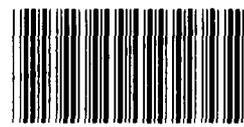119981

BY THE COMPTROLLER GENERAL
Report To The Chairman, Committee On Appropriations House of Representatives
OF THE UNITED STATES

Benchmarking: Costly And Difficult, But Often Necessary When Buying Computer Equipment Or Services

The Federal Government uses benchmarking for most competitive computer procurements to ensure that each vendor's proposal is evaluated fairly. This computer system validation process is necessary-- but costly and difficult for both the vendors and the Government.

GAO studied 73 computer procurements to identify (1) problems agencies have encountered in the benchmark process, (2) alternative evaluation techniques that can be used, and (3) the cost of benchmarking.

The General Services Administration should revise the procurement regulations and issue guidance to help agencies avoid benchmarking problems.

119981

GAO/AFMD-83-5
OCTOBER 22, 1982

523703

B-208077

The Honorable Jamie L. Whitten
Chairman, Committee on Appropriations
House of Representatives

Dear Mr. Chairman:

In your January 23, 1981, letter (app. II), you asked us to
assess the costs of benchmarking in automatic data processing
procurements and determine the need for, cost effectiveness of,
and alternatives to the benchmark process.

This report presents the results of our assessment. As ar-
ranged with your office, unless you publicly release its contents
earlier we will make no further distribution of the report for 30
days. At that time we will send a copy to the Administrator of
General Services and will make copies available to others who re-
quest them.

Sincerely yours,

Comptroller General
of the United States

COMPTROLLER GENERAL'S REPORT
TO THE COMMITTEE
ON APPROPRIATIONS
HOUSE OF REPRESENTATIVES

BENCHMARKING: COSTLY AND
DIFFICULT, BUT OFTEN NECES-
SARY WHEN BUYING COMPUTER
EQUIPMENT OR SERVICES

D I G E S T

One of the evaluation tools used most commonly by
both the Federal Government and private industry
for computer system validation is the benchmark.
It consists of a set of programs and associated
data tailored to represent a particular data proc-
essing workload. The Chairman, House Committee on
Appropriations requested that GAO assess the costs
of performing benchmarks and determine the need
for, cost effectiveness of, and alternatives to
the benchmark process. The Committee is concerned
about whether the high cost of benchmarking justi-
fies its use.

The benchmark process can be a very costly and
difficult undertaking for both the vendors and the
Government. Its success as an evaluation tool,
however, depends on the extent to which the bench-
mark can be made representative of an agency's
projected workload. When properly constructed and
used, the benchmark test

--is acceptable to the computer industry as a fair
  and unbiased test of a vendor's proposed system,

--allows an agency to model its current and pro-
  jected data processing workload with some ac-
  curacy,

--is repeatable within acceptable limits across
  vendor lines, and

--allows vendors to determine the appropriate size
  computer system to propose.

PROBLEMS OF BENCHMARKING

Benchmarking can cause technical and procurement
problems for both the agency and the competing
vendors. Agency benchmarking problems arise be-
cause it is difficult to

--get together talented people who are both ex-
  perienced in benchmarking and knowledgeable
  about the agency's workload;

GAO/AFMD-83-5
OCTOBER 22, 1982

i

--accurately project future workload growth and characteristics beyond 1 or 2 years (see p. 15);

--determine which application programs are representative of the entire workload (see p. 23);

--develop benchmark programs that are portable across vendor lines, without losing workload representativeness (see p. 12); and

--determine the impact that vendor modifications to the benchmark programs have on system performance and cost.

Vendors must deal·with several problems regarding benchmarking in the procurement process. The major problems are (1) agency errors in benchmark programs and poor documentation and (2) difficulties in communicating with the agency to resolve technical issues caused by the first problem. (See pp. 25-28.) Other problems cited by vendors include

--the high cost of benchmarking, which may restrict entry of small businesses into competition;

--the short time agencies often allow vendors for responding to a benchmark requirement;

--agency changes to the benchmark after it has been released;

--unavailability of benchmark programs when the solicitation document is released;

--the significant software conversion efforts required for some benchmarks; and

--agencies using benchmarks to limit competition or preselect a vendor.

BENCHMARKING IS COSTLY,
BUT FEW ALTERNATIVES EXIST

Federal procurement policy requires, to the extent practicable, competitive acquisition of needed goods and services. One of the objectives of this policy is to give all responsible vendors an opportunity to compete. Benchmarking is needed for most competitive computer procurements to ensure fair evaluation of proposals because (1) all responsible vendors are allowed to compete, (2) each vendor's computer system has unique features that

make comparison with another system difficult, and (3) Federal procurement actions are open to public scrutiny, so selection decisions must be defensible. However, for procurements that have a system life contract value of less than $2 million, or for procurements limited to compatible vendors, other evaluation techniques are more appropriate because of the high cost of benchmarking. The other evaluation techniques include paper/technical evaluation, and analytical modeling and simulation. Another technique--more costly than benchmarking--that can be used for major system acquisitions involving high cost and risk is a "compute-off." In this, the Government funds two or more vendors to develop a prototype system. (See p. 8.)

GAO analyzed selected agency procedures for 73 computer procurements, of which 65 had a benchmark requirement. Agency benchmarking costs in the procurements reviewed ranged from about $1,000 for the acquisition of a minicomputer for the Treasury Department's international division to more than $2 million for the replacement of computer equipment in 10 Internal Revenue Service regional processing centers. Most costs fell between $50,000 and $200,000. We did not include inhouse computer time in these figures. Vendor-reported benchmarking costs for these same procurements ranged from $4,000 to more than $2 million. (See app. I.) Although benchmarking costs were substantial in many cases, there was, in GAO's view, no practical alternative to benchmarking for most of the procurements included in its study.

PRIVATE SECTOR ACQUISITION PRACTICES

GAO found that the private sector practice of limiting competition for business to certain vendors is a major difference from Federal procurement policy and influences many facets of the acquisition process, including the use of benchmarking. If an organization's acquisition strategy is to stay primarily with a particular manufacturer's equipment or compatible computer equipment, then a benchmark is usually unnecessary. If a benchmark is used at all, its scope and size can be greatly reduced. This is not true in a fully competitive environment, as is mandated for most Federal acquisitions. Benchmarking is the most common evaluation technique used in the selection of computer equipment and services by the Federal Government because it is the

only technique that can reliably compare the performance of different computers <u>and</u> is generally acceptable to the vendor community. However, GAO believes that for compatible acquisitions, the Federal Government can adopt the private sector practice of limited use of benchmarking. (See p. 19.)

BENCHMARKING PRACTICES CAN BE IMPROVED

GAO found that the use of benchmarking as a performance measurement technique has often been criticized because (1) it can place a significant resource burden on a vendor trying to compete, (2) it can be used by an agency to limit competition, (3) vendors can modify the benchmark programs so that they are no longer representative of the actual workload, and (4) for services procurements, vendors can submit low evaluated-cost proposals based on the benchmark workload, gambling that the actual workload will be substantially different. However, even with these potential shortcomings, GAO believes there is no better alternative to benchmarking for most fully competitive computer equipment or services procurements. (See pp. 22 and 23.)

Vendors encounter great difficulty in trying to resolve the problems stemming from poor benchmarks and difficult communication. GAO concludes that if agencies can improve benchmarking practices in these two areas, the cost and time burden on the vendor can be greatly reduced. To improve the quality of the benchmark before releasing it to vendors, Federal agencies should test it on at least one computer other than the one on which it was developed. Furthermore, the test should be performed by an independent group to ensure that the benchmark instructions and documentation are complete and consistent. To expeditiously resolve technical questions relating to the benchmark, Federal agencies need to be far more open in their communication with vendors. This is essential to increasing competition and can be done without compromising the procurement. (See pp. 25 and 27.)

GSA GUIDANCE IS NEEDED

The Federal Government has been concerned about benchmarking practices since at least 1969 when it was a major topic of a computer acquisition conference sponsored by the Office of Management

and Budget. There have been several studies and workshops on the subject since that time. The General Services Administration is responsible for developing Government-wide guidance for computer acquisitions and has issued guidelines on benchmarking. The National Bureau of Standards has also issued two sets of guidelines. However, other than for teleprocessing services acquisitions, agencies still receive inadequate guidance.

No single answer fits all situations. In most fully competitive procurements, GAO believes benchmarking is the most appropriate evaluation technique. However, for fully competitive procurements of computer equipment with a projected system life contract value of less than $2 million, benchmarking should be discouraged because of its high costs. In other circumstances more cost effective evaluation methods are indicated. Prudent judgement should prevail.

RECOMMENDATIONS

To improve the acquisition of computer equipment, GAO recommends that the Administrator of General Services:

--Develop criteria that will help Federal agencies determine (1) when it is appropriate to benchmark, (2) if benchmarking is needed, what approach is most appropriate, and (3) if benchmarking is not needed, what alternative should be used.

--Revise the Federal Procurement Regulations (FPR 1-4.1109-22) so that benchmarking is discouraged for computer equipment procurements with a projected system life contract value of less than $2 million.

AGENCY COMMENTS

The General Services Administration provided written comments on this report. The agency generally agreed with GAO's conclusions and recommendations, but stated that it lacks the technical expertise to carry them out and will rely on outside assistance. (See app. III.)

C o n t e n t s

ABBREVIATIONS

ADP   automatic data processing

CPU   central processing unit

FEDSIM   Federal Computer Performance Evaluation and
           Simulation Center

FIPS   Federal Information Processing Standards

FPR   Federal Procurement Regulations

GAO   General Accounting Office

GSA   General Services Administration

IBM   International Business Machines

NBS   National Bureau of Standards

OMB   Office of Management and Budget

TSP   Teleprocessing Services Program

VA   Veterans Administration

GLOSSARY

Algorithm

A mathematical formula used by computer services companies to recover costs and determine billable charges for computer resources used by a customer. Depending on the complexity of the computer system and the needs of the company, billing algorithms range from very simple to very complex.

Application software

A set of instructions (called program statements or code) to do a specific job, such as payroll computation, inventory control, and accounting. It is also called application program.

Assembly language

A computer language in which the programmer conveys single machine language statements, or groups of them, by terse mnemonic codes. Programs written in assembly language typically can run only on the make of computer for which they were originally developed. It is also called assembler language.

Benchmark

A set of computer programs and associated data tailored to represent a particular workload and used to evaluate system performance or cost.

Central processing unit (CPU)

That part of a computing system that contains the circuits for interpreting and executing instructions. The CPU includes the control and arithmetic units and an internal storage area.

Code

An ordered list or lists of successive instructions that will cause a computer to perform a particular process.

Compiler

A computer program that translates high-level programming language statements into a form that can directly activate the computer equipment.

Computer network

A complex collection of many types of automatic data processing resources including two or more interconnected computers, data bases, application programs, and special-purpose equipment.

Data base management system	A computer software package that can facilitate the management, manipulation, and control of data.
Functional demonstration	A demonstration of a function or capability without regard to total system performance by a vendor. Often, the demonstration is performed with vendor-provided software and data. It is also called an operational capability demonstration.
High level language	A computer programming language that does not reflect the structure of any given computer or that of any given class of computers.
Internal driver	A benchmarking technique for on-line systems where the remote communications elements, for example, operators, terminals, modems, and lines, are represented by special data and software located within the system being tested.
Live test demonstration	A user-witnessed running of the benchmark on a vendor's proposed computer system, done to validate system performance or cost.
Main memory	The storage component that is considered integral, internal, and primary to the computer system.
Portable software	Software that will require little effort to convert to differing computer systems.
Redesign	Change to application software that involves a change to its functional specifications. The application software will provide new functions and capabilities; it is akin to new development.
Remote terminal emulation	A benchmarking technique for conducting tests of teleprocessing computer systems and services, used when it is impractical to test the entire network of computers, teleprocessing devices, and data communication facilities.
Reprogramming	Any change to application software that conforms to the software's functional specifications but causes changes to the methodology for meeting functional requirements of the user. It is also called technical redesign and rewrite.

Scenario	A system- and vendor-independent description of a group of user workload demands, performed during a benchmark demonstration. It is expressed in terms of user functions.
Software	A set of programs, procedures, and documentation for operating a data processing system. Three categories are (1) application software, (2) operating system software, and (3) utility software.
Software conversion	Making computer programs run on a computer system other than the one for which they were originally devised. Software conversion can be accomplished by translation or reprogramming.
Software package	Computer program(s) and documentation, such as flowcharts and user manuals, prepared as a unit and often sold by a vendor.
Synthetic program	A computer program representing a particular class or function of application programs, used for benchmarking purposes only. Synthetic programs can be oriented toward functions or toward resources.
Translation	A largely automated process of application software conversion in which the functional requirements and design specifications are preserved. It is also called recoding when the process is largely manual.

CHAPTER 1

INTRODUCTION

Federal agency data processing managers are responsible for validating performance and cost before buying a computer system or service. One of the evaluative tools most commonly used by both the Federal Government and private industry for system validation is the benchmark. It consists of a set of computer programs and associated data tailored to represent a particular data processing workload. The benchmark process can be very costly and difficult for both the vendors and the Government. Its success as an evaluative tool, however, depends on the extent to which the benchmark can be made to represent an agency's projected workload.

THE BENCHMARK PROCESS

Benchmarking is an important evaluative technique within the Federal Government and, to a lesser extent, in the private sector for several reasons. When properly constructed and used, the benchmark test

--is acceptable to the computer industry as a fair and un-biased test of a vendor's proposed system,

--allows an agency to model its current and projected data processing workload with some accuracy,

--is repeatable within acceptable limits across vendor lines, and

--allows vendors to determine the appropriate size computer system to propose.

The value and accuracy of benchmarking results can surpass other evaluation techniques. However, benchmarking often costs more than other techniques. Therefore it should be used only when the value of the procurement and the necessity for accuracy clearly justify the expense.

Phases of benchmarking

The Federal benchmark process can be broken down into six phases, as follows:

(1) <u>Workload definition and analysis</u>

The workload to be performed by the new or replacement system is defined. This requires an analysis of the current workload, a projection of future growth, and an estimate of planned new applications.

(2) Design and construction

A benchmark is constructed to represent a defined workload. The agency may select programs from current production applications, develop new ones to represent each workload category, or use a combination of both.

(3) Testing

The agency combines the selected benchmark programs with transactions and data in the combinations required to represent all workload categories. The benchmark should be thoroughly tested on the present system as well as on at least one computer system of a different manufacturer. It is then modified to eliminate errors and major unique features of the present system to the maximum extent possible.

(4) Agency package preparation

The agency prepares the benchmark package--the physical files containing the benchmark programs, data, and transactions--and appropriate documentation for the competing vendors to use in the benchmark demonstration. The agency also prepares and documents the rules for conducting the demonstration.

(5) Vendor preparation

Each competing vendor makes any necessary and allowable changes to the benchmark so that it will operate on the vendor's computer equipment. Each vendor also configures a computer system capable of processing the benchmark within the agency-determined time constraints. During this phase, vendor questions and comments inevitably surface.

(6) Demonstration

The benchmark is performed by each vendor and the results are evaluated by an agency team. The demonstration is usually witnessed by the team. When this is not the case, the vendor sends the demonstration results to the agency for evaluation.

During each of these phases, costs are incurred by the agency, the vendor, or both. (See app. I for benchmark costs incurred for the procurements included in this study.) Production of a high-quality benchmark can be a very expensive process for an agency. Low-quality benchmarks are less costly for agencies, but can be more costly for vendors, as in the case of poorly documented benchmark programs. Also, the risk is higher that the procured system may not satisfy the agency's requirements.

Problems in benchmarking

Although a necessary tool in many competitive procurements, benchmarking can cause technical and procurement problems for both the agency and the competing vendors. These problems arise because it is difficult to

--get together talented people who are both experienced in benchmarking and knowledgeable about the agency's workload;

--accurately project future workload growth and characteristics beyond 1 or 2 years;

--determine which application programs are representative of the entire workload;

--develop benchmark programs that are portable across vendor lines, without losing workload representativeness; and

--determine the impact that vendor modifications to the benchmark programs have on system performance and cost.

Vendors must deal with several problems regarding benchmarking in the procurement process. The major problems are: (1) agency errors in benchmark programs and poor documentation and (2) difficulties in communicating with the agency to resolve technical issues caused by the first problem. Other problems cited by vendors include

--the high cost of benchmarking, which may restrict entry of small businesses into competition;

--the short time agencies often allow vendors for responding to a benchmark requirement;

--agency changes to the benchmark after it has been released;

--unavailability of benchmark programs when the solicitation document is released;

--the significant software conversion efforts required for some benchmarks; and

--agencies using benchmarks to limit competition or preselect a vendor.

Procurement problems arise because very limited guidance exists to help agencies determine (1) when it is appropriate to benchmark and (2) if benchmarking is needed, what approach is best. When properly constructed and used, however, benchmarking can be a valuable evaluative tool. Furthermore, in Federal procurements the benchmark results help defend a vendor selection, as well as make one.

RESPONSIBILITY FOR FEDERAL ADP MANAGEMENT

The Brooks Act (Public Law 89-306), enacted in October 1965, provides for the economic and efficient purchase, lease, maintenance, operation, and use of automatic data processing (ADP) equipment. Responsibilities under the act are assigned to several

3

agencies: The General Services Administration (GSA) is responsible for developing, implementing, and monitoring Government-wide policy for the acquisition, use, and management of ADP resources. The Department of Commerce, primarily through the National Bureau of Standards (NBS), is responsible for providing scientific and technological advisory services and for developing Federal Information Processing Standards (FIPS). The Office of Management and Budget (OMB) is responsible for fiscal and policy control. In addition, each Federal agency has certain responsibilities for managing its own ADP resources.

CONCERNS OF THE HOUSE APPROPRIATIONS COMMITTEE

Computer acquisition cost has been a concern of the Committee for some time. Benchmarking represents a significant portion of these costs. The Chairman wrote to the Comptroller General on January 23, 1981, and requested that we study the use of benchmarking in Federal ADP procurement. (See app. II.) The Committee is concerned that the Government simply has no idea what it costs to carry out the benchmark process or whether this cost will be justified by reduction in future operating difficulties. The Committee believes that benchmarking represents such a significant cost in the ADP procurement process that it can discourage competition. It also believes that benchmarking costs are often hidden and lead to higher equipment costs as vendors try to recoup this Government-directed overhead expense. Specifically, the Committee has requested that we assess the costs of performing benchmarks in typical Federal and private sector ADP procurements and determine the need for, cost effectiveness of, and alternatives to the benchmark process.

OBJECTIVES, SCOPE, AND METHODOLOGY

The objectives of our review were to:

--Identify, for selected computer procurements, the cost to both Federal agencies and vendors of benchmarking.

--Identify private sector benchmarking practices.

--Determine whether benchmarking discourages competition.

--Identify alternatives to benchmarking and when they would be appropriate.

--Find out how vendors recoup benchmark costs.

--Identify benchmark-related problems of both agencies and vendors.

The review was performed in accordance with our current "Standards for Audit of Governmental Organizations, Programs,

4

Activities, and Functions." We reviewed GSA regulations and procedures and NBS guidelines and other publications. We discussed the use of benchmarking in the ADP procurement process with staff members of the House Committee on Appropriations. To address benchmark-related issues across the broad spectrum of computer equipment and services offerings, we analyzed selected agency procedures pertaining to 73 procurements, of which 65 had a benchmark requirement.

The cases we reviewed included:

--32 competitive medium- and large-scale computer procurements,

--14 competitive minicomputer procurements,

--12 procurements limited to IBM-compatible computer equipment, and

--15 teleprocessing services procurements.

We included eight procurements that did not have a benchmark requirement so we could identify alternative performance validation techniques.

We interviewed officials of GSA, NBS, the Federal Computer Performance Evaluation and Simulation Center (FEDSIM), and 57 selected civil and defense agencies to obtain their views on benchmarking practices. We also interviewed officials from two State government organizations and 12 private sector firms that have recently conducted ADP procurements.

We held discussions with representatives from 28 computer industry firms and 1 industry trade association to obtain their views on whether benchmarking discourages competition, how benchmarking costs are recouped, and what alternatives to benchmarking exist.

We researched computer industry trade journals, technical documents, and other publications. We also reviewed our decisions in which the benchmark was a protest issue, and information contained in the workpapers supporting our study, "Non-Federal Computer Acquisition Practices Provide Useful Information For Streamlining Federal Methods" (AFMD-81-104, Oct. 2, 1981).

CHAPTER 2

BENCHMARKING, WHILE COSTLY,

HAS FEW ALTERNATIVES IN THE

COMPETITIVE PROCUREMENT ENVIRONMENT

Federal procurement policy requires, to the extent practicable, competitive acquisition of needed goods and services. One of the objectives of this policy is to give all responsible vendors an opportunity to compete. Benchmarking, while costly, is needed for most competitive ADP procurements to ensure fair evaluation of proposals because (1) all responsible vendors are allowed to compete, (2) each vendor's computer system has unique features that make comparison with another system difficult without benchmarking, and (3) Federal procurement actions are open to public scrutiny, so selection decisions must be defensible. However, for procurements that have a system life contract value of less than $2 million or for procurements limited to compatible vendors, other evaluation techniques are more appropriate because of the high cost of benchmarking. GSA has provided very limited guidance to help agencies determine (1) when it is appropriate to benchmark, (2) if benchmarking is needed, what approach is best, and (3) if benchmarking is not needed, what alternative should be used. Also, the private sector practices of limited use of benchmarking for compatible acquisitions and more open communication with vendors concerning the benchmark can be adopted by the Federal Government.

BENCHMARKING CAN BE COSTLY

Benchmarking represents a significant expense to both the Federal Government and the computer industry. Agency benchmarking costs in the 73 ADP procurements we studied ranged from about $1,000 for the acquisition of a minicomputer for the Treasury Department's international division to more than $2 million for the replacement of the computer equipment in 10 Internal Revenue Service regional processing centers. Most costs fell between $50,000 and $200,000. (See app. I.) We did not include in-house computer time in these figures. Vendor-reported benchmarking costs for these same procurements ranged from $4,000 to more than $2 million. Benchmarking costs of competing vendors can vary significantly on the same procurement for several reasons:

    --The software conversion 1/ effort required by each vendor to make the benchmark programs run on its own equipment can be substantially different.

    --Some vendors may place greater emphasis on fine tuning the benchmark programs to take advantage of their equipment's

----

1/See glossary.

6

capabilities, thereby reducing the equipment configuration necessary to pass the benchmark demonstration test and lowering the evaluated cost.

--Some vendors may have to acquire additional equipment or software packages 1/ to perform the benchmark demonstration test.

Although benchmarking costs were substantial in many cases, there was, in our opinion, no practical alternative to benchmarking for most of the procurements included in our study. However, because the high cost of benchmarking can discourage competition by reducing the vendor's profit margin, we do not believe benchmarking is cost effective for procurements with a contract value of less than $2 million.

AGENCIES HAVE LITTLE GUIDANCE

The Federal Government has been concerned about benchmarking practices since at least 1969 when this was a major topic of an ADP acquisition conference sponsored by the Office of Management and Budget. Several studies and workshops on the subject have taken place since that time. NBS has issued two guidelines on benchmarking and GSA has issued one. 2/ However, other than for teleprocessing services acquisitions, very limited guidance exists to help agencies develop a benchmarking strategy or determine when to use alternative evaluation techniques.

The Federal Procurement Regulations (FPR 1-4.1109-22) do not require agencies to use benchmarking for the selection of computer equipment. The regulations properly discourage the use of benchmarks for low dollar value procurements (purchase value less than $300,000) when computer system performance can be validated by some other means. However, GSA's handbook entitled "Guidance to Federal Agencies on the Preparation of Specifications, Selection, and Acquisition of Automatic Data Processing Equipment Systems" states that, depending on the size and complexity of the processing requirements, the agency will specify either a benchmark demonstration or an operational capability demonstration (also called a functional demonstration), or both.

In many teleprocessing services acquisitions, the benchmark is the primary evaluation tool. GSA's draft Teleprocessing

---

1/See glossary.

2/The guidelines are: (1) FIPS Pub 42-1, "Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment"; (2) FIPS Pub 75, "Guideline on Constructing Benchmarks for ADP System Acquisitions"; and (3) GSA Handbook, "Use and Specifications of Remote Terminal Emulation in ADP System Acquisitions."

Services Program Handbook 1/ and FPR 1-4.1209-3(b)3 require agencies whose teleprocessing costs are expected to exceed $300,000 annually to benchmark all technically qualified vendors. The handbook also recommends benchmarking for requirements under $300,000 annually. The benchmark is used to (1) project system life costs of each competing vendor, and (2) test mandatory technical requirements. It can also be used to monitor the vendor's billing algorithm 2/ and system performance over the life of the contract.

ALTERNATIVE EVALUATION METHODS SHOULD BE CONSIDERED

Benchmarking is the most widely accepted method for evaluating computer systems in the Federal ADP procurement process. It can provide a fair and unbiased live test demonstration of candidate computer systems. When appropriate, however, other evaluation methods should be considered because of (1) the high cost of benchmarking, (2) its impact on competition, (3) the uncertainty of workload projection, and (4) the difficulty of making the benchmark representative of the existing workload. The other evaluation techniques that agencies can consider include paper/technical evaluation, analytical modeling and simulation, and a "compute-off" whereby the Government funds two or more vendors to develop a prototype system.

Figure 1 on page 9, prepared by an NBS official, compares the relative costs of various evaluation methods with their relative accuracy.

Paper/technical evaluation

In a paper or technical evaluation, the agency can base its selection decision on a combination of information from several sources. Known workload requirements can be translated into equipment performance specifications such as main memory size, disk storage, and central processing unit (CPU) 2/ cycle time. A number of ADP users have conducted a wide range of evaluation tests on commercially available computer systems. Many of the findings have been presented to professional groups and published in industry trade journals. The agency can rely on this published data and on data from the manufacturer on the performance and capacity of computer equipment. The agency can also contact users of similar systems to help validate system performance and capacity.

1/The Teleprocessing Services Program (TSP) is centrally managed by GSA and provides agencies with an efficient means of acquiring commercial ADP services under prearranged terms and conditions.

2/See glossary.

FIGURE 1. COMPUTER SYSTEM
EVALUATION TECHNIQUES COMPARED

COMPUTE-OFF

*

BENCHMARKING

*

SIMULATION

*

ANALYTICAL MODELING

*

PAPER/TECHNICAL
EVALUATION

*

COST

| CLASSIFY WORKLOAD
CLASSIFY SYSTEM | MODEL WORKLOAD
MODEL SYSTEM | MODEL WORKLOAD
ACTUAL SYSTEM | ACTUAL WORKLOAD
ACTUAL SYSTEM |

ACCURACY
OF
REPRESENTATION
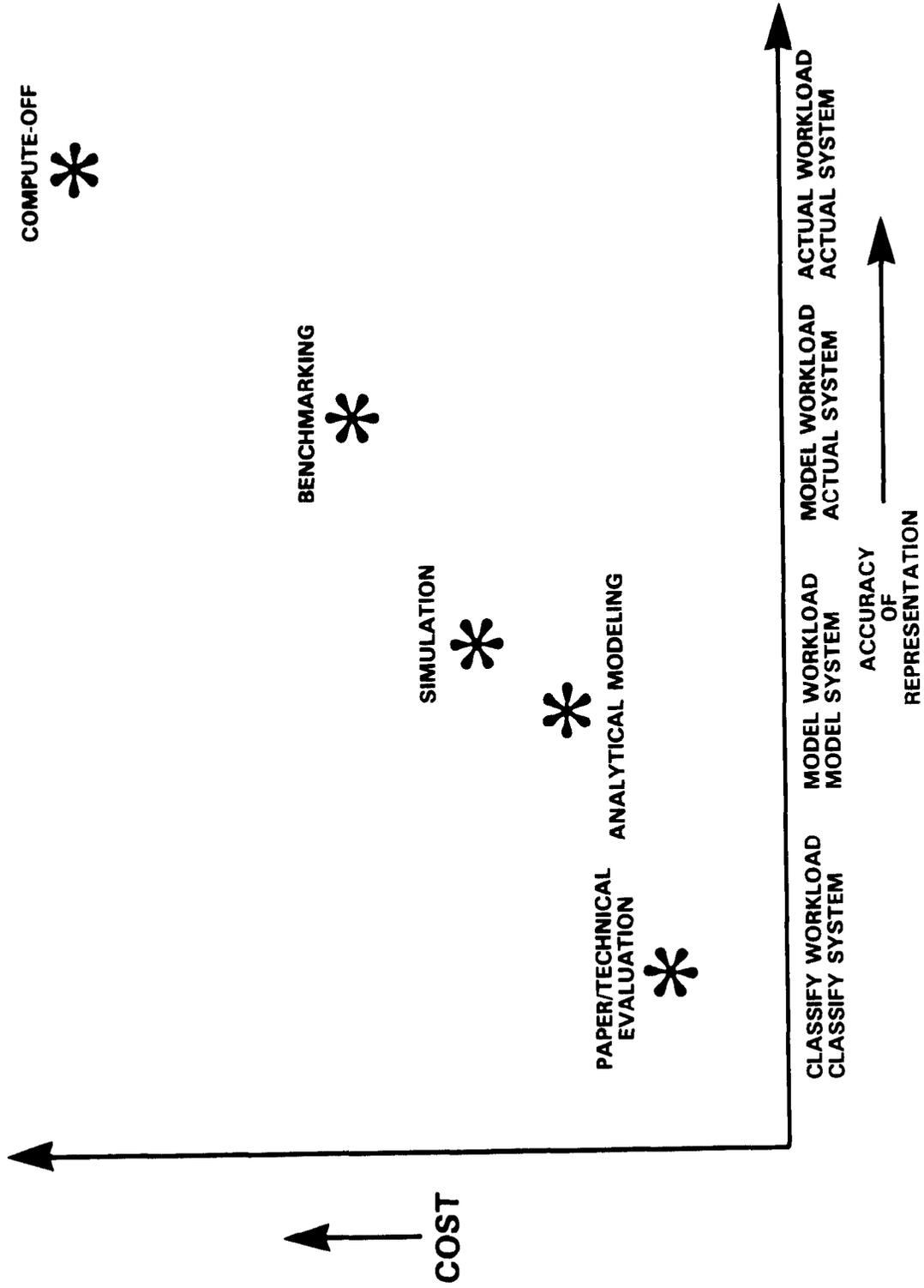
Of the 58 computer equipment procurements included in our study, six relied on a technical evaluation as the basis for selection. Three of the six--St. Louis and San Bruno Postal Data Centers and Air Force Strategic Air Command TRICOMS III--were procurements limited to IBM-compatible equipment. 1/ Two--National Aeronautics and Space Administration (Dynamic Explorer) and Federal Railroad Administration--were fully competitive procurements in which the system life contract value was less than $2 million. The Federal Railroad Administration's evaluation also included a functional demonstration. 2/ The sixth procurement was for two very large-scale scientific computers at the Lawrence Livermore National Laboratory. The Laboratory had gained sufficient knowledge from recent procurements of large-scale scientific computers that a benchmark was not necessary. We believe that these completed procurement actions demonstrate that benchmarking is not always necessary for limited competition or for procurements that have a system life contract value of less than $2 million.

Analytical modeling and simulation

The analytical modeling and simulation techniques are combined here because their advantages, disadvantages, cost, and accuracy are similar. These techniques can be used to determine whether a candidate system can meet the specified performance requirements, as well as provide additional system capacity. Analytical modeling and simulation both use a combination of mathematical expressions to represent the workload and the computer system. The primary difference is that simulation uses a higher level of detail. Both techniques can be highly accurate, within a vendor's product line, for predicting equipment performance. They can also be sufficiently accurate for compatible computers. However, they may not be as accurate across vendor lines.

The Federal Procurement Regulations (FPR 1-4.1109-21) state that simulation will not be used as the only means of describing data processing requirements, and proposals will not be considered nonresponsive or unacceptable solely on the basis of simulation results. The same restrictions apply to modeling. The restriction on simulation has been in effect since 1972 and originates from vendor concern that simulation does not properly represent the vendor's computer equipment. We believe that, for a fully competitive procurement, this concern is valid.

Analytical modeling or simulation was used in conjunction with benchmarking in three of the procurements included in our study, but was not used in the **selection decision.** FEDSIM used simulation to size equipment that vendors might propose for the Veterans

---

1/Several manufacturers offer computers that are compatible with various models of International Business Machines Corporation (IBM) equipment.

2/See glossary.

Administration's "Target" system. The simulation predicted a configuration that had greater capacity than the vendor proposed and used to pass the benchmark demonstration. The Army Corps of Engineers and the Marine Corps both used modeling (with contractor support) to determine what IBM or IBM-compatible computer equipment would meet their requirements, and both wanted to avoid benchmarking primarily because they had limited experience with it. However, the Army and Navy central ADP procurement agencies (Army Computer System Selection and Acquisition Agency and Navy ADP Selection Office) forced the use of a benchmark.

The Navy required the Marine Corps to use a benchmark because of the high dollar value (about $19 million) of the procurement. The Army required the Corps of Engineers to benchmark in order to keep the competition fair. In the latter case, the model predicted the same computer system as the winning vendor used in the benchmark. However, because the Marine Corps' benchmark workload was reduced below that used in the model, the vendor passed the benchmark demonstration using computer equipment that was about 20 percent less powerful than had been predicted by the model.

Because there has been a prohibition against the use of simulation and modeling as the sole evaluation methodology, its use by Federal agencies has been limited. We believe, however, that simulation and modeling can be used along with a paper or technical evaluation to evaluate proposals for limited competitions. It also should be considered to complement benchmarking for large, complex systems such as those using computer networking, 1/ where the cost of benchmarking for the total system would be prohibitive.

Compute-off

An alternative evaluation method agencies can consider for very large procurements is the "compute-off," in which the Government funds two or more vendors to develop a prototype system. Factors to be considered in using this method are discussed in OMB Circular A-109. The circular provides guidance on major system acquisitions that involve high cost and risk to the Government.

The compute-off is much more costly and time consuming than benchmarking but reduces the risk, since a prototype of the actual system is completely developed by each proposing vendor. One procurement included in our study--the Army's Project VIABLE--used this approach. The Army paid two vendors $5.1 million each to develop prototype systems. It also used benchmarking to help determine which two vendors would be funded to develop the prototypes and to help evaluate vendor proposals. We believe this approach should be used for computer acquisitions only when risk to the Government is extremely high.

1/See glossary.

11

MANY FACTORS INFLUENCE BENCHMARKING STRATEGY

The determination of whether a benchmark is appropriate, and if so, what benchmarking approach is most suitable, must be considered within the agency's overall acquisition strategy and be made as early as possible in the acquisition cycle. The benchmarking strategy is influenced by many factors including (1) how the agency plans the transition of its software to the new equipment, (2) the acquisition strategy, (3) the impact on competition, and (4) the uncertainty of workload projection. Figure 2 on page 13 shows how these factors can influence an agency's benchmarking strategy.

Software transition
and acquisition strategies

Application software 1/ costs considerably more than the computer equipment in most systems. How the agency plans the transition of its software to the new equipment should point to the most suitable acquisition and benchmarking strategies. Several approaches, or combinations of approaches, can be taken. An agency can convert its existing software, redesign it, 1/ or, when the equipment is needed to support a new mission or a function not previously automated, develop new application software.

Conversion

Conversion is a key factor in determining benchmarking strategy because (1) conversion costs may be so substantial that competition is effectively limited to compatible vendors (incumbent, other manufacturers, third-party vendors) and (2) how the vendor converts (or optimizes) the benchmark programs can change the benchmark's representativeness. In a limited competition or sole-source acquisition that has been justified, a benchmark is usually not needed; if used at all, its scope and size can be greatly reduced. One of the purposes of the benchmark is to reduce the Government's risk of acquiring a computer system that will not meet its needs. In an acquisition limited to compatible vendors, the agency's risk is greatly reduced simply because the software is compatible; it is a known commodity. Also, if the agency makes a mistake in workload projection or equipment sizing, adequate competition for equipment augmentation will be available.

If the agency determines that full competition is the best acquisition approach and a significant conversion effort is required, the agency should take steps to ensure that the benchmark representativeness is not compromised by the competing vendors.

Two approaches can be used to convert application software: translation and reprogramming.

1/See glossary.
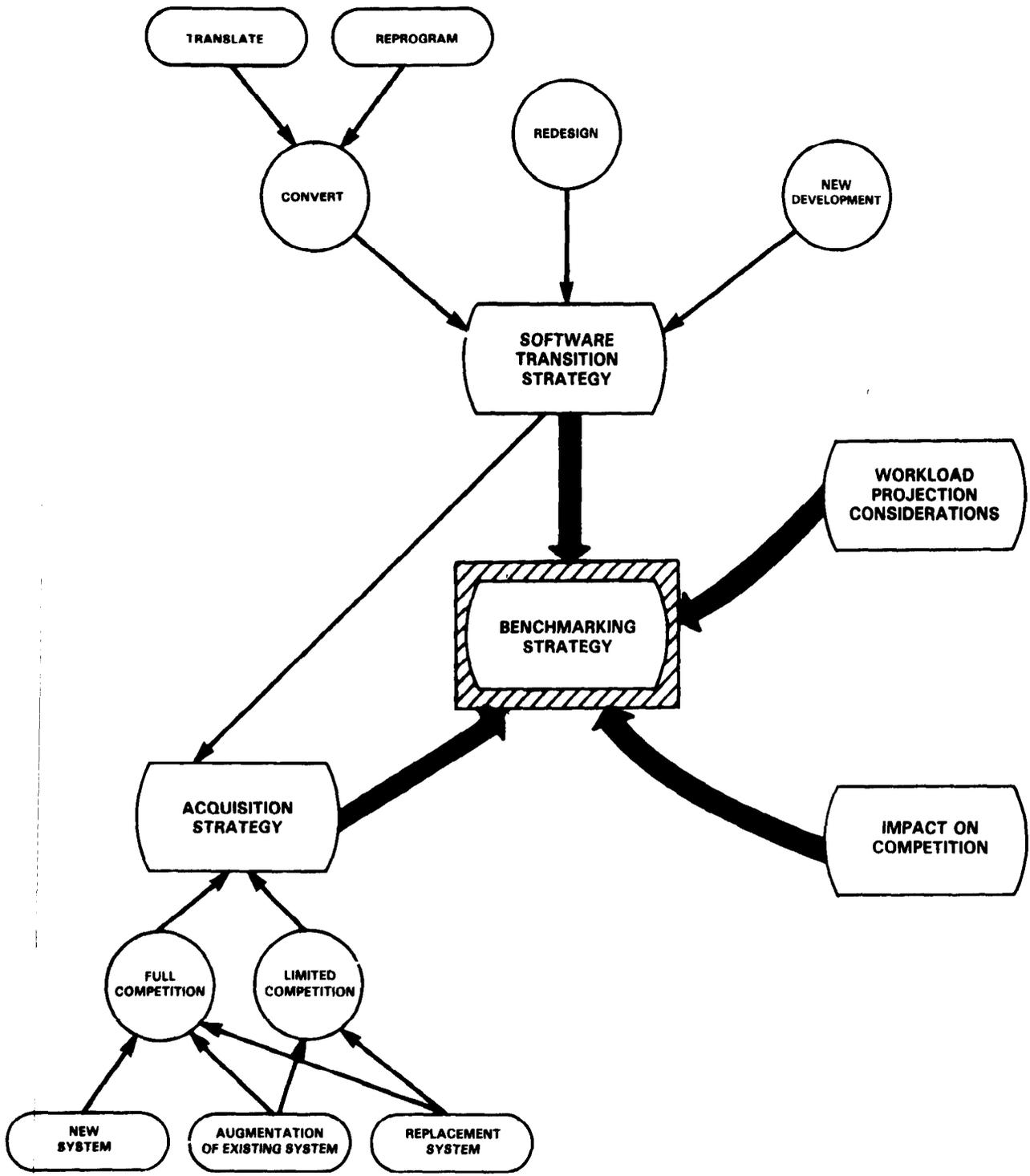
FIGURE 2. FACTORS INFLUENCING BENCHMARKING STRATEGY.

Translation--Application programs can be translated 1/ line for line, meaning they will be changed only enough to make them run on the new computer. Generally, this means that software fine-tuned over the years for the old system will not be as efficient on the new system. This inefficiency can affect the benchmark's ability to properly size the equipment. Agency benchmark programs should be as portable 1/ as possible but most of the agencies' application programs are not, in fact, portable. In addition, the vendor, when converting the benchmark programs, will try to optimize them to take advantage of the existing equipment's features. Often, the vendor's approach to converting the benchmark programs is similar to the reprogramming conversion approach.

Reprogramming--This approach takes full advantage of the new computer system's capabilities. However, the cost, complexity, and time required to reprogram make it unacceptable in most cases for an entire application software inventory.

If the agency doesn't consider the impact of the conversion approach in constructing its benchmark, the resulting computer equipment configuration may not have sufficient capacity to process the actual workload and the agency will have to augment the new equipment much sooner than planned. 2/ If conversion is a factor, the agency can consider requiring the vendor to run the benchmark programs twice; first with the minimum changes necessary to make them run, and the second time by taking advantage of the new computer's capabilities. However, we believe that in most cases, if conversion is substantial, competition will be limited because of the existing large investment in application software.

Federal agency management can reduce conversion-related problems by emphasizing the need to develop better quality application software in the first place. This can be done by using standard high-level programming languages and instituting sound programming and documentation practices. Conversion should be considered during initial application program development. It requires well structured and well documented programs that are maintainable, readable, and understandable. If these quality standards are adhered to, benchmark-related problems can be reduced.

Redesign or new development

Agencies planning to redesign or reprogram their application software or to develop software to support a mission or function

---

1/See glossary.

2/For further discussion of the impact of conversion on operating efficiency see p. 42 of our report, "Conversion: A Costly, Disruptive Process That Must Be Considered When Buying Computers," FGMSD-80-35, June 3, 1980.

not previously automated, should construct their benchmark programs using standard high-level programming languages to increase portability and thereby reduce vendors' benchmark-related costs and problems. If the agency determines that a data base management system 1/ is required, the agency should provide a functional description of the requirements and allow the vendor to choose the manner in which the data base system is implemented for demonstration purposes.

Impact on competition

The decision to use a benchmark and if so, what type of benchmark, can have an impact on which vendors will compete for business. The benchmark should be as simple as possible without compromising requirements or unduly increasing the risk to the Government that the delivered computer equipment will not have sufficient capacity or that services will be unreasonably costly. Benchmark requirements can differ greatly in complexity and, therefore, in cost to the vendor. An agency should compare the risk associated with not using a benchmark or using a less complex one, with the risk of requiring a benchmark and possibly reducing competition. The defense of a selection decision should not be the overriding consideration in developing a benchmark strategy. In chapter 3 we discuss the factors vendors consider when deciding whether to compete and how the benchmark fits into that process. The benchmark can discourage competition in some situations.

Difficulty of workload projection

When developing a benchmark strategy, an agency should consider the difficulty of projecting future workload growth. Many agency officials we interviewed believe it is not possible to accurately project workload growth beyond 1 or 2 years. However, Federal agency benchmarks usually try to represent workload growth over the entire system life, often a 5- to 10-year period. If the workload projection is inaccurate, the benchmark will not be valid for predicting system capacity or cost. The problem of inaccurate workload forecasting is independent of the evaluation technique used.

An agency can compensate to some extent for unexpected increase in the workload by including provisions for fixed-price options that allow the agency to acquire additional equipment at prices determined through competition rather than through sole-source negotiation. Vendors do not like to see fixed-price options as they place a greater risk on the vendor, which could result in a higher proposal price. However, given the Government's

---

1/For further discussion of data base management system software, see our report, "Data Base Management Systems--Without Careful Planning There Can Be Problems," FGMSD-79-35, June 29, 1979.

history of (1) keeping equipment longer than the system life used for evaluation and (2) acquiring more equipment than initially contracted for, we believe it is in the best interests of the Government to include provisions for fixed-price options for additional equipment. For IBM-compatible procurements, fixed-price options are less desirable because ample competition, and therefore lower price, exists among compatible vendors for equipment augmentation.

Workload projection not attempted

In two of the procurements in our study, the agency did not try to represent its workload or project it over the entire system life. The Securities and Exchange Commission obtained from an IBM users group a computer. program developed solely for testing computer performance. The Commission relied on published results of earlier benchmark tests of IBM equipment and required IBM-compatible vendors to perform the same benchmark demonstration. One vendor did, and was awarded the contract. The other procurement was for computer equipment to support the Defense Communications Agency World Wide On-Line System. The primary purpose of the benchmark was to test (1) capabilities and (2) compatibility with existing equipment. Heavy reliance was placed on a technical evaluation. The Agency recognized that it could not accurately predict its workload requirements over the 8-year system life and developed a contract with provisions for expandability.

Workload projection underestimated

The experience of the Veterans Administration (VA) illustrates the difficulty and effect of unanticipated changes in the workload. The VA awarded a contract for about $38.4 million in October 1977 for a computer system to support veterans' compensation, educational benefits, and pensions. The system life was for 8-1/2 years with an option to renew for an additional 7 years. The contract also allowed the VA to purchase an additional $17.1 million worth of equipment at the discounted fixed prices. The first computer was accepted in June 1978. By May 1982, VA had already exercised its fixed-price options for about $12 million worth of additional equipment and estimated that it would spend about $100 million over the system life. According to a VA official, because there is such intense scrutiny of all ADP procurements, the agency includes in the benchmark only those requirements it can "unassailably" defend. Also, the data processing workload has increased significantly because of legislative requirements and new automation projects that were not anticipated at the time the benchmark was constructed. Developments affecting workload include:

--Passage of a new pension law.

--Passage of a new education law.

--New emphasis placed on debt collection, which requires additional ADP support.

16

--New initiatives in the veteran loan guarantee and insurance
programs.

System capacity determined by budget

In one of the procurements included in our study, the funds
available in the budget were the determining factor for sizing the
computer equipment, not the actual workload. The Department of
Commerce's Geophysical Fluid Dynamics Laboratory in Princeton, New
Jersey, awarded a contract for about $43.9 million to support the
Laboratory's atmospheric and oceanic research activities. The
laboratory did not perform a formal workload analysis when con-
structing its benchmark. Laboratory officials stated that the na-
ture of their work called for the most powerful computer available;
that their scientists would use all the computing power that was
acquired. Therefore, the limiting factor was the $45 million in
funds available over the 9-year system life. The benchmark was
used to determine the best performance (most productivity units)
over the system life. The performance or productivity units were
then compared to cost to determine the lowest cost per productive
unit. In this case, the proposal that had the best price/perform-
ance ratio (within budget) was the winning proposal.

BENCHMARKING CRITERIA ARE NEEDED

No single answer fits all situations regarding the use of
benchmarking. In fully competitive procurements, we believe bench-
marking is the most appropriate evaluation technique. In other
circumstances, more cost effective evaluation methods are indi-
cated. Because of its high cost, benchmarking should be discour-
aged for fully competitive procurements of computer equipment with
a projected system life contract value of less than $2 million.
Prudent judgement should prevail. While the Federal procurement
regulations do not require benchmarking, GSA's guidelines strongly
encourage its use for all equipment acquisitions.

We believe GSA should develop criteria that will help Federal
agencies determine (1) when it is appropriate to benchmark, (2)
if benchmarking is needed, what is the most appropriate approach,
and (3) if benchmarking is not needed, what alternative should be
used. The criteria could be incorporated into the next revision
of GSA's handbook, "Guidance to Federal Agencies on the Preparation
of Specifications, Selection, and Acquisition of Automatic Data
Processing Equipment Systems." We envision the guidance as in-
cluding a detailed presentation of the factors that influence the
benchmarking strategy discussed in this report. It should also
present other benchmarking considerations such as

--when a functional demonstration 1/ should be used rather
than a benchmark demonstration;

1/See glossary.

17

--when remote terminal emulation should be used rather than live terminals, simulation, or an internal driver; 1/

--when actual application programs should be used as opposed to synthetic programs, or scenarios; 1/

--when in the procurement cycle the benchmark demonstration should take place; that is, should all technically qualified vendors perform the benchmark demonstration or only the apparent winning vendor;

--when all or a portion of the vendor's benchmarking cost should be funded; and

--when to give credit in the proposal evaluation for improved performance during the benchmark demonstration, as opposed to a pass/fail benchmark test.

SOME PRIVATE SECTOR ACQUISITION PRACTICES CAN BE ADOPTED

In an earlier study of private sector ADP acquisition practices, 2/ we found that full and open competition is not regularly used to acquire computer equipment. Most of the private sector organizations stay primarily with one manufacturer's computer equipment, or compatible equipment. The desire to avoid software conversion costs and the availability of software packages are the reasons given for this. If the organization is willing to obtain equipment from a source other than the incumbent manufacturer, limited price competition can occur among third-party and compatible equipment vendors. We also found that benchmarking is not commonly used to select computer equipment in the private sector. Published performance data and the experience of others are more often used for validation of equipment capacity and vendor reliability.

Limited competition affects private sector benchmarking strategy

The private sector practice of allowing only certain vendors to compete for business is a major difference from Federal procurement policy, and it influences many facets of the acquisition process, including the use of benchmarking. If an organization's acquisition strategy is to stay primarily with a particular

---

1/See glossary.

2/For further discussion of private sector acquisition practices, see our study "Non-Federal Computer Acquisition Practices Provide Useful Information for Streamlining Federal Methods," AFMD-81-104, Oct. 2, 1981.

manufacturer's equipment or compatible equipment, then a benchmark is usually not needed; if a benchmark is used at all, its scope and size can be greatly reduced.

This is not true in a fully competitive environment, as is mandated for most Federal acquisitions. Benchmarking is the most common evaluation technique used in the selection of computer equipment and services by the Federal Government. That is because it is the only technique that can reliably compare performance of different computers and is generally acceptable to the vendor community. However, we believe the private sector practice of limited use of benchmarking for compatible acquisitions can be adopted by the Federal Government.

Several differences exist between private and Federal benchmarking practices

When private sector firms use a benchmark, the results are given less importance in the selection decision than they would be in a Federal procurement. The vendors we met with stated that several differences between the private sector and Federal benchmarking practices make compliance with the Federal requirements much more costly. First, private sector benchmarks tend to be much simpler and shorter in duration; they do not try to represent the entire workload and project it for 5 to 10 years. Secondly, benchmarking is a cooperative effort rather than a seemingly adversary one. The private sector customer works with the vendor to make the benchmark programs work on the vendor's equipment. Communication is always open and scheduling of the benchmark demonstration is at the mutual convenience of the parties. Thirdly, private sector firms are much more flexible in their benchmark requirements. They will grant leeway on noncritical specifications.

We believe these differences in benchmarking practices stem partly from two Government procurement requirements. The first is that each competing vendor must be treated fairly and the second is that minimum requirements must be specified.

An official of a computer equipment manufacturer's benchmark demonstration center best summarized the major differences between Federal and private sector benchmarking practices. He stated, "Federal Government benchmarks are more complex, more time-consuming, more rigid, more restrictive, and tie up more computer and people resources for longer periods of time than do commercial benchmarks, resulting in more expensive efforts to the vendor and to the Federal Government." We believe that the private sector practice of more open communication with the vendor about the benchmark can be adopted by the Federal Government. This issue is discussed in more detail in chapter 3 of this report.

CHAPTER 3
IMPROVED BENCHMARKING PRACTICES
CAN REDUCE VENDOR PROBLEMS AND COST

The use of benchmarking as a performance measurement technique has often been criticized because (1) it can place a significant resource burden on a vendor trying to compete for business; (2) it can be used by an agency to limit competition; (3) a vendor can modify the benchmark programs so that they are no longer representative of the actual workload; and (4) for services procurements, a vendor can submit unrealistically low evaluated cost proposals based on the benchmark workload, gambling that the actual workload will be substantially different. Recognizing these drawbacks, we still believe benchmarking is the best method for most fully competitive computer equipment or services procurements. Also, if agencies improve the quality of their benchmarks and are more open in communicating with vendors on benchmark-related questions, the cost to the vendors can be significantly reduced and the Government will be more likely to obtain equipment more appropriate to its needs.

BENCHMARKING ENCOURAGES
RESPONSIBLE COMPETITION
WHEN PROPERLY USED

Most of the 28 vendors we interviewed felt very strongly that there was simply no alternative better than benchmarking for evaluating proposals in a fully competitive procurement. They feel that by using a benchmark the Government assures itself that it is getting what it asks for--even if not necessarily what it needs. Vendors believe that in most cases the benchmark encourages responsible competition. However, many vendors were also concerned about "sham" competitions and about the practice of some computer services vendors who submit proposals that are evaluated at an extremely low cost but turn out to be much more costly after installation.

Benchmark usually has little impact
on vendor decision to compete,
or on proposal price

In most cases, a vendor's decision to compete for Government business is based on several factors. Rarely, if ever, is the decision not to compete based solely on the benchmark requirement. The probability of the vendor's winning the award far outweighs the benchmark cost in the decision to bid. Questions a vendor must answer in deciding whether to compete include:

   --Does the Government's requirement fit within the vendor's product line?

   --What is the incumbent vendor's advantage (conversion costs)?

   --What other vendors are likely to compete and what are they likely to propose?

--What is the cost to compete (including the benchmark) versus the potential profit?

--Does the vendor have available resources (people and facilities) to respond to the procurement within the Government's specified time constraints?

In some instances benchmarking can discourage competition. For example, in the very large-scale, scientific, "super-computer" environment only two firms, Cray and Control Data, can realistically compete for business. Cray manufactures only the super-computer, which requires an auxiliary computer to handle data input and output. The auxiliary computer is normally a large-scale one and can be of any manufacture. If the agency requires a total system and/or uses remote terminal emulation, 1/ the probability is very high that Cray will not compete for the business because it does not have the resources readily available to perform these types of benchmarks.

In the IBM-compatible environment, too costly a benchmark may discourage firms such as Amdahl and ViON from competing. These firms do not have large benchmark facilities and do not manufacture total systems. To illustrate, even though Amdahl was the incumbent CPU vendor, the firm decided not to compete for Tennessee Valley Authority's recent $30.4 million procurement. The decision was based primarily on the anticipated cost of the benchmark and the impact the benchmark would have on other sales (lost opportunity costs). Amdahl projected that it would lose 17 cents per share on its stock by tying up its benchmark demonstration facility, which is also used for predelivery testing and future products research. Amdahl believes that the "commercially unreasonable benchmark" eliminated all competition in the procurement. For computer equipment that has been in the marketplace for some time, a potential source for the equipment is third-party vendors. However, these vendors usually will not compete if benchmarking is required. Some equipment manufacturers will not compete for low dollar value computer procurements if benchmarking is required. This is because the cost of benchmarking would greatly reduce or eliminate potential profits.

The vendor's cost to perform the benchmark is considered a cost of doing business and generally is not reflected in the vendor's proposal price. One vendor we interviewed stated that while the proposal budget has no line item that states benchmark, the benchmark's cost can affect the percentage of discount given. All other vendors we met with stated that benchmark costs were overhead and were not directly reflected in proposal prices.

We believe that eventually a vendor's benchmark costs must be reflected in equipment prices, as all overhead is ultimately recouped through sales. Most equipment manufacturers would have one or more facilities to demonstrate their products and the issue that

_____

1/See glossary.

remains is "How much do the Government's benchmark requirements drive up vendor prices?" If the Government's benchmark requirements do affect prices, the biggest impact is on the prices the private sector pays for computer equipment or for equipment purchased through sole-source procurement. This is because of (1) the substantial discounts on equipment that the Government achieves through competition, and (2) the burden that Government procurements place on a vendor's benchmark facilities. To illustrate, during fiscal 1981 about 50 percent of the benchmarking resources of a major manufacturer were dedicated to Federal procurements. Federal business accounted for only about 7 percent of the company's revenue. Other companies reported similar data.

"Sham" competitions are a vendor concern

Some vendors we met with were concerned about "sham" competitions in which they allege the agency has already decided which vendor will be awarded the contract when competition begins. The vendors claim that one method is to use the benchmark to purposely limit competition or to preselect the winning vendor. As an example of such allegations of preselection, one vendor cited a recent procurement by a Department of Energy management contractor for its Savannah River Laboratory. The solicitation document required that the processor capacity be rated equal to or greater than 1.7 times an IBM 3033 computer's capacity, and a $1 million credit would be given for 20 percent or more performance improvement in the benchmark demonstration. This equates to 2.04 times faster than the IBM 3033 computer. Five months before the solicitation was issued, IBM announced its new 3081 computer. The product announcement stated that the 3081 computer would be 2.0 to 2.1 times faster than the 3033 for scientific batch applications. Energy officials, after reviewing a bid protest from another vendor, determined that the benchmarking was justified, that the specifications were not unduly restrictive, and that real competition existed. We note that competing vendors could overcome the $1 million credit by providing a larger discount; however, the credit proved to be too formidable an advantage for IBM.

Although agency benchmarks can be consciously or unconsciously slanted toward a specific vendor, benchmarking is usually perceived to be fair. A vendor can usually detect a slanted benchmark only after it has expended some resources reviewing the solicitation document and benchmark requirements. In our view, using the benchmark to preselect a vendor is not in the best interest of the Government or the computer industry.

Vendor "gaming" of benchmark can raise costs

A vendor can use the benchmark to its advantage and to the detriment of the agency conducting the procurement. This is called "gaming" the benchmark and it can be done by either of two approaches. The first approach can be used in both equipment and services procurements. The second can be used only in services procurements. If not detected, either can cause higher costs for equipment or services than were anticipated by the agency. We have no basis on which to judge how widespread the practice of gaming is, but it does occur.

In the first approach, the vendor changes the benchmark programs beyond that allowed by the agency, so that they perform much better than the actual programs will on the vendor's computer. In extreme cases, the vendor could change the compiler 1/ to make it recognize a particular sequence of instructions in the benchmark programs and, upon recognition, provide efficient code 1/ specifically designed to perform the function desired with much greater efficiency. The result could be that the vendor's proposed computer equipment, if acquired by the agency, would not process the actual workload as efficiently as it did during the benchmark and may have to be augmented. In the case of teleprocessing services contracts, the agency might have to pay more than was anticipated for computer services. We found only one instance where a vendor was disqualified from a competition for making substantial changes to the benchmark programs.

The second gaming approach involves the use of "creative pricing strategies" by TSP vendors tailored to the benchmark. If the benchmark 2/ is inaccurate--even slightly--in terms of workload projection, workload mix (processing, storage, connect), or resources used, the actual cost billed under the contract may be much higher than the evaluated cost used in the selection decision. The creative pricing includes credits, decreasing volume discounts, and bundled unit pricing. Of the 15 services procurements we included in our study, the winning vendor used a creative pricing strategy to win the award in 6 of them, and all but 2 are experiencing high cost overruns. The agencies not yet experiencing high cost overruns are the Department of Labor and the Army Corps of Engineers. Labor's contract calls for a fixed price for a fixed amount of computer resources (CPU minutes). Labor has stayed within the fixed amount of CPU minutes. The Army Corps of Engineers found that subsequent to contract award in July 1982, the winning vendor had submitted an apparently unbalanced proposal. The Corps is trying to negotiate a change to the vendor's billing algorithm 1/ before the new contract takes effect.

The experience of the Army and the Navy illustrates the impact vendor gaming can have on actual costs. Both services have acquired teleprocessing support from Boeing Computer Services Company in connection with recruiting efforts. 3/ Both are experiencing very high cost overruns. The Army's initial cost projection for the Boeing proposal was about $8.5 million for the 60-month life of the contract; it now projects a cost of about $120 million.

1/See glossary.

2/In a TSP procurement the benchmark is used to estimate costs.

3/For further discussion of these procurements see our report "Teleprocessing Services Contracts for the Support of Army and Navy Recruitment Should Be Recompeted," AFMD-82-51, March 24, 1982.

The Navy's initial cost projection for the Boeing proposal was about $524,000 for the 42-month life of the contract; it now projects a cost of about $13 million. Both services used benchmarks to evaluate proposals. Neither benchmark adequately represented the actual workload subsequently placed on the system and so was a poor indicator of system life costs.

Computer resources used by both the Army and the Navy greatly exceeded the amount anticipated. Because Boeing submitted a proposal in which commercial rates were charged for teleprocessing services beyond the projected level, both services incurred costs well beyond that expected. This constituted an unbalanced proposal.

GSA has recognized that unbalanced proposals are a problem in the procurement of teleprocessing services. In September 1981, GSA amended the Teleprocessing Services Program Basic Agreement 1/ to include specific clauses that would limit the probability of receiving unbalanced proposals. GSA also issued a report that provided recommendations on

--techniques for evaluating proposals when the workload is not defined,

--ways to contractually minimize unbalanced proposals, and

--ways to recognize unbalanced pricing structures.

The report recognized that when the agency cannot specify the workload, unbalanced pricing can lead to a disproportionate increase in the evaluated prices.

We do not see an alternative to benchmarking for TSP procurements as the program is currently structured. There is no better way to estimate costs because (1) some vendors' billing algorithms 2/ are so complex and (2) some vendors have not released their billing algorithms. Agencies must clearly understand the importance of the benchmark in a TSP procurement. The benchmark not only determines whether the vendor can satisfy an agency's requirements, it

---

1/The Basic Agreement is an agreement between GSA and a number of teleprocessing services vendors. It contains standard provisions, other than technical or cost, that apply to future procurements.

2/TSP vendors charge for their services based on algorithms which consider computer resources used--such as CPU, connect time, and storage. A billing algorithm can have 20 or more components. The algorithms are proprietary and usually are not made available to the customers.

also determines cost and the ultimate winner in most cases. Because of the importance of the benchmark as an evaluative tool and the pricing strategies that have been proposed by the TSP vendors, agencies must take steps to ensure that the benchmark is representative of their processing requirements.

AGENCIES SHOULD EMPHASIZE
BENCHMARK DEVELOPMENT
AND OPEN COMMUNICATION WITH VENDORS

Vendors we interviewed generally shared the view that Federal benchmarks are poorly prepared and documented. Vendors also say they find it difficult to communicate with agencies as they attempt to resolve problems stemming from the poor benchmarks. If agencies can improve practices in these two areas, the burden (cost and time) of benchmarking on the vendor can be greatly reduced.

Benchmarks should be thoroughly
tested and documented

One of the most significant problems cited by vendors we interviewed was the poor construction and documentation of Federal agency benchmarks. One vendor representative stated that in the 4-1/2 years he had managed the company's benchmark support group, no Government benchmark has run right the first time. Similar comments were made by many of the vendors.

NBS guidance on constructing benchmarks states that the benchmark should be tested on one or more systems other than the one for which it was developed. However, few agencies included in our study followed this advice.

Testing the benchmark on other computers can provide valuable information on the benchmark's portability. It can also help to determine the correctness and clarity of the benchmark and its supporting documentation. For example, errors introduced into a benchmark package commonly include incorrectly generated benchmark tapes, inconsistencies between the benchmark programs and the accompanying documentation, incomplete documentation, and even program logic errors. Such errors are likely to be detected if the benchmark is tested on another computer system, especially if performed by personnel other than those who constructed the benchmark.

The experience of one minicomputer manufacturer illustrates the problems a vendor can encounter in responding to a Federal benchmark requirement. The Federal Communications Commission solicited proposals from seven vendors for a minicomputer. The minicomputer was needed quickly to process radio frequency data that were to be presented at an international broadcasting conference. Because of the immediate need for the computer, five of the vendors were given less than 3 weeks (Dec. 29, 1980, to Jan. 15, 1981) to complete the benchmark demonstration. One vendor representative who directed his company's benchmark support group

stated that the first benchmark tape received from the Commission could not be read. The vendor then obtained another benchmark tape, but could not get it to run. The vendor decided to withdraw from the competition because insufficient time remained to resolve the benchmark problems.

In response to an Air Force survey of benchmarking approaches, 1/ a major computer equipment manufacturer stated that while the benchmarking approach is not material, improvements could be made to the process that could aid the vendors. These included the following:

--The programs and input data should be free of errors and agree with the documentation.

--The expected output should be provided and should be correct.

--Documentation should be furnished describing the function of a program and its subprograms. In complex programs, flow diagrams should be provided.

--Accurate formats of data and program tapes should be provided.

--In general, the program and data tapes should be in a simple, sequential format.

--Statistics on the live test demonstration should be provided to help in sizing and identifying errors. Some examples of useful statistics are CPU time, input/output time, lines printed, cards read, cards punched, terminal characters input, and terminal characters output.

--When vendor system functions or assembly language 2/ programs are included, a complete description of the functions and input/output parameters should be provided.

--It should not be mandatory that the benchmark programs be converted to process exactly as they were processed on the incumbent's equipment. The vendor should be required to perform the function to provide the desired result.

1/In March 1981, the Air Force solicited industry comments on four approaches to the use of live test demonstrations in the selection of computers. Two of the approaches were to (1) require each vendor who is determined to be in the competitive range to perform the benchmark demonstration before contract award; and (2) require the "apparent winner" to perform the benchmark demonstration after submission of the best and final offer, but before contract award.

2/See glossary.

--The time required to perform the steps in a solicitation could be shortened. For instance, the current sequence is (1) Request-for-Proposal release, (2) benchmark materials release, (3) proposal submission, and (4) live test demonstration. In spite of this, the final configuration submitted with the proposal depends on the live test demonstration. Only rarely does the benchmark demonstration precede proposal submission. It would be of great benefit if the benchmark materials could be released before the request for proposal. At a minimum, this would allow conversion of the benchmark programs to be started earlier.

We agree with the NBS instructions that Federal agencies should test the benchmark on at least one computer other than the one on which it was developed and the test should be performed by an independent group to ensure that the benchmark instructions and documentation are complete and consistent. Adequate guidance exists but agencies are not following it. If an agency does not have sufficient talent in-house to develop high quality benchmarks, it can and should obtain assistance from recognized centers of benchmarking expertise such as FEDSIM or the Air Force Computer Acquisition Center.

More effective communication is needed between agency and vendor

Agency and vendor exchange of technical information is one of the most important ways to minimize benchmark discrepancies. It maximizes competition and reduces the time and cost of an acquisition for both the agency and the vendor. However, most vendors we interviewed cited the difficulty of communicating with the agency to resolve technical benchmark issues. By limiting or complicating communication, the agency is enhancing the already favorable advantage of the incumbent vendor as well as increasing the costs of all vendors participating in the procurement.

The vendors' point of contact is normally the agency's contracting officer, who usually is not able to answer technical benchmark questions. Also, vendors are usually required to put their questions in writing and wait for a written response. The questions and answers are distributed to all vendors participating in the procurement. The cycle can take up to a month and can affect a vendor's ability to respond to the benchmark within the agency-specified time. Because of the lack of standard terminology in the computer industry, the agency's written response could lead to the vendor asking another question, thus beginning another cycle. According to GSA officials, opening up the communication between the procuring agency and the vendor after the solicitation document has been released would only lead to more bid protests. The GSA officials believe that the vendors would try to influence the outcome of the award if communication were open.

The experience of the Internal Revenue Service illustrates that communication on technical benchmark issues can be open without compromising the procurement. The Service awarded a contract for about $102.5 million in June 1981 for computers to replace aging equipment in its service centers. The Service had an open line to the vendors to resolve technical questions regarding the benchmark. The vendors were required to follow up verbal questions in writing. When the vendor failed to follow up in writing, the Service did it for them and made the questions and answers available to all competing vendors. Part of the training given to the benchmark team was how to deal with vendor questions without compromising the procurement. We believe agencies can be far more open in their response to vendors' technical questions without compromising the procurement. Effective communication is essential to increasing competition and reducing costs. However, if agencies placed greater emphasis on producing high-quality benchmarks fewer questions would be asked by vendors.

CHAPTER 4

CONCLUSIONS, RECOMMENDATIONS,

AND AGENCY COMMENTS

CONCLUSIONS

Benchmarking represents a significant cost to both the Government and the computer industry. However, there are few alternatives in the competitive ADP procurement environment.

The private sector practice of limiting the vendors who may compete for specific business is different from that of the Federal Government, which emphasizes open competition. This major difference influences many facets of the acquisition process, including the use of benchmarking. However, some private sector acquisition practices can be adopted by the Government to reduce the benchmarking burden on the vendors.

Federal agencies should place greater emphasis on developing well documented, representative, and reasonable benchmarks. Benchmarks should be as simple as possible without compromising requirements or unduly increasing risk to the Government that the delivered computer equipment will not have sufficient capacity to process the workload or that services will be obtained at unreasonable rates. This is a very difficult undertaking which requires the agency's most talented computer specialists, and perhaps assistance from other organizations with benchmarking expertise.

The greater the effort agencies put into the benchmark, the easier and less costly it will be for the vendor to perform the benchmark demonstration and the higher the likelihood that the Government will obtain computer equipment or services appropriate to its needs. While in some instances benchmarking can discourage competition, most vendors feel strongly that no better method exists for evaluating proposals in a fully competitive procurement and that benchmarking encourages responsible competition.

Federal procurement regulations do not require benchmarking, although GSA's guidelines strongly encourage its use for all equipment acquisitions. Available guidance provides little help to agencies for determining when a benchmark is needed, what benchmarking approach is best, or what alternatives are suitable. Agencies need guidelines that include a detailed presentation of the factors influencing benchmarking strategy, as discussed in this report.

RECOMMENDATIONS

To improve the acquisition of computer equipment, we recommend that the Administrator of General Services:

--Develop criteria that will help Federal agencies determine (1) when it is appropriate to benchmark, (2) if benchmarking

is needed, what approach is most appropriate, and (3) if
benchmarking is not needed, what alternative should be used.

--Revise the Federal Procurement Regulations (FPR 1-4.1109-22)
so that benchmarking is discouraged for computer equipment
procurements with a projected system life contract value of
less than $2 million.

AGENCY COMMENTS

In commenting on our draft report (see app. III), the Deputy
Administrator, General Services Administration concurred with our
recommendations. He stated that GSA does not have the breadth of
technical expertise required to develop the recommended guidelines,
but will rely on outside assistance, probably the Federal Computer
Performance Evaluation and Simulation Center, to develop them.

He also stated that our second recommendation to modify the
Federal Procurement Regulations should await technical investiga-
tion to determine under what circumstances benchmarking benefits
outweigh risk, cost, and competitive disadvantage. We believe that
further investigation is not needed to make the recommended regu-
lation change. The data gathered as a result of our study support
the raising of the threshold from $300,000 to $2 million for dis-
couraging benchmarking. Should further investigation indicate that
additional regulation changes are needed, they can be made at that
time.

ADP PROCUREMENT ACTIONS

REVIEWED DURING BENCHMARK STUDY

The 73 ADP acquisitions we reviewed during our study of bench-marking practices included:

--32 competitive medium- and large-scale computer procure-ments,

--14 competitive minicomputer procurements,

--12 procurements limited to IBM-compatible computer equip-ment, and

--15 procurements for teleprocessing services.

The tables in this appendix detail the system life contract value at time of award, system life, agency benchmark cost, competing vendors, and vendor benchmark costs for each of the procurement actions included in our review. For teleprocessing services procurements, we have included the projected system life contract value after installation as well as the contract value estimated at the time of award because the benchmark is used to determine cost in those procurements. Agency benchmark costs do not include in-house computer time to develop or test the benchmark. Vendor benchmark costs usually do include a cost for computer time. Only the winning vendor is identified. The number of other vendors who also performed the benchmark demonstration is shown.

FULLY COMPETITIVE PROCUREMENTS
(Medium- and large-scale computers)

Agency (Procurement)	Contract value	Competing vendors	System life	Agency benchmark cost	Vendor benchmark cost (note a)
	(millions)		(years)		
Energy Dept. Richland Computer Center	$10.4	Univac 1 other	6	$ 51,157	$500,000 to 999,000
Geophysical Fluid Dynamics Laboratory	43.9	Control Data 1 other	9	79,100	-
Navy Fleet Numerical Oceanographic Center	14.2	Control Data 1 other	10	87,610	-
Bureau of Government Financial Operations	21.3	Honeywell 2 others	8	123,935	250,000 100,000 to 499,000
Goddard Space Flight Center (Mgt. Services)	1.1	IBM 1 other	7	14,721	-
Goddard Space Flight Center (GLAS)	31.2	Control Data 1 other	6	11,000	-
Goddard Space Flight Center (Dynamic Explorer)	1.6	IBM 2 others	5	No bench-mark	-
Treasury Dept. (Service Center)	8.7	Univac 1 other	6	168,750	100,000 to 499,000
Air Force (IDHS-80)	17.3	IBM 3 others	6 plus two 3-yr. options	305,608	395,000 100,000 to 499,000

a/Not all vendors provided benchmark costs. The costs shown are representative vendor bench-mark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

Agency (Procurement)	Contract value	Competing vendors	System life	Agency benchmark cost	Vendor benchmark cost (note a)
	(millions)		(years)		
Tennessee Valley Authority	$ 30.4	IBM	6	$ 135,400	$ —
Forest Service, Ft. Collins, CO	b/11.5	Univac	6	73,265	500,000 to 999,000
Bureau of Indian Affairs	12.1	Burroughs 1 other	6	227,094	50,000 to 75,000
Naval Underwater System Center	42.8	Univac	10	107,600	500,000 to 999,000
Customs Service Washington, DC	9.7	IBM 2 others	8	264,560	50,000 to 75,000 250,000
Jet Propulsion Laboratory	14.5	Univac 1 other	10	136,000	500,000 to 999,000
Sandia Laboratory Livermore, CA	12.2	Cray 1 other	8	129,475	—
Federal Railroad Administration	1.5	Honeywell 1 other	7	c/No benchmark	—
Internal Revenue Service	102.6	Univac 1 other	8	2,054,010	1,000,000 1,500,000 to 1,999,000

a/Not all vendors provided benchmark costs. The costs shown are representative vendor bench-mark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Projected. The Department of Agriculture could not provide documentation supporting the evaluated contract value at time of award.

c/Required a functional demonstration.

Agency (Procurement)	Contract value	Competing vendors	System life	Agency benchmark cost	Vendor benchmark cost (note a)
	(millions)		(years)		
Veterans Adminis- tration (Target)	$ 38.4	Honeywell 1 other	8-1/2 plus 7- yr. opt.	$503,000	$ 250,000 1,000,000
Bureau of Economic Analysis	2.8	Honeywell 3 others	7	89,806	250,000 100,000 to 499,000
Patent and Trademark Office	1.1	Burroughs 4 others	5	73,400 .	250,000 50,000 to 75,000 100,000 to 499,000
Housing and Urban Development (Service Center)	17.7	Univac 1 other	7	90,264	1,000,000 500,000 to 999,000
Lawrence Livermore Laboratory (Admin- istrative)	1.9	Univac 4 others	6	91,944	750,000 50,000 100,000 to 499,000
Lawrence Livermore Laboratory (Scien- tific #1)	12.5	Cray	8	50,255	–
Lawrence Livermore Laboratory (Scien- tific #2)	12.9	Cray	5	53,107	–
Lawrence Livermore Laboratory (Scien- tific #3)	12.9	Cray	5	47,224	–

a/Not all vendors provided benchmark costs. The costs shown are representative vendor bench-
  mark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

Agency (Procurement)	Contract value	Competing vendors	System life	Agency benchmark cost	Vendor benchmark cost (note a)
	(millions)		(years)		
Lawrence Livermore Laboratory (Scientific #4)	$ 15.5	Cray	5	$ 77,844	$ –
Lawrence Livermore Laboratory (Scientific #5)	36.7	Cray	5	No benchmark	–
Air Force Military Personnel Center	50.9	Honeywell 2 others	8 plus 4-yr. opt.	231,107	500,000 2,000,000 to 2,499,000
Army (Project VIABLE)	616.4	EDS 1 other	10	386,000	c/ 100,000
Naval Surface Weapons Center	(b)		10	d/207,670	–
Sandia Laboratory Albuquerque, NM	14.1	Cray 1 other	7	57,000	–

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Pending award.

c/Each vendor was funded $5.1 million to demonstrate the operational capability of the proposed system. The amount shown reflects the vendor's benchmark cost to be eligible to participate in the compute-off.

d/The amount shown reflects agency benchmark costs up to the live test demonstration.

FULLY COMPETITIVE PROCUREMENTS
(Minicomputers)

Agency (Procurement)	Contract value (millions)	Competing vendors	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
Army (JACS) Military Pay	$ 22.7	Honeywell 1 other	8	$ 56,790	$ 50,000 750,000
Treasury Dept., International Div.	0.2	Burroughs 1 other	1-2/3 mos.	980	-
Army Corps of Engineers (CPMS)	7.5	Harris	3	5,527	-
Federal Compiler Testing Service	0.2	(b)	5	1,064	-
Army (DAS-3)	130.0	GE 4 others	8	239,000	200,000 170,000
Marine Corps (ADPE-FMF)	43.7	IBM 1 other	8	47,120	1,000,000
Labor Dept., Productivity and Technology	0.6	Harris	5	7,914	-
Defense Communications Agency (WWOLS)	6.3	IBM	8	5,938	-
Dept. of Housing and Urban Development (Network)	30.7	Honeywell 1 other	8	115,000	1,000,000

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Relied on benchmark results from the Army DAS-3 procurement. The Federal Compiler Testing performed the compiler validation of the competing vendors in that procurement.

Agency (Procurement)	Contract value	Competing vendors	System life	Agency benchmark cost	Vendor benchmark cost (note a)
	(millions)		(years)		
Postal Service, (DDP Prototype)	.9	Digital 2 others	7	$ 22,990	$125,000
Federal Communications Commission	(b)	Digital 1 other	1-1/2	3,273	–
Defense Dept. (TRI-RAD)	11.5	National Computer Systems 1 other	8	98,718	40,000
Defense Dept. (WWMCCS Graphics)	19.9	AYDIN Corp. 3 others	7	77,902	28,900 75,000 39,000
Navy (SNAP I, Phase II)	602.9	Honeywell 2 others	10 plus 10-yr. opt.	445,321	(c)

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Because of benchmark results, an award was not made. Agency entered into an agreement with vendor to lease time on a computer system used for customer demonstrations.

c/Each qualified vendor was authorized to receive $1 million to perform the benchmark demonstration. One of the three vendors declined to accept the money.

COMPETITION LIMITED TO IBM-COMPATIBLE COMPUTERS

Agency (Procurement)	Contract value (millions)	Competing vendors	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
NASA, Langley (Administrative System)	$ 0.5	Control Data 3 others	5	$ 7,688	$ 30,000
Marine Corps (Service Center Replacement)	19.1	Federal Data 2 others	6	108,236	-
Securities and Exchange Commission	0.4	SMS 1 other	6	2,757	b/ 5,000
Army Corps of Engineers (CROHMS)	3.8	Federal Data 1 other	8	148,700	110,000
Goddard Space Flight Center (Scientific)	7.2	IBM 2 others	8	c/ 24,162	75,000
National Institutes of Health	92.3	IBM	5 plus 5-yr. opt.	113,168	-

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Only manufacturers other than IBM were required to perform benchmark demonstrations.

c/No live test demonstration was held.

Agency (Procurement)	Contract value (millions)	Competing vendors	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
Stanford Linear Accelerator Center	$ 8.9	IBM 2 others	5	$ 41,361	$75,000
Air Force, Tinker AFB	11.9	Federal Data 1 other	8 plus 7-yr. opt.	191,526	—
Air Force (TRICOMS III)	10.7	Federal Data 1 other	4 plus 4-yr. opt.	No benchmark	—
Naval Postgraduate School	9.9	IBM	8	91,800	—
Postal Service, St. Louis, MO	5.0	IBM 4 others	7	No benchmark	—
Postal Service, San Bruno, CA	4.5	Amdahl 3 others	5	No benchmark	—

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

TELEPROCESSING SERVICES

Agency (Procurement)	Estimated costs at time of contract award	Competing vendors	Projected costs after installation	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
Social Security Administration (OPO)	$ 1,838,258	CSC 6 others	$12,500,000 to 17,500,000	5	$112,731	$200,000 80,000 to 100,000 187,000
National Defense University	0	BCS 7 others	700,000	5	27,860	140,000 28,000 80,500 34,000
Pacific Southwest Forest and Range Experiment Station	480,000	ISD 2 others	288,000	4	No benchmark	–
Public Health Service (PHAMIS)	2,909,661	MMDS 3 others	(b)	5	107,345	250,000 215,000
Health and Human Services (HEWCAS)	689,342	CSC 5 others	586,260	5	16,569	28,000 17,000 36,000

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/Public Health Service accepted equipment or teleprocessing services proposals. The program the computer system was designed to support was cut from the Federal budget.

Agency (Procurement)	Estimated costs at time of contract award	Competing vendors	Projected costs after installation	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
GAO (Audit)	$ 726,102	COMNET 8 others	$ 635,352	3	$ 68,142	$ 12,500 150,000 60,000 28,000
Navy Recruiting Command (PRIDE)	523,969	BCS 5 others	13,000,000	3-1/2	42,208	60,000 27,300 80,000 to 100,000 45,000
Mare Island Naval Shipyard	470,400	UIS 2 others	800,000	3	7,700	30,000 to 50,000
Sacramento Air Logistics Center	527,957	CSG 3 others	500,000	1	No benchmark	–
Army Corps of Engineers (Nationwide Engineering Support)	25,640,803	BCS 3 others	(b)	5	69,470	30,300

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/After contract award in July 1982, the Corps of Engineers determined that the winning vendor had submitted an apparently unbalanced proposal. The Corps is negotiating a change to the vendor's billing algorithm.

Agency (Procurement)	Estimated costs at time of contract award	Competing vendors	Projected costs after installation	System life (years)	Agency benchmark cost	Vendor benchmark cost (note a)
Congressional Budget Office	$ (b)	SBC	$ 4,500,000	3	$ 7,860	$ —
Labor Dept. (Admin. Services)	4,800,000	BCS	c/4,800,000	4	327,283	—
Agriculture Economic and Statistical Services	9,425,275	MMDS 1 other	16,000,000	5	53,414	500,000 212,000
Army Military Personnel Center (Request-Retain)	8,493,934	BCS 3 others	120,000,000	5	87,450	160,000 80,000 to 100,000 34,600
Housing and Urban Development	8,829,330	Datacrown 7 others	12,115,038	6	10,000	4,000

a/Not all vendors provided benchmark costs. The costs shown are representative vendor benchmark costs and do not necessarily reflect the costs of the corresponding vendor in col. 3.

b/The agency renegotiated an existing contract. The benchmark was used to specify and monitor performance.

c/Labor's contract calls for a fixed price for a fixed amount of resources (CPU minutes).

Congress of the United States

House of Representatives jc

Committee on Appropriations

Washington, D.C. 20515

January 23, 1981

Honorable Elmer B. Staats
Comptroller General of the United States
General Accounting Office
Washington, D. C. 20548

Dear Mr. Staats:

During the past year, the Financial and General Management Studies Division completed a review for the Committee on Appropriations concerning automatic data processing acquisitions, emphasizing conversion and related costs. This report has been extremely helpful to the Committee in its continuing study of ADP procurement practices in the Federal government. Like conversion costs, "benchmarking" represents another significant cost in the ADP procurement process which can, from time to time, be so substantial as to discourage competition. Also, benchmarking costs are often hidden and lead to higher equipment costs as vendors try to recoup this government-directed overhead expense.

It would be appreciated if the General Accounting Office would study the use of the benchmark process in Federal ADP procurement. This study should assess the costs of performing benchmarks in typical Federal and private sector ADP procurements and determine the need for, cost effectiveness of, and the alternatives to the benchmark process.

There is some concern that the government simply does not have any idea of the cost of carrying on the benchmark process or whether this cost can be justified in terms of reducing future operating difficulties.

Your cooperation would be appreciated.

Sincerely,

Chairman

AGENCY COMMENTS

GSA General
Services
Administration   Washington, DC 20405

_____

AUG 15 1982

Honorable Charles A. Bowsher
Comptroller General of the
United States
U.S. General Accounting Office
Washington, DC   20548

Dear Mr. Bowsher:

Thank you for affording us the opportunity to review the draft
of your proposed report entitled "Benchmarking: Costly and
Difficult, but Often Necessary when Buying Computer Equipment or
Services."

It is a well organized and thorough document which clearly
identifies the positive and negative aspects of benchmarking and
highlights the fact that each circumstance must be evaluated on
its own merit.

We support your proposed recommendations. However, GSA does not
have the breadth of technical expertise required to cover every
point of your recommendations. In order to develop the necessary
guidelines that will help Federal agencies to determine their
need for benchmarking as well as the most appropriate approach
and possible alternatives, we will have to rely on some outside
capabilities. The Federal Computer Performance Evaluation and
Simulation Center (FEDSIM), operated by the Air Force under GSA
delegation, is a likely source for this assistance.

Cost and time required to incorporate the proposed guidelines in
GSA's handbook, "Guidance to Federal Agencies on the Preparation
of Specifications, Selection, and Acquisition of Automatic Data
Processing Equipment Systems", can only be estimated after the
technical problems have been addressed. Also, the revision of
Federal Procurement Regulation (FPR) 1-4.1109-22 to modify
existing policy should await technical investigation to determine
under which circumstances benchmarking benefits outweigh risk,
cost, and competitive disadvantages.

Sincerely,

Ray Kline
Deputy Administrator

(913673)

23364.

AN EQUAL OPPORTUNITY EMPLOYER

UNITED STATES
GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300