United States General Accounting Office

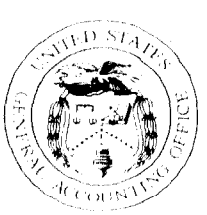
GAO

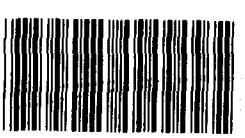
Report to Congressional Committees

May 1991

FEDERAL PERSONNEL

Review of Evaluation of Personnel Demonstration Project at Commerce





GAO/GGD-91-93

GAO

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

General Government Division

B-240429

May 14, 1991

The Honorable John Glenn
Chairman, Committee on Governmental Affairs
United States Senate

The Honorable Ernest F. Hollings
Chairman, Committee on Commerce, Science and Transportation
United States Senate

The Honorable William L. Clay
Chairman, Committee on Post Office and Civil Service
House of Representatives

The Honorable George E. Brown, Jr.
Chairman, Committee on Science, Space and Technology
House of Representatives

Public Law 99-574 required us to report to you on the personnel management demonstration project at the Commerce Department's National Institute of Standards and Technology (NIST). It also required the Office of Personnel Management (OPM) to hire a qualified evaluator to determine whether project objectives were achieved. Congress instituted the project to enhance NIST's ability to hire and retain capable employees, especially scientists and engineers.

This report, which responds to the statutory requirement, provides our assessment of the evaluation's design and implementation during 1988 and 1989, the evaluation's first 2 years. The evaluator for those years was University Research Corporation (URC) of Bethesda, Md.; the current evaluator is HumRRO International, Inc., (HII) of Alexandria, Va.

Results in Brief

The evaluation of the NIST project's first 2 years was not sound. Although URC proposed a relatively strong research design for its study, its implementation of that design was flawed. For example, URC selected certain Commerce units as comparison sites but failed to demonstrate in a convincing way that these sites were indeed suitable comparisons to NIST. Additionally, no priority was given to collecting and analyzing information about nonproject changes—mission, organizational, and

management—occurring at NIST that may affect project results. Such shortcomings, we believe, created serious problems for identifying and understanding the project's effects.

We also believe OPM's funding of the evaluation, which was $176,000 for the first year and $191,000 for the second year, was less than required for a full-fledged evaluation of a complex project. On the basis of our experience with other evaluations, we believe funding was insufficient to permit adequate staffing of the evaluation.

We discussed our concerns with OPM officials, and they acted on them. For example, OPM has required HII, which began in late summer 1990 to evaluate the project's third year, to examine the suitability of the comparison group further and to take into account nonproject changes in interpreting project data. OPM also significantly increased evaluation funding; funding for the evaluation's third year is 93 percent greater than the second year.

We cannot precisely forecast how successful OPM and HII will be in addressing our concerns or what other problems they may encounter. We believe, however, that the chances for a more methodologically sound and informative evaluation have been enhanced by OPM's actions to improve the evaluation and increase its funding.

Background

NIST provides scientific and technological services to industry and government. The demonstration project covers about 3,000 employees at NIST's two locations: Gaithersburg, Md., and Boulder, Colo. These employees include scientists, engineers, technical support staff, and administrative and clerical personnel. The project does not include wage-grade (blue collar) employees because the act specifically excludes them.

Unlike most other personnel management demonstration projects, Congress specifically mandated the NIST project. It gave NIST authority in Public Law 99-574 to set salaries competitive with those available outside the government and to adjust compensation on the basis of an individual's performance (or merit). In setting salaries, Congress (1) required NIST to annually make compensation comparisons to the private sector and (2) authorized it to make up part or all of any compensation deficiency. Congress provided these authorities as a means of addressing a problem it saw across government in attracting and keeping qualified personnel in high technology fields.

The act required NIST and OPM to jointly design the project and for NIST to carry it out. Using personnel management changes set out by the act and by adding others, NIST and OPM designed the project to

- improve hiring and allow NIST to compete for high-quality researchers more effectively through such means as agency-based and direct hiring of job applicants (instead of hiring through OPM), selective use of higher entry salaries, and selective use of recruiting bonuses;
- motivate and retain staff through such means as higher pay potential, pay-for-performance, and selective use of retention allowances, which employees can receive to remain with NIST rather than accept nonfederal job offers;
- strengthen the managers' role in personnel management through delegation of certain personnel authorities to them;
- increase the efficiency of personnel systems through such means as installation of a simplified job classification system and the reduction of guidelines, steps, and paperwork;
- keep total employee compensation at the level it would have reached under the governmentwide system without the project; and
- serve as a model that can be replicated in whole or in part by other federal agencies.[1]

The evaluation contractor is to determine whether personnel management changes implemented through the project caused the project's objectives to be met. Public Law 99-574 required evaluation results to be reported annually to Congress.

OPM hired URC as the evaluator about the time the project began in January 1988. And, under a series of OPM-issued work orders totaling about $367,000, URC remained in that role for the evaluation's first 2 years. In the summer of 1990, OPM hired HII, on a work order basis, to evaluate the remaining years of the 5-year project.

Objective, Scope, and Methodology

The objective of our review was to assess whether the evaluation was likely to provide data that would be sufficient to judge whether the demonstration project enabled NIST to improve employee hiring and

[1]Major pay reform legislation—the Federal Employees Pay Comparability Act of 1990—has been enacted since authorization of the NIST project. Key features of the November 1990 act include the introduction of a locality pay system in January 1994, the exploration of a pay-for-performance system for all General Schedule employees, and the addition of such employee recruitment and retention tools as recruitment bonuses and retention allowances. According to the Chief of NIST's project office, NIST has begun analyzing the act's provisions to determine how they pertain, or could pertain, to the demonstration project.

retention.[2] Our review was done intermittently between March 1989 and March 1991. During that period, the first 2 years of the evaluation, covering calendar years 1988 and 1989, were completed and the resulting evaluation reports were sent to Congress.[3]

To assess the evaluation, we obtained information on URC's plans for evaluating the project and how those plans were implemented in 1988 and 1989. We discussed the project and evaluation with project managers at OPM, URC, and NIST and reviewed pertinent documents such as URC's plan for managing the evaluation and the two evaluation reports. We also reviewed HII's September 7, 1990, plan for managing the evaluation's final 3 years.

In a May 1988 report on another demonstration project,[4] we outlined what we consider to be the elements of an adequate evaluation plan. We used those same elements, along with other criteria commonly accepted among professional evaluators, to determine the adequacy of the NIST evaluation over its first 2 years.

We followed generally accepted government auditing standards in doing our work, which was done in and around Washington, D.C. We obtained informal comments on the information contained in this report from project officials at NIST, OPM, and URC. Their comments are summarized on page 11.

Evaluation Design and Implementation

To test the effects of the demonstration project, URC planned to use a particular type of research design called "nonequivalent control group" design. The objective of this design is to determine project impacts by making before-and-after comparisons between sites participating in the demonstration project and comparable nonparticipating sites. If demonstration and comparison groups have sufficiently similar characteristics, this is one of the stronger designs for establishing causal linkages between project changes and effects. However, in our view, the intended

[2]Before this review, we issued a fact sheet about the demonstration project: Federal Workforce: Information on the National Bureau of Standards Personnel Demonstration Project (GAO/GGD-88-59FS, Apr. 5, 1988).

[3]The evaluation reports are titled Implementation Report (Aug. 1989) and Second Annual Evaluation Report (Aug. 1990). The evaluation contractor writes the report and, after review by OPM, NIST, and the Department of Commerce, the OPM Director submits it to Congress.

[4]Federal Personnel: Observations on the Navy's Personnel Management Demonstration Project (GAO/GGD-88-79, May 3, 1988).

research design was not properly implemented, creating difficulty in understanding project impacts.

Appropriateness of Comparison Group Not Convincingly Demonstrated

URC selected several Department of Commerce sites in Boulder, Colo., to serve as the comparison group to NIST. URC compared NIST and comparison group workforces by gender, age, years of government service, educational level, and career path.[5] These comparisons led URC to conclude that, from a practical perspective, employee characteristics were sufficiently similar that the Boulder sites would be an appropriate comparison to the NIST sites.

However, there is no indication that examinations were made of other characteristics that might also affect the study's findings. These include employee characteristics (such as grade level, quit rate, promotions, performance ratings); organizational characteristics (such as management structure, work environment, personnel processing times); and geographic characteristics (such as local economic conditions, supply of good job candidates).

The URC project director said many of the characteristics we suggest that could have been examined were addressed in the second year evaluation report. That report included analyses of some of the characteristics we listed. However, those analyses did not compare the characteristics as they existed at NIST and the comparison sites before the demonstration project was implemented. Such comparisons would be necessary to identify how alike and different NIST and the comparison sites were at the start of the project.

It is unreasonable to expect a perfect match between demonstration and comparison groups in the real world context of evaluations. However, preexisting differences that may affect results need to be identified and probed more thoroughly than was done for the NIST evaluation. If the differences can be measured, they should be incorporated into analyses of the data. If the differences cannot be measured, studying them is still worthwhile because they can assist in interpreting how the results may be biased.

[5]Occupations at NIST have been grouped into four career paths—scientific and engineering, scientific and engineering technician, administrative, and support. NIST and comparison site employees were compared by these career paths.

OPM has acknowledged this issue of comparability. It is requiring HII to further investigate the comparability of NIST and the comparison group. HII's evaluation approach continues the before/after, demonstration/comparison group design.

Need for Preproject Data

URC's evaluation design required the gathering of data on NIST and comparison group activities as they stood before the project began. These data are essential to determining how much net change occurred at NIST during the project. They provide a basis—where NIST and the comparison group were before the project—to compare against.

In the second evaluation report, URC concluded that "[t]hrough the use of direct hire and agency-based hiring, NIST has reduced the time required to fill position vacancies." This conclusion was not based on preproject measures. Rather, it was based on averages of the number of days that NIST and the comparison group took to fill vacancies in 1989. However, without preproject measures, there is no way to reliably determine whether NIST's hiring time actually changed. For example, hiring times may have differed between NIST and the comparison sites before the project began. Without consideration of preproject hiring times, we believe the evaluator's conclusion is questionable.

The URC project director said the conclusion is appropriately qualified elsewhere in the second evaluation report and that other information in the report, such as comments made by supervisors during "focus group" discussions, supports the notion that hiring time was being reduced. The second evaluation report said that a thorough examination of changes in the time required by the various hiring methods and the steps within those methods requires more data than were currently available. A statement was also made later in the report that there was "some evidence suggesting" that direct hire and agency based staffing reduced hiring times. These statements notwithstanding, the study does not include analysis of preproject measures against which to accurately and reliably gauge whether hiring times have changed and the significance of any change.

We do not know the extent to which preproject hiring and other data for NIST and the comparison sites are available or can be reasonably reconstructed. However, we believe reaching sound conclusions on many project issues will be hampered considerably, if not made impossible, without reliable measures of what the situations at NIST and the comparison sites were before the demonstration project began.

Questionable Comparability of Employee Surveys

Perceptions and attitudes of employees were being ascertained through periodic surveys at project and comparison sites. Such surveys are intended to track the attitudes of project and comparison group employees and to measure how they change over time.

Through the project's first 2 years, three surveys had been done, including one shortly before the project began. NIST employees were initially surveyed in September-October 1987 in order to obtain a "baseline" measure of attitudes. NIST employees were surveyed a second time in April-May 1989. Comparison group employees were first surveyed in September-October 1988, establishing a baseline for comparison group employees. URC did the 1988 and 1989 surveys; another firm, under a NIST contract, did the 1987 survey.

Because of differences in when the survey data were collected, we had concerns about using the 1988 (baseline comparison) and 1989 (second NIST) surveys for comparison purposes. For example, the comparison group survey, through no fault of URC, was done 1 year after the NIST baseline survey. The comparison sites became part of the evaluation after URC proposed their inclusion in July 1988; OPM had not required a comparison group. In any event, differences between the results of the two baseline surveys may be attributable to actual differences in attitude, to the passage of time, or to a combination of the two. The same is true when comparing the second NIST survey with the comparison group survey. The second NIST survey was done 6 months after the comparison group survey.

We were also concerned about making comparisons between the different surveys because different methods were used to collect the data. The comparison group survey and the second NIST survey were done by mail (written questionnaire); the NIST baseline survey was done by telephone. Because a person may respond differently to the same question if asked by a telephone interviewer versus a written questionnaire, the difference in data collection methods may have contributed to attitudinal differences found when comparing the written surveys to the telephone survey.

Finally, we were concerned about the comparability of the employee surveys because response rates to the mail questionnaires were relatively low—59 percent for the comparison group survey and 57 percent for the second NIST survey. When response rates are low, characteristics of respondents and nonrespondents, such as occupations and pay grade, should be compared to each other. If the profiles of the two groups on

relevant characteristics are sufficiently similar, the respondents' views may be presented as representative of all persons targeted in the survey. In connection with the comparison group survey and the second NIST survey, we are unclear as to whether respondents were representative of all targeted employees.

Because of the methodological difficulties, OPM has decided, in effect, not to permit the comparison group survey and the second NIST survey to be used for further comparison purposes. In addition, OPM has instructed HII to do future employee surveys by telephone.

OPM has required HII to do at least one future survey of NIST employees and to do so in 1992, the year in which the project ends. OPM has instructed HII to obtain responses from at least 95 percent of the NIST employees.

Changing Conditions at NIST Not Adequately Considered

About 9 months after the project began, Congress renamed the National Bureau of Standards NIST and gave it the additional mission of assisting "industry in the development of technology and procedures needed to improve quality, to modernize manufacturing processes, to ensure product reliability, manufacturability, functionality, and cost-effectiveness, and to facilitate the more rapid commercialization . . . of products based on new scientific discoveries." NIST has made numerous other organizational changes since January 1988, including creating, abolishing, and realigning organizational units. In addition to these changes, a new NIST director, who came from within NIST's managerial ranks, took office in February 1990.

None of these changes were made because of the demonstration project. However, such changes can affect evaluation findings by producing effects that may be misinterpreted as project effects. How much of an impact, if any, the organizational, mission, and managerial changes will have on evaluation results is unknown. Examining, documenting, and measuring their effects was not a priority during the first 2 years of the evaluation. However, OPM has acknowledged the need to understand the project in the context of NIST's changing conditions. It is requiring HII to monitor changes at NIST and interpret project data in light of those changes.

Design Lacked Adequate Analysis Plan

As we noted in our May 1988 report, a data analysis plan is an essential element of an evaluation because it sets out how data will be used once it is collected. A well developed plan identifies what data will be collected, when collection will occur, and how the data collected will be analyzed. Because an analysis plan should serve as a data collection blueprint, it can help ensure the collection of appropriate data at appropriate times.

The analysis plan that existed for the NIST evaluation was only in "draft" form and prepared in December 1989, long after the evaluation began. In our view, it was not detailed enough to be of practical use in guiding either data collection or analysis. For example, the specific items that would be incorporated into each analysis were not identified. URC officials said they were unable to develop a full-fledged analysis plan because other tasks that also had to be done within the evaluation's budget had greater priority.

The consequence of an inadequate analysis plan is that OPM cannot know in advance how the evaluator will treat the data, nor can it monitor easily the evaluator's progress in carrying out the numerous components of the evaluation. OPM is requiring HII to submit a detailed analysis plan before starting work each evaluation year.

Other Useful Information Should Be Collected

We said in our 1988 report on the Navy demonstration project (see footnote 4) that to understand how a successful program might be replicated, as well as why a program might have failed to achieve its objectives, data on project effects are not enough. The evaluation, in addition, should document the specific components of the program being studied and monitor the extent to which the program was implemented. We said collection of this qualitative implementation information can help identify the project's strengths and weaknesses. Such identification would (1) enable critical linkages to be made between project changes and effects and (2) provide a substantive basis for improving and refining future demonstrations.

We believe this is true as well for the NIST evaluation. Data on implementation should involve an examination of the organizational changes and support mechanisms that were necessary to put project activities into use. For example, documenting decisions made and obstacles overcome in awarding recruitment and retention bonuses, delegating authorities, and simplifying classification procedures would be useful. How project implementation differed between NIST's Gaithersburg and Boulder sites

and among units within NIST and across occupational groups and pay bands should also be documented.

Gathering information on project implementation was not a sufficiently major focus during the first 2 years of the evaluation. That has changed, however. HII plans to do a "treatment/process analysis," which will assess how the project was implemented, the extent to which it was implemented, and the consistency with which it was implemented across NIST. The treatment/process analysis is a major part of HII's evaluation approach.

OPM's Management of the Evaluation

Our review focused on the design of the evaluation and how that design was being carried out. However, we also reviewed collected information to see if we could ascertain underlying reasons for the methodological problems we found. Despite its active oversight of the contractor's work, OPM's funding of the evaluation may have contributed to the problems.

OPM issued work orders to URC totaling $176,000 in 1988 and $191,000 in 1989. A portion of these funds, about one-fourth of the 1989 amount, for example, were for a URC subcontractor. Over 90 percent of the work order costs were for labor, and URC was required to finish the tasks assigned for the labor amounts (dollars) in the work orders. This included tasks URC assigned to the subcontractor.

URC project officials said the funding level prevented URC from providing the degree of effort necessary to do a good evaluation. They said the funding level required them to prioritize what had to be done, and the priorities were, in a sense, data driven, such as getting the employee surveys done. This meant that other activities, such as the analysis plan, may not have been done as quickly or as properly as they should have been or may have gone undone.

We cannot be certain that an evaluation free of the problems we found would have been accomplished with greater OPM funding. However, on the basis of our experience with other evaluations, we believe the evaluation's funding level was less than what was needed for a full-fledged impact evaluation of a complex personnel program. We do not believe it permitted enough staff to be placed at NIST and the comparison sites to collect the kinds and amounts of data an evaluation of this complexity requires.

OPM now has greatly increased funding for the evaluation. Scheduled contractor funding for the third-year evaluation is approximately $369,000. Estimated funding for the evaluation's fourth and fifth years is approximately $280,000 and $309,000, respectively.

Conclusions

Doing a scientific study of the impacts of a personnel management demonstration project in a dynamic environment is a difficult but worthwhile undertaking. At the start, OPM compounded the difficulty of obtaining a useful NIST evaluation by underfunding it. Nevertheless, the evaluation had a particular type of research design—nonequivalent control group—that is very appropriate to answering the types of questions posed by Congress concerning the effectiveness of the NIST project. In our opinion, however, that research design was not properly carried out. For example, the comparison group was not convincingly demonstrated to be a suitable comparison to NIST. Additionally, we believe the evaluation's design should have placed greater emphasis on gathering information on project implementation.

OPM acted promptly on our concerns. It directed its new contractor to address those concerns and provided additional funding for the evaluation. Whether the evaluation will now achieve its objective remains an unknown, however. We cannot precisely forecast how successful OPM and HII will be in addressing our concerns or what other problems they may encounter. For example, we do not know what difficulty the evaluator will face in developing baseline data with which to measure the demonstration project against. Nevertheless, we believe that chances for a methodologically sound and informative evaluation would have been poor had OPM not acted and that OPM's prompt responses have increased the chances for a more methodologically sound and informative evaluation. Accordingly, we are not making recommendations.

Comments From Project Officials

Project officials from NIST, OPM, and URC reviewed the information in this report. The NIST official suggested certain technical changes, which we made. The OPM official said OPM had no substantive comments but was skeptical that greater funding over the first 2 years would have improved the evaluation. We continue to believe that higher funding would have allowed greater staffing levels, thereby increasing the chances for better designing and implementing the evaluation. The URC official said, in summary, that OPM and URC both overreached on what pieces of the project to evaluate and, as a result, URC was spread too thin in doing the evaluation.

We will provide copies of this report to NIST, OPM, URC, and the Department of Commerce. We will also provide copies to others upon request. The major contributors to this report are listed in the appendix. If you have any questions, please call me on (202) 275-5074.

Bernard L. Ungar
Director, Federal Human Resource
Management Issues

Major Contributors to This Report

General Government Division, Washington, D.C.

Larry H. Endy, Assistant Director, Federal Human Resource Management Issues
Eva L. Rezmovic, Assistant Director, Design Methodology and Technical Assistance Group
Anthony Assia, Evaluator-In-Charge
Richard B. Groskin, Evaluator
Leah Cates, Reports Analyst

Ordering Information

The first five copies of each GAO report are free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

U.S. General Accounting Office
P. O. Box 6015
Gaithersburg, MD 20877

Orders may also be placed by calling (202) 275-6241.

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