

135786

United States General Accounting Office

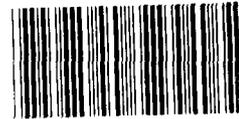
GAO

Report to the Chairman, Subcommittee on
Federal Services, Post Office, and Civil
Service

May 1988

FEDERAL PERSONNEL

Observations on the Navy's Personnel Management Demonstration Project



135786

042119/135786

Blank header section with a solid horizontal line at the top and a dashed vertical line on the left side.

Main body of the page, mostly blank, with a solid horizontal line at the top and a dashed vertical line on the left side.

General Government Division

B-203022

May 3, 1988

The Honorable David H. Pryor
Chairman, Subcommittee on
Federal Services, Post Office,
and Civil Service
Committee on Governmental Affairs
United States Senate

Dear Mr. Chairman:

This is in response to your Subcommittee's request that we report on the Navy's personnel management demonstration project operating at two of its laboratories. These are the Naval Weapons Center (NWC) in China Lake, California, and the Naval Ocean Systems Center (NOSC) in San Diego, California.

The Navy's project instituted procedural changes in the areas of position classification, performance appraisal, and pay. It intended to demonstrate that the effectiveness of federal laboratories could be enhanced by (1) allowing greater managerial control over personnel functions and (2) expanding opportunities available to employees through a more responsive and flexible personnel system. The project began in July 1980 and covered 7,655 employees as of January 1988. The Office of Personnel Management (OPM) has estimated that salary costs at the demonstration laboratories have increased by about 1 percent per year since the project began.

The proposed Defense Industry and Technology Act of 1988 (S. 2254), which was introduced on March 31, 1988, would, among other things, enable the Department of Defense (DOD) to apply the alternative personnel system developed at the Navy's two laboratories to up to 300,000 employees in DOD acquisition and logistics organizations. The act's goal is to resolve recruitment and retention problems in these DOD organizations by giving DOD managers new flexibility in managing and rewarding their work forces.

As requested, we have focused this report on matters particularly relevant to the goal of S. 2254. Thus, our report addresses several key "lessons learned" from the Navy's project that Congress, OPM, and other federal agencies should take into account when authorizing, funding, designing, and evaluating future personnel system projects.

Appendix I explains our findings and observations. Appendix II contains our letter to the Chairman of the Senate Committee on Armed Services answering six questions relating to S. 2254 that were asked by Senator John Glenn. The answers provide additional insight into the matters discussed in this report. In particular, our response to the third question elaborates on what we believe would constitute an adequate evaluation design for a project like NOSC and NWC.

Results in Brief

The Navy's project demonstrates that a pay-for-performance system with such process features as broad pay bands, simpler position classification, and closer linkages between performance and pay can be implemented to the general satisfaction of many managers and employees. Although we found no significant deficiencies attributable to the project, insufficient data were available to support a conclusion that the project had successfully met its objectives of enhanced laboratory effectiveness, greater managerial flexibility to assign work, and improved employee recruitment and retention.

Given the inherent difficulties of doing a rigorous, controlled evaluation of a complex program in a dynamic environment, it would not have been reasonable to expect OPM's evaluation of this one project to have fully answered all questions. However, OPM's evaluation did represent an important step in building a knowledge base on systemic methods of improving personnel management and performance. By conducting more small-scale projects of this nature and constantly improving evaluations of them, the weaknesses of individual studies will become less prominent and the applicability of findings to other settings will be enhanced.

Unlike the Navy's project, the Defense Industry and Technology Act proposes implementation in a budget-neutral manner with no net increase in salary and administrative costs. Given the large number of employees to be involved and the fact that a budget-neutral design has not yet been sufficiently tested (see p. 20), Congress may instead wish to further experiment by authorizing an approach allowing several projects with different project designs and smaller numbers of employees to be tested in a carefully planned and monitored manner.

Objectives, Scope, and Methodology

Our objectives were, to the extent practical, to (1) determine whether the results expected from the personnel system changes at NOSC and NWC had been achieved and (2) provide information that would assist the

Subcommittee in considering additional proposals to demonstrate or implement the project's procedures in other federal agencies.

We reviewed various documents that explained the policies and procedures for implementing the demonstration project and described the methods used by OPM and the Navy for evaluating the results of the personnel changes brought about by the project. We primarily relied on the evaluation plan and the 10 reports prepared by OPM on the project and statistical data provided by NOSC and NWC. A list of the 10 reports prepared by OPM is included in appendix III. We did not verify the accuracy of the information in the evaluation reports or the statistical data provided to us by NOSC and NWC.

To obtain additional views of persons responsible for implementing, monitoring, and assessing project results, we interviewed officials at OPM, Navy headquarters, NOSC, and NWC. We also judgmentally selected and interviewed certain supervisors and employees at NOSC and NWC for their insight into how the demonstration project operated and how it affected them.

In addition, we interviewed selected officials at the Naval Air Development Center in Warminster, Pennsylvania, and the Naval Surface Warfare Center in Dahlgren, Virginia. The Navy selected these Centers as control laboratories so that determinations could be made of how the demonstration project affected personnel operations at NOSC and NWC compared to the control laboratories where such changes did not occur.

We did our study between March 1987 and December 1987 in accordance with generally accepted government auditing standards.

Summary of Findings and Observations

The Navy's demonstration project showed that a pay-for-performance system with revised personnel processes to classify, appraise, and pay federal employees is workable. The project also showed that line managers could be given authority and responsibility for making personnel decisions—a factor the Navy considered to be essential in implementing the revised system. However, given the magnitude of missing data and the differences between the demonstration and control laboratories, we cannot assess whether the major outcome benefits cited by OPM are attributable to the change in personnel practices, to pre-existing differences between laboratories, or to outside factors.

Had it been fully implemented, the evaluation design developed for the Navy's demonstration project should have produced a clearer indication of the project's effect on such outcome objectives as recruitment and retention. However, it should be recognized that neither we nor OPM have identified any significant problems or serious adverse consequences attributable to the project. Furthermore, sufficient information is available to draw reasonable inferences on some key elements of the project. For example, it apparently did not give managers more flexibility to assign work. On the other hand, the project did produce a simpler, less burdensome, and less time-consuming position classification process. Further, it showed that a new personnel system with closer linkages between performance and pay could be implemented to the general satisfaction of many employees. However, because of data limitations, we cannot determine whether these aspects of the project, by themselves, warranted the expenditure of additional funds.

Even if the data on the Navy's project were conclusive and showed that the project was successful, this would not necessarily mean that it would be workable in other locations because of various differences between the Navy's laboratories and other organizations. Also, unlike the Navy's project, other proposals for alternative personnel systems call for their implementation in a budget-neutral manner—that is, they do not provide for increases in total costs. This is a major difference in design from the Navy's project and has not yet been sufficiently tested.

Generally, because of problems such as noncomparable pay between the public and private sectors, we believe it would be difficult to implement pay system alternatives that are budget neutral, especially if one of the goals is to improve recruitment and retention. Accordingly, we have strong reservations about endorsing the expansion of the Navy's project on a permanent and large-scale basis as proposed in S. 2254. Because of the untested nature of budget-neutral systems, the large number of employees potentially affected, and the difficulties many organizations have in developing good, agreed-upon productivity measures and in operating accurate and reliable management information systems, we suggest the Subcommittee proceed slowly and cautiously in considering proposals to expand the Navy's project to other locations in the manner envisioned in S. 2254.

Recommendation to the Subcommittee on Federal Services, Post Office, and Civil Service

Regarding future alternative personnel system projects, we recommend that the Subcommittee consider a series of projects smaller than those envisioned in S. 2254 in several locations throughout the country to test various design features and determine the applicability of the findings to other settings. Legislation should require that these projects be systematically evaluated and that the adequacy of the evaluation be carefully monitored.

In cases where projects are authorized on a test or demonstration basis, we suggest that the projects be designed so that within a given organization, some employees are covered while others are not. This would permit more valid comparisons and better measurement of project results.

For future projects that are authorized on a permanent basis, rather than as a test or demonstration, we suggest they be designed to eventually apply to all employees within an organization to avoid disparate treatment of employees. We recognize that this may require a process for phasing in coverage of various groups of employees.

Recommendation to the Director of OPM

We recommend that the Director ensure that future personnel system demonstration projects are carefully evaluated and fully documented. In addition, we recommend that the Director arrange for such evaluations to be closely monitored to ensure that they are implemented as designed, and that appropriate alternatives are developed and implemented if there are problems in executing the original evaluation.

Agency Views

As arranged with the Subcommittee, we did not obtain agency comments on a draft of this report. We did, however, discuss our findings and observations with officials of the Navy, DOD, and OPM. Navy and DOD officials generally concurred that there were shortcomings in the evaluation of the Navy's demonstration project. However, they believed that better data are now being gathered and indicated a willingness to take over the responsibility for evaluating this project in the future. OPM officials concurred that there were limitations in the evaluation of the Navy's project and with our recommendation that future demonstration projects be carefully evaluated and fully documented. OPM officials said that some of the baseline data needed for the evaluation were not available and that other problems were encountered in the early phases of the evaluation. The officials also stressed that in the future, adequate time should be provided for designing an evaluation before projects are implemented.

As requested, we plan no further distribution of this report until 30 days after its issuance unless you publicly announce its contents earlier. At that time, we will send copies to the Director, Office of Personnel Management; the Secretary of the Navy; and to other interested parties upon request.

Sincerely yours,



Richard L. Fogel
Assistant Comptroller General

Contents

Letter		1
Appendix I		10
Findings and	Background	10
Observations on the	The Project Is Workable but Its Outcomes Are Unclear	13
Navy's Personnel	Factors to Consider in Future Demonstration Projects	18
Management		
Demonstration Project		
Appendix II		23
Letter to the	Responses to Questions From Senator Glenn	24
Chairman, Senate		
Committee on Armed		
Services, Answering		
Six Questions on		
Senate Bill 2254		
Appendix III		33
List of OPM Reports		
Evaluating the Navy		
Demonstration Project		
Figure	Figure I.1: Career Paths and Pay Bands Compared to General Schedule	11

Abbreviations

DOD	Department of Defense
GAO	General Accounting Office
GPA	Grade point average
GS	General Schedule
NOSC	Naval Ocean Systems Center, San Diego, California
NWC	Naval Weapons Center, China Lake, California
OPM	Office of Personnel Management

Findings and Observations on the Navy's Personnel Management Demonstration Project

Background

The Naval Weapons Center (NWC) is located about 150 miles north of Los Angeles, California, in the Mojave Desert. At that location, personnel officials cited problems in (1) recruiting the numbers of qualified personnel needed because starting salaries were not competitive with private industry and (2) retaining qualified senior personnel because of limited promotional opportunities above the journeyman (i.e., GS-12) level. At the Naval Ocean Systems Center (NOSC), located near downtown San Diego, personnel officials described a different set of problems. These officials told us that line managers at NOSC lacked flexibility when assigning work to their employees because the General Schedule (GS) position classification process required too much paperwork and time. NOSC officials also expressed concern that the classification process was in the hands of personnel specialists rather than the line managers. It was their view that this situation made managers' jobs more difficult because it hindered their ability to effectively administer personnel resources.

The demonstration project, which involved 3,076 employees at NOSC and 4,579 at NWC as of January 1988, instituted changes to the GS system in the areas of classification, performance appraisal, and pay. It intended to demonstrate that the effectiveness of federal laboratories, such as NOSC and NWC, could be enhanced by (1) allowing greater managerial control over personnel functions and (2) expanding opportunities available to employees through a more responsive and flexible personnel system. The project's plan was approved by OPM in April 1980 and was implemented by the two laboratories beginning in July 1980.

Classification

The changes in the classification system primarily involved combining the 18 separate GS grade classifications into broad pay bands and streamlining the classification process. In lieu of GS grades, the project established five separate career paths, and within each path employees were placed into one of several broad pay bands. Each of these bands included at least two previous GS grades. Figure I.1 shows the general structure of the career paths, the pay bands established under the demonstration project, and the GS grades that the bands replaced.

**Appendix I
Findings and Observations on the Navy's
Personnel Management
Demonstration Project**

Figure I.1: Career Paths and Pay Bands Compared to General Schedule

Career Path							
Professional	GS	1-4	5-8	9-11	12-13	14-15	16-18
	Pay band	A	I	II	III	IV	V
Technician ^a	GS	1-4	5-8	9-10	11-12		
	Pay band	A	I	II	III		
Technical specialist	GS	1-4	5-8	9-10	11-12		
	Pay band	A	I	II	III		
Administrative specialist	GS	1-4	5-8	9-10	11-12		
	Pay band	A	I	II	III		
General ^b	GS	1-3	4-5	6-7	8-9	10-11	
	Pay band	A	I	II	III	IV	

^aAt NOSC, GS-8 technicians were included in band I; at NWC, GS-8 technicians were included in band II.

^bNOSC employees were included in the project as of September 1982. NWC general personnel were not included in the project until November 1987.

As part of changing the classification process, individual narrative position descriptions of GS occupations were replaced by broader classification standards and shorter, less individualized position descriptions. At NOSC, generic descriptions of the responsibilities of the positions in each band were used, while at NWC, position-specific descriptions were constructed from a menu of generic items describing employee duties, responsibilities, and difficulty in levels of work in each band.

Performance Appraisal

The demonstration project featured a performance appraisal system similar to the performance appraisal process required by the Civil Service Reform Act (5 U.S.C. 4302) for GS and other federal employees. Specifically, the performance appraisal process under the demonstration project consisted of three phases—development of a performance plan, interim reviews, and a final appraisal at the end of the appraisal period, usually 1 year.

The performance plan establishes the expectations of employee performance during the upcoming appraisal period. The expectations are written as objectives that employees are to accomplish. The objectives could be specific tasks related to the project to which the employee has been assigned (e.g., analyze the Soviet threat to the United States' fleet in the Indian Ocean) or ongoing duties and responsibilities associated with the employee's position (e.g., planning work, supervising staff, and evaluating work progress). In some cases, an employee's plan includes both types of objectives. At least one interim review is required during the appraisal period. During the interim review, employees and their supervisors discuss how the employee is performing in relation to the performance plan and whether changes to the plan are needed.

At the end of the rating period, the employee's performance is evaluated in relation to the objectives contained in the plan. Although NOSC and NWC used different narrative descriptors for the different performance levels, each had five levels—a level equivalent to fully successful, two levels above fully successful, and two levels below. The supervisor recommends that the employee be rated at one of these levels, and the recommendation is reviewed and discussed with higher level managers. Although the employee and his/her supervisor have discussed the employee's performance, a final rating is not provided until the managers have discussed and agreed on the employee's appropriate performance level.

If employees disagree with final performance ratings, they may request that the ratings be reconsidered. An official at a level higher than the manager who approved the final rating reviews the facts and decides whether the rating should be changed.

Pay

Because the pay bands incorporate at least two GS grades, adopting this system gives NOSC and NWC more discretion in determining the starting

salary to be offered to new employees than previously existed. Subsequently, employees' pay levels are adjusted annually on the basis of performance and can include a salary increase within the same band, a one-time bonus or performance award, or a combination of both. Also, employees can receive pay increases through promotions to a higher band. In addition to performance-based pay increases and bonuses, employees in the two laboratories are also eligible for the same general pay adjustments (comparability) granted to employees under the GS system. Employees with ratings of fully successful or higher receive a full general increase, while employees with below fully successful ratings receive either one-half or none of the general increase.

The Project Is Workable but Its Outcomes Are Unclear

The Navy's demonstration project showed that a pay-for-performance system with revised personnel processes to classify, appraise, and pay federal employees can be implemented to the general satisfaction of many managers and employees. The project also demonstrated that line managers could be given authority and responsibility for making personnel decisions—a factor the Navy considered to be essential in implementing the revised system.

OPM summarized its views on the project's success in a February 1986 report entitled A Summary Assessment of the Navy Demonstration Project. The OPM report acknowledged that the project's original goal—improved laboratory effectiveness—could not be demonstrated primarily because of the inherent difficulty in defining and measuring such effectiveness. Instead, OPM maintained that the true measure of the project's success was that it showed that the revised personnel management procedures could be implemented and that these procedures were superior to the procedures NOSC and NWC used before the project was established. According to OPM, the use of simplified classification, the establishment of a more direct link between performance and pay, and increased managerial control over personnel functions would help NOSC and NWC to achieve improvements in

- managerial flexibility over workload assignments,
- recruiting employees, and
- retention of quality employees.

To test the effects of the new personnel processes, a nonequivalent control group evaluation design was used. The objective of this design was to determine project impacts by making before-and-after comparisons

between sites participating in the demonstration project and similar non-participating sites. If demonstration and nonequivalent control groups have sufficiently similar characteristics and the required data are collected, this is one of the stronger evaluation designs for ruling out external events as explanations of study findings.

It is apparent from our analysis of OPM's design that it put considerable effort into conceptualizing the implementation of a methodologically rigorous impact evaluation. Strengths of the evaluation design for the Navy project were (1) specification of the objectives of the project, (2) recognition of the importance of comparing project sites to sites that did not participate in the project, and (3) specification of empirical measures of outcome and data sources. It would be useful to others who evaluate alternative personnel systems to review the Navy demonstration design in these areas and build on work already done.

The evaluation also had certain weaknesses. These flaws included significant differences between the demonstration and control laboratories and substantial missing data. With respect to the former, demonstration and control sites were located in different geographic regions, operated in different economic and organizational environments, and were differentially affected by other major changes in personnel practices. The sites were subject to such different external influences that it is unlikely that any amount of statistical manipulation could fully correct the problem.

With respect to missing data, problems with data availability prevented OPM from obtaining all the data called for in the evaluation plan. Missing data included absent baseline measures, incomplete project data, and limited responses to annual employee surveys.

Given the magnitude of missing data and the differences between demonstration and control laboratories, we cannot assess whether the project benefits cited by OPM related to its major outcome objectives are attributable to the change in personnel practices, to pre-existing differences between laboratories, or to outside factors. The following sections discuss the benefits reported by OPM and illustrate the data problems encountered.

Managerial Flexibility Over Workload Assignments

The project's classification system was designed so that line managers at the demonstration laboratories would have greater authority to make classification decisions and spend less time justifying classification actions. As a result, line managers were expected to have greater flexibility in assigning work to their employees. OPM's evaluation model identified various measures and data sources, such as managers' time involved in classification activities and supervisors' perceptions, which would indicate whether managerial effectiveness had increased.

OPM's evaluation report on classification noted several positive aspects of the revised classification system that appeared to be the basis for OPM's conclusion about the project's success. OPM noted that (1) supervisors spent much less time on classification matters, (2) supervisors perceived themselves as having increased ability to influence classification decisions, and (3) staff spent less time processing classification actions. In addition, OPM identified cost savings resulting from simplified classification procedures at NOSC and NWC that, in 1984, ranged from approximately \$770,000 to \$870,000 at each of the two laboratories. However, OPM noted that the savings identified were not necessarily concrete amounts that could be subtracted from total laboratory expenditures. In commenting further, OPM pointed out that personnel resources spent on such functions as classification were redirected to other personnel management activities.

OPM observed that little evidence existed to support a conclusion that managerial flexibility to assign work had been enhanced by the project. According to OPM, almost no difference existed between responses from demonstration and control laboratory managers when asked if they were subject to controls over the way they assigned work. Further, supervisors at the demonstration laboratories did not perceive they had additional flexibility to assign work to employees. However, OPM stated that information existed indicating that these perceptions were influenced by such outside constraints as funding limitations and demands from the organizations for which NOSC and NWC did work, rather than the project's classification system.

Recruitment of Employees

The broad pay bands established under the project were designed to allow the laboratories to set higher starting salaries than those allowed by the GS system so that NOSC's and NWC's ability to recruit high-quality employees would be improved. As part of its evaluation plan, OPM established measures that were intended to demonstrate whether recruitment of quality employees had improved under the project. These measures

included comparisons between the demonstration and control laboratories to assess (1) ratio of acceptances to offers; and (2) attributes of new hires, including education levels, performance ratings, and grade point averages (GPA).

We found only limited information in the OPM reports to measure the occurrence of recruitment successes. OPM acknowledged in one of its reports that more complete and consistent measures of employee recruitment would have permitted more extensive analyses to be done concerning the project's results in enhancing recruitment. Also, OPM mentioned that limited GPA information was available to assess the quality of new scientists and engineers.

Because we could not obtain recruitment data from the OPM reports, we obtained information from NOSC and NWC personnel officials on new junior professionals that included (1) the ratio of acceptances to offers and (2) new employees' GPAs. The data indicated that some changes had occurred at NOSC and NWC in these two areas. For example, in 1985, the GPAs of newly hired junior professional employees were slightly higher than they had been in 1980. At NOSC, the GPAs of new professionals had changed from an average of 3.52 to 3.53. At NWC, the average GPA of these new employees had increased from 3.14 to 3.20. However, NOSC and NWC officials told us that they believed that recruiting results could also have been influenced by managerial decisions that were unrelated to the project. For instance, NOSC personnel managers said that from 1980 to 1985, they concentrated on attracting graduates with relatively high GPAs and were less concerned about recruiting the numbers of employees they required. In that regard, the ratio of the number of persons hired to the number required declined at NOSC from 75 percent to 68 percent during that time period. At NWC, personnel managers expanded the level of their recruiting efforts in 1979 before the project was implemented. Based on this information, we were unable to discern the effect of the project on the recruitment of quality employees.

Retention of Quality Employees

One objective of the project's performance-based compensation system was to encourage the retention of quality employees. Specifically, NOSC and NWC expected that the greater salary increases and bonuses available to employees with performance ratings above the fully successful level would make it more likely that they would stay at NOSC and NWC than employees rated below fully successful who did not receive similar increases and bonuses.

**Appendix I
Findings and Observations on the Navy's
Personnel Management
Demonstration Project**

OPM's evaluation model called for determining how the project affected employee turnover by using such measures as (1) the number of employees who left NOSC and NWC for other federal and nonfederal positions, as well as those who left due to illness, retirement, or death; (2) a comparison between the demonstration and control laboratories of actual employee turnover by performance rating; and (3) employee attitude data on such matters as their level of satisfaction with jobs, organizational climate, and supervision.

Although OPM included some turnover information in its summary report, OPM acknowledged that more complete and consistent measures of employee turnover would have allowed more extensive analyses of the project's impact on employee retention. For example, the report stated that turnover was about 30 percent lower at the demonstration laboratories than the control laboratories. However, this information was only reported for 1984 and 1985.

To further explore the turnover issue, OPM reviewed employees' intent to turnover. In one of its reports, OPM stated that because actual turnover and intent to turnover were closely related, an analysis of intent to turnover would be meaningful. To address intent to turnover, annual surveys were sent to all supervisors and a sample of employees at the four laboratories. Among other things, the surveys sought information on job satisfaction, a factor OPM considered to be the strongest predictor of intent to turnover. Based on employee responses to the attitude surveys, OPM found that employees felt the project's effect on job satisfaction was generally favorable in matters such as pay equity and pay-for-performance, and that intent to turnover decreased over time at the demonstration laboratories.

However, on average, about 2,000 employees, or 61 percent of the survey sample, responded between 1979 and 1984. Because it is not known if and how the attitudes of nonrespondents differed from those of the respondents, the applicability of survey results to the overall work force at the four laboratories is unknown. Furthermore, OPM found that job satisfaction was affected less by pay and more by the work itself and the amount of control employees exercised over the work—issues that the project did not address.

During the course of our review, NOSC and NWC personnel officials provided us statistical data on employee turnover rates from 1982 to 1985 that they believed indicated that the expected retention improvements had been attained. The data showed that the percentage of employees

who received ratings below the fully successful level and subsequently left was greater by about 22 percent at NOSC and by 47 percent at NWC than the percentage of employees rated above fully successful who subsequently left. While we believed these percentages were impressive, we could not determine from available data whether turnover rates were any different from those that existed before the project.

One area with missing data involved performance ratings. The four Navy laboratories were supposed to send computer tapes to OPM annually containing performance rating data. However, for 1979 and 1980, no such data were obtained from the control laboratories and only partial data were obtained from the demonstration laboratories. In later years, significant gaps in ratings data persisted. Without the requisite data, it is very difficult to demonstrate a causal linkage between performance ratings and outcomes, such as salary, work attitudes, and turnover.

An analysis of this information was also affected by the project's establishment of five performance levels in the NOSC and NWC performance appraisal systems. Before the project, three performance levels were used—outstanding, satisfactory, and unacceptable. Thus, it was not possible to determine how many employees who were rated as satisfactory performers under the previous three-level system would have been placed in the levels directly above and below satisfactory in the new five-level system.

According to an OPM official, OPM is currently preparing a report that will further address the issue of turnover.

Factors to Consider in Future Demonstration Projects

We recognize that optimal conditions for scientific study of complex programs rarely exist. Those evaluating the Navy demonstration project encountered the difficulties of trying to do a controlled study in a dynamic environment. While no evaluation can definitively discern all program effects, each one affords an opportunity to identify areas that will play major roles in future projects and to learn how future studies may be improved.

Considering the above, Congress, OPM, and federal agencies need to be alert to several factors when considering the design, implementation, and evaluation of future alternative personnel systems. The following sections discuss these factors.

A Credible Performance Appraisal System Is Helpful

One factor that can enhance a personnel project's chances of success is a credible performance appraisal system. Elements of such a system include clear performance expectations, ratings that can discriminate between different levels of performance, and feedback on the extent that expectations have been achieved.

In its summary report, OPM noted that in a review of a sample of demonstration employees' performance plans, certain problems—such as unclear standards and the lack of a consistent formula for deriving summary ratings from ratings on individual elements—could negatively affect employees' perceptions about the fairness of performance appraisals. However, OPM reported that these shortcomings did not appear to adversely affect employees' perceptions.

According to OPM, employees at the demonstration laboratories perceived the performance appraisal system to be fair and objective. Demonstration laboratory employees also expressed satisfaction with better communication between supervisors and employees, clearer performance expectations, and the appropriateness of actions taken toward high performers as well as low performers.

Funding Level Is Important

In its evaluation, OPM reviewed salary costs and found that as of January 1986, employees at the demonstration laboratories were paid salaries that were 6 percent higher than those received by employees at the control laboratories. OPM estimated that this increase amounted to an additional \$15 million in salary costs. OPM attributed this increase primarily to (1) the higher starting salaries paid to entry-level scientists and engineers at the demonstration laboratories and (2) the larger average salary increases given to employees both within a pay band and through promotions between the bands at nearly all levels. OPM estimated that increased salary costs have occurred at NOSC and NWC at the rate of approximately 1 percent each year since the project was established in July 1980. OPM also stated that barring any changes in policy at the demonstration laboratories, salary costs under the project could be expected to continue to increase.

OPM stated that the increased costs at NOSC and NWC were attributable not only to the project but also to an additional monetary investment made by the two laboratories. This investment included such items as payments to employees at the demonstration laboratories for (1) within-grade increases that would have been due within a specified period under the old system and (2) increased salaries for existing employees to

bring their salaries in line with those that were paid to new employees under the project. OPM acknowledged that if this investment had not been made, the results of the project could have been less favorable.

The proposed Civil Service Simplification Act (S. 1545), which authorizes the implementation of the Navy's project in other federal agencies, and the proposed Defense Industry and Technology Act both impose a new requirement: budget neutrality. At present, however, little information is available on the feasibility or likely impacts of implementing an alternative pay system in a budget-neutral manner. For example, the project at the National Bureau of Standards was designed to be budget-neutral; however, that project has only recently been implemented.

Some budget-neutral proposals provide for applying anticipated productivity savings to reduce the costs of new personnel systems. These proposals impose a significant level of risk for two reasons. First, it is unclear whether any budgetary savings will be achieved, and the project could adversely affect employees if such savings were relied upon to fund the new system. Generally, we believe it would be difficult to implement reasonable pay system alternatives that are budget neutral if the goal is to help improve recruitment and retention, as is the case in S. 2254. Second, although good productivity measures and accurate and reliable data on performance can be developed for some operations, this is often not the case for many governmental activities. For example, National Bureau of Standards officials said they do not plan to evaluate the effect of their project on productivity because they believe it would not be possible to measure it.

Whether and how budget neutrality affects the outcomes of new personnel programs is an empirically testable question. If budget neutrality is to be an objective of future demonstration projects, the range of its intended and unintended effects should be carefully tested, monitored, and documented along with projects that are not designed to be budget neutral.

Transferability of Successful Projects Is a Function of Many Factors

It would be tenuous to use the results of the Navy's project as a basis for full implementation of the project in other organizations even if (1) optimal evaluation conditions had existed; (2) the data were entirely reliable, valid, and complete; and (3) the findings were conclusive. The fact that a particular program is found to be successful does not mean that its findings can be generalized across different people, settings, and times.

OPM has pointed out that various factors would influence results obtained in other organizations, including

- the existence of outside factors, including labor market competition and employee mobility;
- agencies' implementation of the project's features in a different manner from the way they were implemented at NOSC and NWC;
- employees' understanding of and attitudes toward the new system; and
- the amount of money agencies would be willing to invest to establish the project.

We agree with OPM and believe that the demonstration project would need to be replicated elsewhere to ascertain its applicability to other settings.

Evaluation Considerations

As a single study of a particular form of personnel management, it would not have been reasonable to expect the evaluation of the Navy project to have answered all questions.

The evaluation, however, represents an important step to building a knowledge base on systemic methods of improving personnel management and performance. Through repeated small-scale tests of this sort, the weaknesses of individual studies will become less prominent and applicability of findings to other settings would be enhanced. The most sound basis for formulating better personnel management policy decisions will come from accumulating evidence across studies done on different work force populations in different settings at different times.

In addition, data collection efforts should be monitored, studies should adequately respond to changing conditions, and the components of the program should be documented. Practical, procedural, and administrative problems often arise during the course of an evaluation that impede collection of the required data. To preserve the integrity of the study design, it is necessary to have a feedback mechanism signalling the existence of these problems so that evaluators can take compensating actions to get the data collection back on track. Therefore, active monitoring of data collection should be an ongoing component of future evaluations.

Full-back options should also be available if critical data cannot be obtained. The original evaluation plan should include alternative measures that could be used if anticipated data do not materialize. If alternative measures are not feasible, then provisions should be available for refocusing the study so that meaningful information can still be obtained. Adapting the evaluation to changing circumstances can help strengthen the conclusions reached.

To understand how a successful program might be replicated, as well as why a program might have failed to achieve its objectives, data on outcomes are not enough. In addition, the evaluators should document the specific components of the program (e.g., what new recruitment efforts are to take place) and monitor the extent to which the program was implemented (e.g., number of staff days spent on the new activity). Monitoring at both demonstration and control sites also facilitates awareness of outside factors (e.g., economic conditions) that can affect program results. The collection of qualitative process information can help identify strengths and weaknesses of the project. This would enable critical linkages to be made between program inputs and outputs and provide a substantive basis for improving and fine-tuning future demonstration projects.

Letter to the Chairman, Senate Committee on Armed Services, Answering Six Questions on Senate Bill 2254

GAO

United States
General Accounting Office
Washington, D.C. 20548

General Government Division

April 19, 1988

The Honorable Sam Nunn
Chairman, Committee on Armed Services
United States Senate

Dear Mr. Chairman:

Pursuant to a request from Senator Glenn, enclosed are our answers to six questions on a proposed new personnel system for Department of Defense acquisition personnel. Please call me on 275-4232 if you need further clarification.

Sincerely yours,

Bernard L. Ungar
Bernard L. Ungar
Associate Director

Enclosure

RESPONSES TO QUESTIONS FROM SENATOR GLENN

Question 1

You stated you have concerns with the component aspect of the personnel system proposal. If the proposal covers entire organizations within the Department of Defense (DOD) so that equal treatment would be given to all DOD components, would GAO support its passage?

GAO Response

We would not support its passage if the only change to the proposal is to cover entire organizations within DOD. We are concerned about two other elements of the proposal in addition to the organizational versus occupational issue. These other elements are the budget-neutral nature of the proposal and the size of the work force that would be covered--up to 300,000 DOD employees. Our response to question 2 further explains these concerns.

If the Congress decides to authorize an alternative personnel system for DOD or other agencies on a permanent rather than a demonstration or test basis, GAO would favor its implementation for all white-collar employees of an organization rather than for certain occupations. This would avoid the administrative problems that could result from multiple, occupationally-based personnel systems and assure uniform treatment for all employees.

We believe that section 1643(a) of S.2254 is unclear as to whether the proposed personnel system would cover entire DOD components or certain positions within components. The term "component" in this section is open to interpretation. It could be interpreted as an entire organization involved in acquisition and logistics, such as one of the service's inventory control points or a systems command. On the other hand, it could be interpreted as those units within an organization that are directly involved in acquisition and logistics.

If the Subcommittee decides to take further action on the bill, we suggest that section 1643(a) be revised as follows (suggested clarifying language is in bold type and suggested deletions are in brackets):

Section 1643(a) In General - The Secretary, with the approval of the Director of the Office of Personnel Management, may establish and implement an alternative personnel management system [for the] **to cover all** professional, administrative, and technical positions of those [components] **organizations** of the Department of Defense that, as determined by the Secretary, primarily perform acquisition and logistics functions.

Appendix II
Letter to the Chairman, Senate Committee on
Armed Services, Answering Six Questions on
Senate Bill 2254

Question 2

In your prepared statement, you expressed concern about the budget-neutral nature of the proposal. In your remarks, you stated that implementing the pay band plan could be difficult on a budget-neutral basis; however, it could be done with adequate testing and evaluation. Could you expand on that last point? Are you saying that, with testing and evaluation, GAO would support the budget-neutral aspect of the proposal?

GAO Response

We would not support adopting or testing a budget-neutral alternative pay system involving as many employees as would be allowed under the bill at this time. As many as 300,000 DOD employees could be included in the system proposed by S.2254. Generally, we believe it is very difficult to implement reasonable pay system alternatives that are budget neutral.

The budget-neutral feature of the proposal has not yet been sufficiently tested to determine how it would work or its results. The Navy's demonstration project at China Lake and San Diego was not designed to be budget neutral. According to OPM, as of January 1986, overall salary costs for this project were 6 percent higher than at counterpart laboratories used as control facilities. Further, little information is available on the results of the two other federal alternative pay projects--the National Bureau of Standards and McClellan Air Force Base--which were only recently implemented and involve a much smaller number of employees than contemplated under S.2254. (See our response to question 5.)

With so little information available on the results of a budget-neutral system, we do not believe it would be advisable to place up to 300,000 Defense employees in such a system, particularly if it turns out to have undesirable results. If the Subcommittee wishes to further test a budget-neutral alternative pay system, we suggest it be done on a much smaller scale involving a smaller number of employees and that it be very carefully planned, monitored, and evaluated.

Question 3

You made a point in your remarks that DOD needs to have an adequate test and evaluation plan to insure the effective implementation of the personnel proposal. What would constitute an adequate test and evaluation plan? Does the China Lake demonstration project provide a sufficient test for many aspects of the DOD proposal? If not, why not?

GAO Response

Elements of an adequate test and evaluation plan

The evaluation plan should act as a blueprint for how to carry out the study so that proper conclusions can be drawn at the end. It should provide clear guidance on the purposes of the program, the study design, the information to be collected and how data collection will be done, data analysis procedures, and the management plan. Major components of an adequate plan to test and evaluate program impacts include the following:

-- Specification of the objectives of the program.

Both short- and long-term objectives should be specified clearly and in measurable terms. Development of the objectives should include consideration of what - in practical terms - would constitute successful outcomes. For example, how much increase in the retention of high performers is necessary to conclude that the program is successful in retaining the "best and brightest"?

-- Specification of the study design.

The role of the study design is to insure that program variables, rather than unrelated factors, are responsible for producing the observed results. One major aspect of design is the identification of an appropriate comparison group to test program results against. The comparison group should be as similar to the experimental group as possible, with the exception of the new program. If the experimental group is not equivalent to the comparison group (e.g., there are differences in age which may affect turnover rates), then data should be collected on those initial differences that might affect outcomes. Statistical analyses should use these data to try to correct for the pre-existing differences between experimental and control groups.

-- Development of empirical measures of outcome.

The extent to which the objectives of the new system have been met should be measurable using reliable and valid

Appendix II
Letter to the Chairman, Senate Committee on
Armed Services, Answering Six Questions on
Senate Bill 2254

measures of outcome. Since program outcomes can rarely be measured perfectly or completely, the evaluation plan should include multiple measures of the same outcome. (For example, quality of new hires might be measured using grade point average, first year performance ratings, and end of first-year interviews with supervisors.) The weight of evidence from multiple measures can significantly enhance the interpretability of findings.

-- Identification and verification of data sources.

The evaluation plan should specify the data sources for the study and provide for quality checks of those data. It should not be left to chance that after the study begins, anticipated sources of data do not materialize. If an existing data source is found to have unreliable or incomplete data, the evaluation plan should either call for supplementing the data source with better data or for creating a primary data base specific to the study. Active monitoring of the data collection - at both experimental and control sites - is then necessary to ensure high data quality.

-- Specification and monitoring of activities at both experimental and comparison sites.

To understand how a successful program might be replicated, as well as why a program might have failed to achieve its objectives, the same types of information are necessary: documentation of the specific components of the program (e.g., what new recruitment efforts are to take place) and monitoring of the extent to which the program was implemented (e.g., number of staff days spent on the new activity). Monitoring at both experimental and control sites also facilitates awareness of outside factors (e.g., economic conditions) that can affect program results. The evaluation plan should include collection of such process information because it adds richness and depth to the study and enables critical linkages to be made between program inputs and outputs.

-- Development of a data analysis plan.

The purpose of an analysis plan is to set out how data will be used once collected. It should specify what analytical techniques will be used and how analysis results will be presented. It can streamline data collection by reducing tendencies to collect unnecessary information, and it helps keep the focus on the final product of the evaluation study.

**Appendix II
Letter to the Chairman, Senate Committee on
Armed Services, Answering Six Questions on
Senate Bill 2254**

-- Specification of roles, responsibilities, and tasks involved in performing the study.

A detailed breakdown of evaluation tasks should be set out along with cognizant personnel responsible for carrying out those tasks. This will minimize potential for confusion over the order in which tasks are to be carried out and who is to do what.

Every effort should be made to develop detailed information in the evaluation plan on the methodological components cited above. The evaluators should strive to follow the plan as closely as possible. Because social programs take place in dynamic environments, however, it can be extremely difficult for an evaluation study to fully carry out all aspects of the methodology as planned. Evaluators should recognize that new events and external influences might arise which could thwart plans for a "pure" study. In such a case, fall-back options should be considered and provisions made for adjusting the study so that it is sensitive to the new conditions.

Applicability of the China Lake evaluation to the DOD proposal

Our information on China Lake comes from reviewing relevant regulations and guidance, the project proposal, the original evaluation plan, the ten evaluation reports issued between 1984 and 1987, interviews with OPM and Navy officials, and site visits to the project's experimental and control sites. We began our review seven years after the start of the demonstration project and were faced with all the limitations associated with conducting such a retrospective inquiry. These limitations notwithstanding, the accumulated evidence indicates that the data from the China Lake evaluation are not strong enough to determine whether or not the project met its objectives. If the full complement of data identified in the evaluation plan had been collected as planned, there could have been a stronger basis for drawing conclusions about effects.

Based on the available information, and using the evaluation framework cited above, we are able to make the following statements about the test of the China Lake demonstration project which may be helpful to potential tests of the proposed DOD system:

The strengths of the China Lake evaluation were its specification of the objectives of the project, its awareness of the importance of comparing project sites to sites that did not participate in the project, and its specification of empirical measures and data sources. In these areas, it is apparent that considerable effort was put into conceptualizing the implementation of a rigorous impact evaluation. It would behoove others who evaluate alternative personnel systems to review the China Lake

**Appendix II
Letter to the Chairman, Senate Committee on
Armed Services, Answering Six Questions on
Senate Bill 2254**

study in these areas and build on the work already performed.

The overall weakness of the China Lake evaluation was that when all was said and done, the volume of data that were either missing or non-comparable was quite large. While we do not know all the reasons behind the data problems, we can determine that they were of such magnitude that firm conclusions about project effectiveness cannot be drawn.

One major problem originated from the non-comparability of demonstration and control sites. They were located in different geographical regions, operated in different economic and organizational environments, and were differentially affected by other major changes in personnel management. On the one hand, given the decision to implement the demonstration project laboratory-wide, the evaluators appear to have done their best in selecting control sites. On the other hand, the differences between sites were so great that no amount of statistical manipulation could correct the problem. Officials should consider implementing future demonstration projects in units within an organization, rather than organization-wide. While this might result in some added administrative burden, it would afford a stronger test of program impacts by enabling the selection of a highly similar comparison group from within the same organization.

Another major problem resulted from data that were supposed to be collected but were not. For example, the two demonstration labs and the two control labs were supposed to send computer tapes to OPM annually containing performance rating data. However, for 1979 and 1980, no such data were obtained from the control labs and only partial data were obtained from the demonstration labs. In subsequent years, significant data gaps persisted. It is obvious that failure to obtain such data impairs before- and after comparisons and clouds the ability to interpret the findings.

Finally, the available evaluation reports on China Lake provide little information on how and to what extent the project was implemented at the demonstration sites. Qualitative information on implementation processes could have identified strengths and weaknesses of the project which, in turn, could have been used to improve and fine-tune future demonstration projects.

We share the concern about the quality of the work force and support tests of specific reforms targeted to alleviating the problem. We caution against advocating specific reforms as though they were certain to succeed, however. We cannot know beforehand the range of direct and indirect outcomes from a specific complex reform. The

sufficiently strong to draw unequivocal cause-effect conclusions about the effectiveness of its personnel demonstration project. Our commitment should be to testing (either sequentially or in combination) several possible solutions to the work force quality problem, rather than to a specific reform.

Question 4

I understand GAO recently completed a review of the China Lake demonstration project. What lessons did we learn from that demonstration and are any of those lessons particularly relevant to the DOD proposal? What were China Lake's accomplishments?

GAO Response

The goal of the alternative personnel system proposed in the bill is to give DOD acquisition managers more flexibility in managing and rewarding their work forces, thereby resolving the recruitment and retention problems evident in DOD acquisition organizations. DOD would be able to apply the alternative personnel management system developed at the Navy laboratories at China Lake and San Diego to DOD acquisition organizations.

We are now completing a review of the Navy's personnel management demonstration project for the Subcommittee on Federal Services, Post Office, and Civil Service, Senate Committee on Governmental Affairs. Several observations we have made during this review would be relevant to the DOD proposal.

- First, the Navy's demonstration project was not implemented in a budget-neutral manner. According to OPM, as of January 1986, overall salary costs for the Navy's demonstration project are 6 percent higher than at counterpart laboratories used as control facilities. Thus, accomplishments associated with the Navy's project were not achieved under a budget neutral system.
- Second, the demonstration project at the two Navy laboratories in China Lake and San Diego, California, showed that a pay-for-performance system with such process features as broad pay bands, simpler classification procedures, and closer linkages between performance and pay can be implemented to the general satisfaction of many managers and employees. The project also showed that line managers could be given direct authority and responsibility for making pay and reward decisions for their employees, which was considered an essential factor in effectively implementing the revised system.
- Third, although we did not identify significant deficiencies with the Navy's project and OPM believes it to be successful, problems with data availability and usefulness preclude a

**Appendix II
Letter to the Chairman, Senate Committee on
Armed Services, Answering Six Questions on
Senate Bill 2254**

conclusive assessment of the project's results relative to the outcome objectives established. These objectives, which are generally similar to the goals set forth for DOD acquisition personnel, include improved recruitment and retention.

- Fourth, according to OPM, managers at China Lake and San Diego perceived that they did not have greater flexibility to assign work to employees, which was one of the objectives of the revised classification procedures. However, OPM found that managers felt they (1) had sufficient authority to influence classification decisions; (2) encountered much less difficulty in getting positions classified; and (3) generally spent less time on classification matters due to the establishment of simpler, less burdensome classification procedures. In addition, the project's broad pay bands provided more latitude in making pay distinctions which OPM considered an important aspect of any performance-based compensation system.

Question 5

Please describe other budget-neutral alternative pay projects implemented by the federal government. What has been the size and scope of these projects? What is the precedent for implementing a budget-neutral pay band project for a work force of 300,000 employees?

GAO Response

In January 1988, the National Bureau of Standards (NBS) implemented a personnel demonstration project which included all of NBS' 3,050 white-collar employees. Although not required to do so by law, NBS has set an objective to keep the project budget neutral throughout the 5-year demonstration. To meet this objective, NBS officials said they expect to pay for the higher salaries anticipated to result from the project by employing fewer or lower-graded personnel. In addition, in early 1988, the Air Force implemented a personnel demonstration project at McClellan Air Force Base on a budget-neutral basis involving about 2,000 white- and blue-collar employees. The project includes a gainsharing program whereby the Air Force will identify gains in employee productivity and share the savings generated between employees and the government.

We are unaware of any budget-neutral pay band projects that cover a work force of 300,000.

Question 6

There are two aspects of the personnel proposal that your prepared statement was silent on about which I would like your opinion. One deals with the use of documented savings from improved productivity to reduce the aggregate costs of the new system, and the other is providing for special pay for certain scientific and technical positions. Both of these were not included as part of the China Lake demonstration project so we have no experience on which to determine the viability of either proposal. While the benefits may be apparent, given GAO's work on a broad range of issues, can you identify problems with either of the proposals that the committee should be aware of as it decides on the bill's passage?

GAO Response

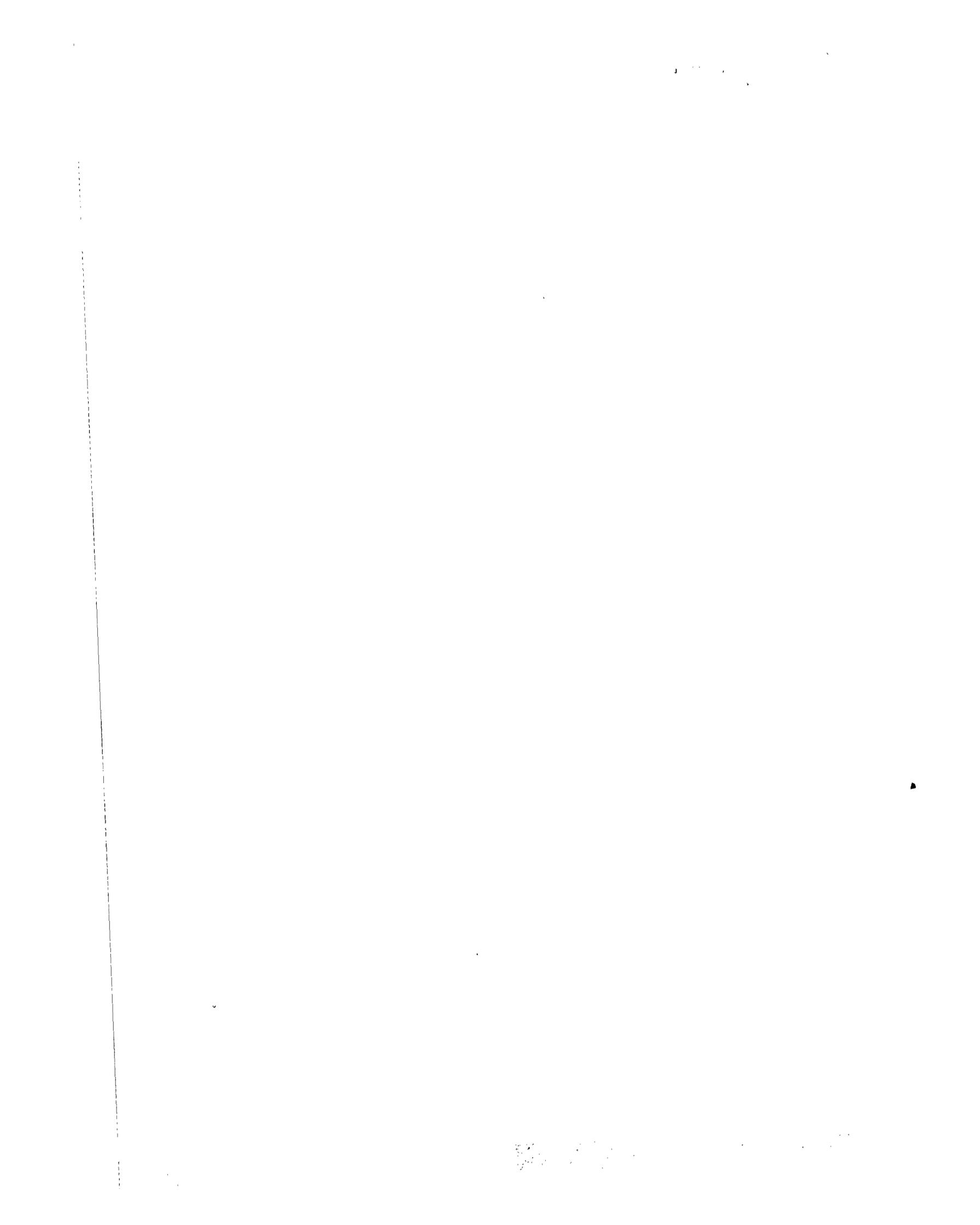
We have some concerns about the proposal to use documented savings from improved productivity to reduce the aggregate costs of the new personnel system. First, our work over the last several years has shown that agencies often have difficulty developing valid, agreed-upon productivity measures. This lack of good output measures for many operations, along with shortcomings in many agencies' management information systems, raises questions about how reliable such a system would be.

The situations at the Navy's China Lake and San Diego project and at the National Bureau of Standards also help shed some light on this issue. OPM reported that the two Navy laboratories realized some savings in administrative areas, such as reductions in personnelists' time spent on classification activities. However, a comprehensive evaluation of productivity savings resulting from the project at China Lake and San Diego was apparently not done. NBS officials said they do not plan to evaluate the effect of the project on productivity because they believe it would not be possible to measure it.

We have no objection to using special pay rates for certain scientific and technical positions, where warranted. Our work has shown that the use of special pay rates for certain occupations has been effective in making the government's compensation package more competitive in attracting and retaining people in hard-to-fill positions.

List of OPM Reports Evaluating the Navy Demonstration Project

Report title	Report issue date
Status of the Evaluation of the Navy Personnel Management Demonstration Project (Management Report I)	March 1984
Evaluation of the Navy Personnel Management Demonstration Project: Analysis of Survey and Interview Results, 1979 to 1983 (Management Report II)	July 1984
Effects of Restructured Compensation System on Salaries of Scientists, Engineers and Other Professionals, 1980 to 1984 (Management Report III)	Jan. 1985
Navy Personnel Management Demonstration Project: The Effects of Performance-Based Pay on Employee Attitudes (Management Report IV)	June 1985
Effects of Restructured Compensation System on Salaries of Technicians, 1981 to 1984 (Management Report V)	June 1985
Evaluation of the Navy Personnel Management Demonstration Project: Demonstration Classification Systems (Management Report VI)	Sept. 1985
An Analysis of Tradeoffs in the Navy Personnel Management Demonstration Project (Management Report VII)	Nov. 1985
Evaluation of the Navy Personnel Management Demonstration Project: Performance-Based Pay Systems (Management Report VIII)	May 1986
A Summary Assessment of the Navy Demonstration Project (Management Report IX)	Feb. 1986
Salary Costs and Performance-Based Pay Under the Navy Personnel Management Demonstration Project: 1986 Update (Management Report X)	Dec. 1987



Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are \$2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use \$300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100