growth in legislative programs, newer techniques will have to be used. For the first time in the history of the Senate, really, we are getting a computer capability that most Federal agencies had 15 years ago and most of private industry had 20 years ago. It's a growth process. There is a realization and understanding that these techniques are going to have to be used, constraining or not. So to that extent, Congress, like a lot of other groups, is being dragged kicking and screaming into the 20th century. Thank you.

END