of Commerce

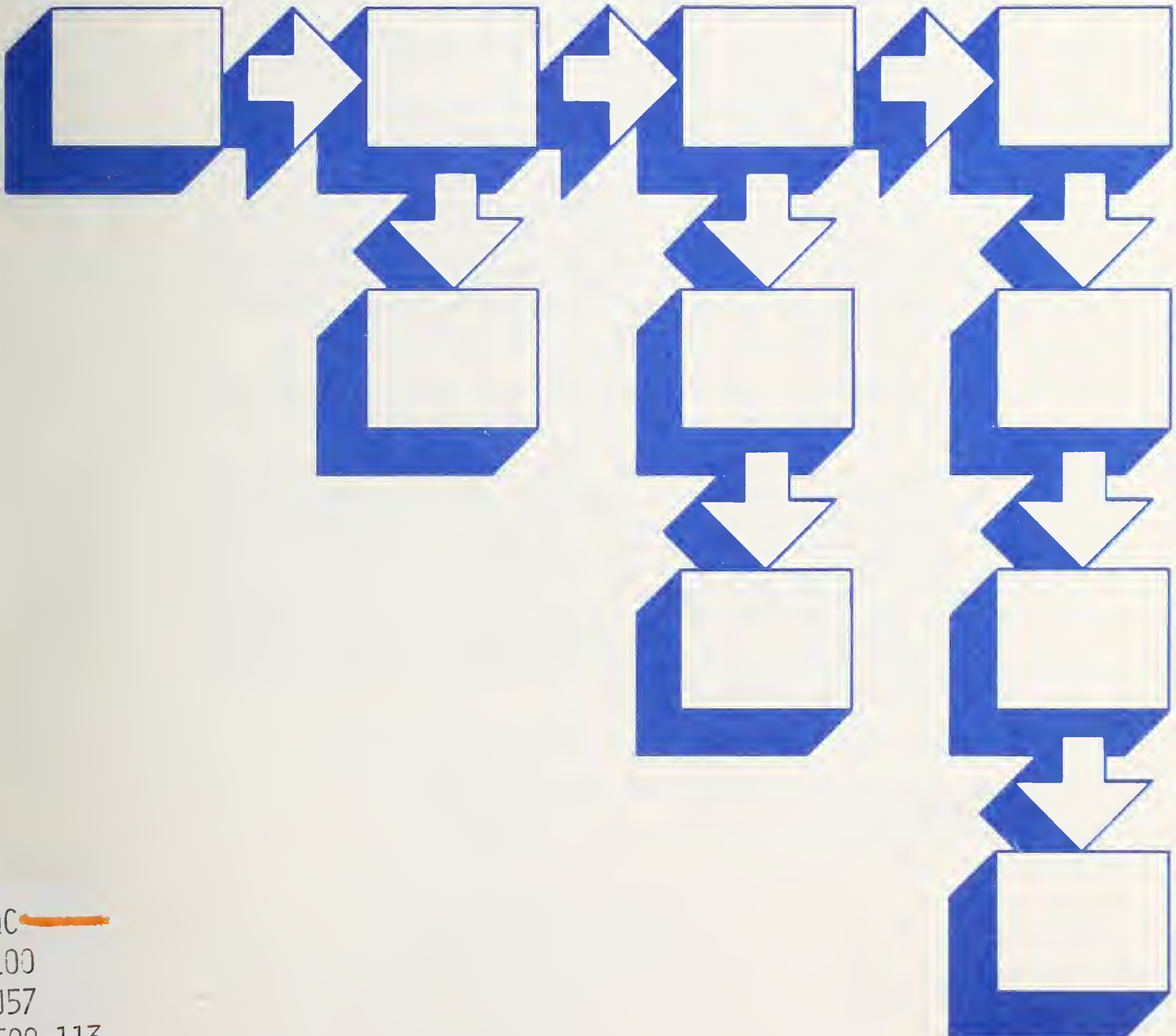
National Bureau
of Standards

Computer Science and Technology

NBS PUBLICATIONS

NBS Special Publication 500-113

Assessment of Techniques for Evaluating Computer Systems for Federal Agency Procurements



NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress on March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

THE NATIONAL MEASUREMENT LABORATORY provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities[2] — Radiation Research — Chemical Physics — Analytical Chemistry — Materials Science

THE NATIONAL ENGINEERING LABORATORY provides technology and technical services to the public and private sectors to address national needs and to solve national problems; conducts research in engineering and applied science in support of these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering[2] — Manufacturing Engineering — Building Technology — Fire Research — Chemical Engineering[2]

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides scientific and technical services to aid Federal agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following centers:

Programming Science and Technology — Computer Systems Engineering.

[1]Headquarters and Laboratories at Gaithersburg, MD, unless otherwise noted;
mailing address Washington, DC 20234.
[2]Some divisions within the center are located at Boulder, CO 80303.

Computer Science and Technology

NBS Special Publication 500-113

Assessment of Techniques for Evaluating Computer Systems for Federal Agency Procurements

Helen Letmanyi

Center for Programming Science and Technology
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, DC 20234

Based in Part on the Contract
Report (NB81SBCA0810)
of the MITRE Corporation

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Bureau of Standards
Ernest Ambler, Director

Issued March 1984

Reports on Computer Science and Technology

The National Bureau of Standards has a special responsibility within the Federal Government for computer science and technology activities. The programs of the NBS Institute for Computer Sciences and Technology are designed to provide ADP standards, guidelines, and technical advisory services to improve the effectiveness of computer utilization in the Federal sector, and to perform appropriate research and development efforts as foundation for such activities and programs. This publication series will report these NBS efforts to the Federal computer community as well as to interested specialists in the academic and private sectors. Those wishing to receive notices of publications in this series should complete and return the form at the end of this publication.

Library of Congress Catalog Card Number: 84-601012
National Bureau of Standards Special Publication 500-113
Natl. Bur. Stand. (U.S.), Spec. Publ. 500-113, 34 pages (Mar. 1984)
CODEN: XNBSAV

U.S. Government Printing Office
Washington: 1984

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402

TABLE OF CONTENTS

ASSESSMENT OF TECHNIQUES FOR EVALUATING COMPUTER SYSTEMS FOR FEDERAL AGENCY PROCUREMENTS

Helen Letmanyi

ABSTRACT

The primary purpose of this document is the identification and qualitative assessment of computer system evaluation techniques for use during acquisition of computer systems. Also addressed is the identification of several criteria by which these alternative evaluation techniques may be compared and selected. The concepts presented in this study are applicable to all sizes of general purpose computers, from microcomputers to mainframes. Embedded or single-purpose computers, such as those used in weapon systems, have been excluded.

Keywords: Acquisition; benchmarking; evaluation; instruction timing analysis; modeling; prototyping; rating charts analysis; system selection.

1. INTRODUCTION

1.1 Purpose

The primary purpose of this report is the identification and qualitative assessment of computer system evaluation techniques for use during acquisition of computer systems. Also addressed is the identification of several criteria by which these alternative evaluation techniques may be compared and selected. A future NBS guideline will address related issues dealing with acquiring computer services.

Within the general goal of obtaining and managing the most suitable and cost-effective computer systems to meet users' requirements, evaluation techniques may be used for several reasons. They include:

1. Determination of whether a candidate system can meet the specified functional and performance requirements for the anticipated workload. The performance requirements are usually expressed by such attributes as:

   (a) response time (a specified time in which a minimum percentage of responses are made under specified conditions);

   (b) maximum time to process a specified workload;

   (c) workload processed in a given time.

2. Determination of the amount of additional capacity, beyond the stated requirements, that is available on a proposed system. Such additional capacity may be measured as:

   (a) percentage of CPU power not used;

   (b) potential increased throughput, i.e.; additional interactive transactions which may be processed within the specified response time.

3. Comparative ranking of candidate systems in a competitive acquisition.

4. Identification of potential bottlenecks in a candidate system.

5. Determination of the appropriate size of a candidate system.

6. Incorporation in acceptance test procedures.

7. Monitoring the performance of an installed system.

While all of these reasons may be useful and valid, this study is primarily focused on the determination of required functional and performance capability and available additional capacity on the vendors' proposed system as part of the acquisition process. The other uses listed have been considered only in terms of additional benefit to be gained from using a given technique.

With the rapid advances in the cost/performance of microcomputer-related technology, the issue of end-user productivity becomes increasingly important. This issue will only be indirectly addressed in this report. However, it is important to realize that, as new ways of using computers become established, it will become necessary to address end-user productivity more directly in computer performance evaluation. This issue is addressed by the National Bureau of Standards in a series of reports including a recently published document [GI83] on agency experiences with microcomputers.

1.2 Background

The objective of any procurement is the identification and acquisition of the most appropriate and cost-effective computer systems available to meet the specified requirements. Within the context of an emphasis on fostering competition, a number of approaches have been used to evaluate candidate computer systems. One of these approaches is benchmarking.

Benchmarking (the measurement of the performance of a candidate system under actual or simulated workload) is the most widely accepted method of evaluating computer systems for Federal agency procurements. It is generally considered to provide a fair and unbiased live test demonstration of candidate computer systems.

However, the growth in numbers of smaller and less expensive systems and the increasing use of distributed systems has raised questions about whether or not benchmarking is cost-effective. The length of the acquisition cycle in the Federal government has also made benchmarking less useful, due to the lower long-range accuracy of workload forecasting and representation.

It is the recognition that benchmark costs are increasing, in addition to their questionable accuracy, that has promted this study. The concepts presented in this study are applicable to all sizes of general purpose computers, from microcomputers to mainframes. Embedded or single-purpose computers, such as those used in weapon systems, have been excluded.

The information presented in this guide is based on an extensive review of the relevant literature, both technical and regulatory (Appendix A), and on a series of interviews with representatives of Federal agencies and vendor organizations (Appendix B) with experience in using benchmarking and other evaluation techniques.

1.3 ADP Acquisition Process

A detailed description of the ADP system acquisition process is not within the scope of this report. However, it is important to identify how the selection of an evaluation technique(s) fits into this process. The selection of evaluation technique(s) is performed as an integral part of the Evaluation Plan and Strategy phase of the acquisition process. In general, the acquisition process involves six main components:

1. Studies and Approvals. Feasibility studies, approvals, resource sharing and consolidation studies, funding studies, etc. are generally performed as the first step, often in response to internal and/or external regulations.

2. Definition of User Requirements and Technical Specifications. User requirements provide the basis for the Request for Proposal (RFP), and for the evaluation and selection procedures. Development of technical specifications (based on user requirements), which will be released to all interested vendors, is a crucial part of the process.

3. Evaluation Plan and Strategy. An evaluation plan describes the cost and technical factors that are to be evaluated and the strategy for conducting the evaluation. As part of this phase, the objectives of the evaluation should be clearly defined, that is, the agency requirements or technical specifications the agency is intended to evaluate. Once the evaluation objectives are identified, the technique(s) for testing them can be selected.

4. Preparation and Release of the RFP. The RFP combines the user requirements and technical specifications with the evaluation criteria, evaluation package, and

contractual requirements. The RFP is released, usually followed by vendor questions and subsequent amendments to the RFP.

5. Evaluation of Proposals. Proposal evaluation is the process by which the procuring agency determines the extent to which the hardware and software configurations proposed by the vendors meet the requirements stated in the RFP. Various techniques are necessary to validate those requirements that cannot be sufficiently evaluated from the vendor's written proposal.

6. Selection and Contract Award. After an evaluation of each vendor's written proposal and, where appropriate, performance testing (e.g., benchmarking), negotiations are held with qualifying vendors. Subsequently, best and final offers are usually solicited. A contract is then awarded to the vendor who meets the requirements in the RFP, and who offers a system that is most advantageous to the procuring agency in terms of technical capabilities and expected life cycle cost.

More information on these acquisition components can be obtained from the General Services Administration, Office of Information Resources Management, Washington, D.C. 20405.

1.4 Planning for Uncertainty

This study is focused on the selection of evaluation techniques. However, a short discussion of contractual flexibility is included, since it is advisable to plan for the nearly inevitable gap between the forecasted and actual workloads.

Since uncertainties must be expected in any computing environment, the use of evaluation techniques discussed in the following sub-sections should be combined with contractual safeguards. Inaccuracies in the workload forecasting - and, for some evaluation techniques, the workload representation - on which the evaluation is based must be adjusted and accounted for during the system life. Additionally, shifts in the economy or in other external factors (including the impact of technological change) may alter the size or the composition of the workload. In the Federal sector, furthermore, changes in the law may have similar effects.

Since the length of the Federal ADP procurement cycle renders frequent procurements of large scale systems impractical, the uncertainty in future workloads may be compensated for by:

1. An analysis of the proposed systems to determine the sensitivity of their costs and performance to workload fluctuations.

2. A set of contractual arrangements providing for system growth as needed.

The arrangements suggested above should include safeguards for both the procuring agency and the vendor(s) to insure an appropriate rate of system growth. RFP and contract clauses should cover the means of determining the points at which system growth is desirable and the nature of the appropriate price adjustments. The General Services Administration (GSA) provides suggested RFP and contract clauses for these purposes in their "Guidance to Federal Agencies on the Preparation of Specification, Selection, and Acquisition of Automatic Data Processing Equipment Systems."

2. CURRENT CONSTRAINTS IN EVALUATING COMPUTER SYSTEMS

The use of evaluation techniques in the Federal government during acquisition of computer systems is constrained by Federal procurement regulations and GSA guidelines. Constraints may be defined as those factors which limit a procuring agency's choice of evaluation techniques. They include:

1. Federal procurement regulations and guidelines show a preference toward benchmarking for large systems.

   (a) Federal Procurement Regulations (FPR 1-4.1109-21) state that simulation will not be used as the only means of describing data processing requirements. Also, offers should not be considered non-responsive or unacceptable solely on the basis of simulation results. The same restrictions apply to modeling. This regulation essentially prevents the use of simulation and modeling as a substitute for benchmarking by placing restriction on their use.

   (b) GSA's "Guidance to Federal Agencies on the Preparation of Specification, Selection, and Acquisition of Automatic Data Processing Equipment Systems", Section D states that, depending on the size and complexity of the processing requirements, the agency will specify either a benchmark or an operational capability demonstration, or both.

2. There is a significant Congressional desire to foster competition among vendors.

3. Most vendors and Federal agencies show a preference toward benchmarking, especially in fully competitive procurements.

In the private sector [GE81], much less use is made of benchmarking and more reliance is placed on rating charts and on the experience of others with similar systems. These tendencies are facilitated by the following factors:

1. A full and open competition is not regularly used to acquire computer systems.

2. A shorter procurement cycle makes errors correctable in less time, due to simpler procedures for acquiring computer systems.

Since these factors do not apply to the Federal sector, it is unlikely that the techniques used in the private sector can be directly adopted by Federal agencies.

3. FACTORS AFFECTING THE CHOICE OF EVALUATION TECHNIQUES

The choice of a technique or a set of techniques for evaluating a candidate computer system should be based on the nature of the planned system, the workloads, and the type of procurement. Also, the choice should be based on the objectives to be met by the use of a given evaluation technique.

3.1 Agency-Dependent Factors

The following is a list of those agency-dependent factors which may affect a procuring agency's choice of evaluation technique:

1. The size, complexity, and cost of the system;

2. The importance of the system in allowing the agency to fulfill its mission;

3. The system architecture/concept (centralized vs. distributed, batch vs. interactive);

4. The type of applications to be handled (e.g., compute-heavy, real-time, high degree of I/O, balanced mix);

5. The degree of change from the current system (e.g., CPU change only, computerization of currently manual applications);

6. The type of procurement (e.g., sole source, compatible only, fully competitive, multi-vendor buy);

7. The degree of anticipated uncertainty;

8. The nature and level of the evaluation skills which are possessed by the procuring agency staff or which are readily available to the agency from other sources.

3.2 General Factors

This section identifies general criteria (non-agency dependent) for selecting one or more evaluation techniques to be used in a given procurement.

3.2.1 Conformance with Federal Procurement Regulations

Conformance with federal procurement regulations is the degree to which the use of a given technique for a specific procurement adheres to the regulations and/or guidance promulgated by OMB, GSA, and GAO.

3.2.2 Accuracy

Accuracy is the degree to which the results of an evaluation technique approximate the behavior of the system under actual conditions. In the extreme, the most accurate evaluation technique would consist of running the full workload on the candidate system for the entire system life. However, the aim of an evaluation should not be the greatest degree of accuracy but, rather the greatest degree which is cost-effective.

Accuracy depends on the nature of the technique (e.g., benchmarking may be inherently more accurate than simulation because the real computer system is used) and the quality and effectiveness with which the technique is implemented. Accuracy contributes to perceived fairness and affects the total system cost (via the savings associated with an accurately selected system or, conversely, the additional cost of an inaccurately selected one).

The accuracy of an evaluation technique may be estimated on the basis of empirical tests of the technique and of past experience with that technique for similar systems.

3.2.3 Cost

The cost of using an evaluation technique is the total amount of money spent, by both the vendor and the procuring agency, to apply it to a candidate system. It is clearly desirable to minimize the total system cost (over the expected system life) rather than just the evaluation cost. The evaluation technique selected on grounds of evaluation cost may not be the least expensive, overall. An inaccurately selected system can be more costly than a suitable one.

The cost of using an evaluation technique is affected by:

1. The ease of using the technique; i.e., the amount of effort (preparation, training and application) required to apply it to a candidate system.

2. The time needed to use the technique, i.e., the amount added to the procurement time in order to apply the technique.

3. The flexibility of the technique; i.e., its ability to be used on different types of systems, on different sizes of systems (expandibility) and/or at different stages (such as selection, sizing, acceptance and operation) of a system's life cycle. All else being equal, a more flexible technique will result in lower cost over the long term, due to the distribution of training and other costs over several applications, and should thus be preferred.

The cost of applying a given evaluation technique may be a deciding factor in the acceptability of the technique to the vendor(s) and the procuring agency. The cost to both the vendor and the procuring agency of using a specified technique in a given instance may usually be estimated with reasonable accuracy. However, the eventual savings resulting from this expenditure are often harder to determine.

3.2.4 Perceived Fairness/Acceptability to Vendors

Perceived fairness is the degree to which an evaluation technique is considered not to favor any one vendor. The perceived fairness is a subjective factor; the most accurate evaluation technique may not necessarily be perceived to be the fairest one possible.

An evaluation technique is acceptable to a vendor if that vendor will not protest its use and is willing to participate in procurements in which the technique is used. A technique acceptable to vendors should be: (1) perceived to be fair and, (2) economical enough to the vendor(s) to be affordable over a series of procurements in which some are lost. Acceptability to vendors contributes to acceptability to the procuring agency by minimizing protests.

3.2.5 Ease of Understanding

Ease of understanding is the clarity with which an evaluation technique is comprehended by someone not trained in that technique. (For example, such techniques as equating the quality of a system with its speed and judging speed by instruction cycle time are usually very easy to understand.)

The ease of understanding an evaluation technique depends on the nature of the technique and on the degree to which the system being procured differs from the one being upgraded/replaced. It contributes to perceived fairness and to the flexibility and expandability of a technique. Since it is a subjective factor, it may be judged by those who are responsible for using the results of an evaluation.

4. ASSESSMENT OF CANDIDATE EVALUATION TECHNIQUES

This section presents an appraisal of several evaluation techniques with regard to the parameters defined in Section 3.2. The techniques to be examined are:

1. Proposal data analysis;
2. Applying experience of the evaluator(s);
3. Instruction timing analysis;
4. Rating charts analysis;
5. Analytic modeling and simulation;
6. Benchmarking; and
7. Prototyping.

While the degree to which specific evaluation techniques conform to Federal Procurement Regulations and to GSA guidance is usually clear, the relative values of the other parameters, particularly accuracy and cost, are less well known.

4.1 Proposal Data Analysis

Proposal data may be defined as the pricing information, configuration descriptions, and performance guarantees (i.e., the guarantees that the proposed systems will perform the specified functions at the the specified levels of speed and accuracy) contained in the vendors proposal(s).

The decision to use only the information contained in the proposal(s) submitted may, in some circumstances, be very appropriate. This approach provides the lowest (no additional) cost for evaluating vendors' proposed systems and may tend to decrease the length of the procurement. It is particularly suitable for low-cost systems, where the cost of using additional evaluation techniques may exceed the benefit to be gained from it. In such a case, it is particularly important to incorporate considerable flexibility into the contract, as discussed in Section 1.4.

4.2 Applying Experience of the Evaluator(s)

The experience of the evaluator(s) consists of the knowledge of the candidate system(s) that they have when the evaluation is begun and their opinions of these system(s) based on this knowledge.

The success of using this technique depends exclusively on the ability of the evaluator(s). Therefore, its value in predicting performance and capacity is likely to be most questionable.

This technique is easy to understand, quick and easy to use, and comparatively low in cost. It does not generally conform with current Federal Procurement Regulations or GSA guidance. It is applicable to many sizes and types of systems at many stages in their life cycles. It is likely to be less usable for newer systems, for which less experience is available.

4.3 Instruction Timing Analysis

Instruction timing techniques are designed to provide a measure of CPU speed, based on the assumption that such a measure bears some relationship to system capacity. Instances of the technique include CPU cycle time comparison, instruction execution timing, and instruction mixes. The first of these methods is simple, and straightforward and will not be discussed further. The second and third are more complex and will be defined below.

Instruction execution timing (also called the cycle-add technique) is usually the comparison of arithmetic instruction (normally add or multiply) execution times. Instruction mixes involve the computation of a weighted average of the execution times for a mix of instructions which are typical of the intended applications. The weights are derived from the measured or assumed frequencies of instructions in the actual or planned applications. For example, a scientific instruction mix would emphasize arithmetic operations, while a business mix would be weighted toward instructions used in moving and editing data.

Unless the planned system will focus on heavily compute-bound applications, instruction execution timing is not likely to provide a good measure of whether a candidate system can meet the specified functional and performance requirements. This technique is not likely to indicate the amount of additional capacity available on a candidate system even if the system is simply a more powerful version of the one currently used; i.e., only the CPU is being upgraded.

Except in the circumstances noted, instruction execution timing has not proven to be an accurate measure of performance. It is easy to understand, quick and inexpensive, and relatively easy to use. It generally does not conform to Federal Procurement Regulations or GSA guidance, although its use may be acceptable in low dollar value procurements. While it may be used in the source selection phase of a system's life, even before the system itself is available, it offers no new information which might prompt its use during a system's operational life. It may be used on any type or size of system, but, as noted, above, such use may not be accurate.

Instruction execution timing will probably not be perceived as fair, except in the limited circumstances discussed above, and thus will probably be generally unacceptable to vendors. It does have the advantage of not requiring workload representation. Instruction execution timing becomes steadily less applicable as the use of networking and distributed processing increases. In these processing modes, the importance of the CPU in total system efficiency is decreasing [BO79].

4.4 Rating Charts Analysis

Rating charts are tables listing such computer system characteristics as CPU cycle time, speed of arithmetic operations, memory access time, word size, and I/O rates. They may also include measures of power based on a standard set of benchmark problems and/or instruction mixes. Examples are Computerworld ratings [CO--], Auerbach ratings [AU--], and Adams's Charts [AD--].

Like all of the evaluation techniques, rating charts require proper use. For a system which is heavily biased toward one performance factor (such as numerical computation speed or tape input/output), rating charts may provide some assistance in predicting both performance and available additional capacity. In larger, more complex or less centralized systems, rating charts are likely to be less useful.

Rating charts are relatively easy to understand and to use. For the most part, their use does not conform with Federal Procurement Regulations or GSA guidance. They are most useful before a system has been obtained apply to a range of system types and sizes. Their use is not likely to lengthen the procurement cycle or add much to its cost. Rating charts are sometimes perceived to be fair, depending on the nature of the system, and will, therefore, vary in acceptability to vendors.

4.5 Analytic Modeling and Simulation

Analytic modeling is a mathematical description of computer system behavior. Models may be implemented with paper and pencil or by a computer program. The method(s) may be statistical, probabilistic (usually based on queuing theory), graphical, or algorithmic (algebraic). Because of the mathematical nature of analytic modeling, it would be unrealistic to think in terms of developing an analytic model from scratch. Most analytic modeling is done with the aid of preprogrammed analytic modeling packages. Such packages require that the characteristics of the system be described in terms of some input language. Four commercially available analytic modeling packages in general use are [KE83]: BEST/1, SNAP, THEsolver, and CADS. Another package, ACMS [AC82] was developed by the Federal government.

Simulation involves the representation of the processing flow of a computer system. This representation may be accomplished by using simulation packages or by using a simulation language to develop a model of the specific system to be evaluated. Such development may be accomplished in a special-purpose system simulation language (e.g., ECSS), a general-purpose simulation language (e.g, GPSS, SIMSCRIPT II.5) or a general-purpose programming language (e.g., FORTRAN, PL/I). ECSS is one of the most widely used simulation languages for modeling computer systems. ECSS was developed by the Rand Corporation and enhanced by FEDSIM for use within the Federal government. Further information on the use of ECSS can be obtained from: FEDSIM, Department of the Air Force, Washington, DC 20330.

These techniques have been combined here because their advantages and drawbacks are virtually identical. Analytic modeling or simulation can be used to determine whether a candidate system can meet the specified functional and performance requirements for the expected workload, as well as the amount of additional capacity of the system. They can be highly accurate within vendor lines, but may be much less so across them.

The construction and use of these techniques may be somewhat difficult to understand for those not trained in the technique(s). For this reason, and because of the difficulty of validating a model across different computer architectures, an analytic model or a simulation may not be perceived as fair when used in a fully competitive procurement. The use of analytic modeling or simulation does not conform to procurement regulations or GSA guidance when used in a fully competitive procurement, although Federal Procurement Regulations (see Section 2) indicate that such use is permissible in a small or medium size system procurement (regardless of the degree of competition).

Analytic modeling and simulation are often relatively costly, due to their complexity. Because they may be used before an actual physical system is available, they are particularly useful early in a system's life cycle. In addition, they may be applied later in a system's life for such purposes as predicting the impact of changing a system before implementing the change. They may be used on many different sizes and types of systems, although the scope of any specific model or simulation may be more limited. Because they lack accuracy and perceived fairness across vendor lines, analytic modeling and simulation may not be acceptable to vendors in a fully competitive procurement [BO79].

4.6 Benchmarking

Benchmarking is a common test by which different vendor systems can be evaluated. It facilitates the verification of the proposed system as to the time required to perform the workload within certain predetermined service level requirements. Benchmarking may also be used during a functional demonstration to verify that a system has certain functional capabilities. Appendix C of this document identifies available guidelines for benchmarking.

4.6.1 Timed Benchmark Tests

Benchmarking involves measuring performance of an actual candidate computer system under a benchmark which is designed to stress the system in the same way as the forecasted workload. The workload may be represented by a set of real and/or synthetic benchmark problems (batch programs, online activities). While most benchmark problems are designed to represent a certain workload category at a given organization, some attempts have been made to develop standard benchmark problems that may be used repeatedly. Such benchmark problems are usually designed to represent a given category of workloads either in terms of functional or resource usage characteristics.

Since benchmarking involves the use of actual candidate hardware and system software, it is inherently more accurate than simulation or analytic modeling. However, it requires more precise and detailed workload forecasting than these other techniques. This technique can be a good means of determining whether a candidate system can perform the forecasted workload at the required service level. On the same basis, benchmarking can also be used to determine the amount of additional capacity available on a given system.

Actual benchmarks are relatively easy to understand; synthetics are slightly less so. Benchmarking is easiest to apply to systems which are centralized and batch-oriented. Since most systems today are terminal driven, remote terminal emulator (RTE) was developed to benchmark online workloads. The RTE is an independent computer system used to emulate the terminal workload on a candidate computer system. The "Use and Specifications of Remote Terminal Emulation in ADP Acquisitions" [GS79] provides information on when and how to use RTE during the acquisition of systems requiring an online component(s).

This technique conforms to Federal Procurement Regulations, particularly for large systems. It may be applied to a system only after the system physically exists. Benchmarking typically adds significantly to the length and cost of the procurement cycle.

Benchmarking is usually perceived to be fair, although benchmarks may well be biased (deliberately or unconsciously) toward a specific vendor. It is a relatively costly technique for both the vendor and the procuring agency.

The growth in numbers of smaller and less expensive systems and the increasing use of distributed systems have made benchmarking less cost-effective than it was for centralized mainframe-based computer systems. The length of the acquisition cycle in the Federal government has also made benchmarking, like the other system performance evaluation techniques (simulation and modeling), less useful, due to the lower long-range accuracy of workload projection and representation.

4.6.2 Functional Demonstrations

Functional demonstrations are usually designed to test certain mandatory requirements or desirable features that cannot be satisfactorily evaluated from vendor proposals or would not be appropriate for inclusion in a timed benchmark test. This evaluation technique can also be used in combination with the techniques discussed above. The growth in numbers of smaller and less expensive systems make this evaluation technique more acceptable both for vendors and procuring agencies. Also, the increasing use of special-purpose application packages and systems makes functional demonstration a viable evaluation alternative.

This technique conforms to regulations and GSA guidance, depending on the size and complexity of the system being procured.

4.7 Prototyping

Prototyping is an alternative evaluation technique, in which the procuring agency funds selected vendors to develop a prototype system. This evaluation technique should be used only when the risk to the government is extremely high. Factors to be considered using this method are discussed in OMB Circular A-109. Prototyping is much more costly and time consuming than other evaluation techniques. However, it reduces the risk of acquiring inappropriately sized systems, since a prototype of an actual system is completely developed by each vendor.

5. USE OF EVALUATION TECHNIQUES

Table 1, is a summary of the qualitative assessment of those evaluation techniques which are described in Section 4, as to their relative accuracy, cost, and suitability. Prototyping is not included in this table, because it is applicable only in special cases and its use is governed by OMB Circular A-109. The use of these alternatives might require years to gain acceptance both by Federal agencies and the vendor community. However, completed Federal procurements indicate [GE82] that benchmarking is not always necessary for limited competition (e.g.; compatible system only) of procurements that have under $2 million estimated life cycle cost.

No cost data is available on the use of the different evaluation techniques in the same procurement. However, it is well known that the cost of using benchmarking in evaluating computer systems increases. Therefore, agencies might consider the use of evaluation techniques other than benchmarking for evaluating computer systems in their procurement process.

The desired results of applying any evaluation technique are significantly impacted by the availabilty of up-to-date information on the agency's workload requirements. If an agency is to succeed in the acquisition process, the agency should have an on-going procedure for determining their requirements for computing resources. The determination and forecasting of these requirements should

ATTRIBUTE	PROPOSAL DATA ANALYSIS	EXPERIENCE OF EVALUATOR(s)	INSTRUCTION TIMING	RATING CHARTS ANALYSIS	ANALYTIC MODELING	SIMULATION	BENCHMARKING
ACCURACY	VARIABLE	VARIABLE	LOW	VARIABLE	MODERATE	MODERATE TO HIGH	HIGH
COST	LOWEST	COMPARATIVELY LOW	LOW	LOW	MODERATE	HIGH	HIGH
SUITABILITY	LOW COST SYSTEMS	OLDER SYSTEMS WITH RELATIVELY STANDARD APPLICATIONS	COMPUTE-ORIENTED SYSTEM WHERE ONLY COMPATIBLE CPU IS BEING PROCURED	SIMPLER, SMALLER, MORE CENTRALIZED SYSTEM	COMPATIBLE ONLY PROCUREMENT	1) MULTI-VENDOR SYSTEM 2) DISTRIBUTED SYSTEMS	1) FULLY COMPETITIVE PROCUREMENT 2) CENTRALIZED SYSTEMS

TABLE 1. QUALITATIVE ASSESSMENT OF RELATIVE ACCURACY, COST AND SUITABILITY OF EVALUATION TECHNIQUES

be an integral part of agencies' planning process. Having up-to-date information on the agency's workload requirements would shorten the acquisition cycle, and would reduce the cost of the evaluation process.

5.1 Use of Benchmarking

It is widely accepted [NA80] in the performance evaluation community that benchmarking can provide an unbiased and fair demonstration of the vendors' proposed systems. However, this does not imply that an agency is necessarily getting the most cost-effective system to perform the workload. Presently no widely accepted system-independent unit is available to measure [KE83] the workload at the level required to represent the workload in the benchmark. The lack of this unit of measure can lead to the acquisition of over- or under-sized systems because the workload is measured and represented in the present system's capabilities and not in the candidate system's.

A procuring agency can acquire appropriately sized systems by forecasting its workload with relatively high degree of accuracy and representing its workload in the benchmark in terms of:

1. Job origin (e.g., on-line, remote batch, batch),

2. ADP operations performed (e.g., edit, update),

3. Time distribution of ADP operations performed,

4. Operational requirements (e.g., priority, security).

However, creating a high quality benchmark is an expensive undertaking. In procurements under $2 million estimated life cycle cost, the benefits to be gained from the use of benchmarking should be carefully evaluated. For large dollar volume procurements, the agency should be aware of the importance of benchmark representativeness in terms identified above.

5.2 Use of Alternative Evaluation Techniques

Athough no quantitative information is available on the cost-effectiveness of the evaluation techniques currently used in the same procurement, it is widely accepted that benchmarking can be expensive and the results can be quite inaccurate. There are certain drawbacks, such as of system-dependent units of measure to express the workload categories, that are often difficult to overcome. This problem coupled with other deficiencies in the benchmark construction process, may make the estimated level of obtainable accuracy unacceptable.

The use of simulation and modeling as the sole evaluation technique is prohibited by Federal Procurements Regulations. However, simulation and modeling can be used along with proposal data analysis, experience of evaluators, and rating charts analysis for limited competitions. Simulation and modeling should also be considered to complement benchmarking in evaluating complex systems with networking requirements, or for validating the representativeness of the benchmarks. Functional demonstrations should also be considered in combination with other evaluation techniques where the vendor demonstrates certain prescribed capabilities without regard to total system performance.

Although, it has not been discussed as a separate evaluation technique, the experience of other organizations with similar systems can be used as an input for validating equipment capacity in combination with other alternatives described in this document.

6. SUMMARY

In light of the prevailing Federal Procurement Regulations, GSA guidance, and the advantages and disadvantages of the evaluation techniques discussed, there is no one best technique for evaluating computer systems in the acquisition process. Benchmarking is very expensive both for vendors and agencies during the procurement process. However, there are few alternatives for evaluating medium and large scale computer systems in the Federal government's competitive procurement environment.

The techniques discussed vary in complexity, accuracy, cost, and suitability. Their applicability can only be determined on a case-by-case basis. The agency-dependent (including application-dependent) factors and the general factors discussed in this document should provide agencies with guidance for determining the most appropriate evaluation technique for a specific procurement.

In general, the selection and use of a given evaluation technique should be governed by its cost-effectiveness to the organization as a whole, including the cost to the vendors, which is usually reflected back in higher cost to the government over the long term. The resources to be expended in using an evaluation technique should be commensurate with the expected life-cycle cost of the planned system. In some cases, the criticality of the system in enabling the agency to fulfill its mission might be a deciding factor over cost considerations.

APPENDIX A

SELECTED REFERENCES AND READINGS

[AC82] "ACMS Users Guide." FEDSIM, 1982.

[AR66] Arbuckle, R. A., "Computer Analysis and Thruput Evaluation," Computers and Automation, Vol. 15, No. 1, January 1966, pp. 12-15, 19.

[AD--] Charles Adams Associates, Computer Characteristics (Formerly: Computer Charcteristics Review, Currently: Computer Review), GML Corp., Lexington, Massachusetts, 1961 - Present.

[AU--] Auerbach Standard EDP Reports, Auerbach Info., Inc., Philadelphia

[BA78] Bayraktar, A. Nevzat, Computer Selection and Evaluation, Naval Postgraduate School, Monterey, CA, June 1978.

[BO79] Borovits, I. and Newman, S., Computer Systems Performance Evaluation, Lexington Books, Lexington, MA, 1979.

[BO75] Boyse, John W. and Warn, David R., "A Straightforward Model for Computer Performance Prediction," Computing Surveys, Vol. 7, No. 2, June 1975, pp. 91-93.

[CL80] Clark, Jon D. and Golladay, Robert M., "Empirical Investigation of the Effectiveness of Several Computer Performance Evaluation Tools," Performance Evaluation Review, Vol. 9, No. 3, Fall 1980, pp. 31-36.

[CL79] Clark, Jon D. and Reynolds, Thomas J., "Computer Performance Evaluation - An Empirical Approach," Performance Evaluation Review, Vol. 8, No. 12, Spring - Summer 1979, pp. 97-101.

[CO--] Computerworld, Framingham, Massachusetts.

[FE77] Feldmeyer, Bruce A., Computer Performance Evaluation During System Acquisition, MTR-3290, The MITRE Corporation, Bedord, MA, January 1977.

[GA80] Gay, A.R., "Benchmarking a Multi-access System," Software Practice and Experience, Vol. 10, 1980, pp. 45-55.

[GE81] General Accounting Office, Non-Federal Computer Acquisition Practices Provide Useful Information for Streamlining Federal Methods, AFMD-81-104, October 2, 1981.

[GE82] General Accounting Office, Benchmarking: Costly and Difficult, but Often Necessary when Buying Computer Equipment and Services, AFMD-83-5, October 22, 1982.

[GI83] Gilbert, Dennis; Parker, Elizabeth; Rosenthal, Lynne; "Microcomputers: A Review of Federal Agency Experiences," National Bureau of Standards. 1983 June. 146p. NBS Special Publication 500-102.

[GR78] Graham, G. Scott, "Queueing Network Models of Computer System Performance," Computing Surveys, Vol. 10, No. 3, September 1978, pp. 219-224.

[GS79] General Services Administration (GSA) Automated Data and Telecommunications Service. "Use and Specifications of Remote Terminal Emulation in ADP System Acquisitions." Washington, DC:GSA/ADTS(CDD); 1979 August; FPR 1-4.11.

[HO80] Hodgins, Bart Dallas, A Computer Evaluation Technique for Early Selection of Hardware, Naval Postgraduate School, Monterey, CA, December 1980.

[HO77] Howard, Phillip C., "Performance Considerations for Distributed Data Processing Systems," Proceedings of the 1977 SIGMETRICS/CMG VIII Conference on Computer Performance: Modeling, Measurement and Management, 1977, pp. 237-246.

[JO77] Joslin, Edward O., Computer Selection: Augmented Edition, The Technology Press, 1977.

[KE83] Kelly, John, C., "Capacity Planning A State of the Art Survey," Datametrics Systems Corporation. NBS Contract Report, May 1983.

[KN66] Knight, Kenneth E., "Changes in Computer Performance," Datamation, Vol. 12, No. 9, September 1966, pp. 40-54.

[KO79] Kochhar, A. and Burns, N., "Micros - How to Make the Right Choice," Machinery and Production Engineering, Vol. 135, No. 3485, 31 October 1979, pp. 41-44.

[LE77] Lehman, Richard S., Computer Simulation and Modeling: An Introduction, Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[LE83] Letmanyi, Helen, "Guide on Workload Forecasting." Work in Progress, NBS, 1983.

[LI80] Lias, Edward J., "Tracking the Elusive KOPS," Datamation, November 1980, pp. 99-105.

[LU71] Lucas, Henry C., Jr., "Performance Evaluation and Monitoring," Computing Surveys, Vol. 3, No. 3, September 1971, pp. 79-91.

[MA79] Mamrak, Sandra A. and Abrams, Marshall D., "A Taxonomy for Valid Test Workload Generation," Computer, December 1979, pp. 60-65.

[NA77] National Bureau of Standards, "Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment," FIPS Pub. 42-1, May 15, 1977.

[NA80] National Bureau of Standards, "Guideline on Constracting Benchmarks for ADP System Acquisitions," FIPS Pub. 75, September 18, 1980.

[RE76] Reiser, Martin, "Interactive Modeling of Computer Systems," IBM Systems Journal, Vol. 15, No. 4, 1976, pp. 309-327.

[RI80] Ringland, Gill and Standing, Phil, "Measure for Power... Money for Measure," Computing, Vol. 8, No. 33, 14 August 1980, pp. 14-15.

[SA81] Sauer, Charles H. and Chandy, K. Mani, Computer Systems Performance Modeling, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.

[SC67] Schneidewind, Norman F., "The Practice of Computer Selection," Datamation, Vol. 13, No. 2, February 1967, pp. 22-25.

[SP75] Spiegel, Murray, R., "Probabilty and Statistics", McGraw-Hill Book Company, 1975.

[ST77] Stevens, David F., "Obfuscatory Measurement," Proceedings of the 1977 SIGMETRICS/CMG VIII Conference on Computer Performance: Modeling, Measurement, and Management, 1977, pp. 33-39.

[VA80] Vallone, Antonio, "Structured Procedure for Comparison and Selection of Computer System Designs," AFIPS Conference, 1980, pp 801-806.

[WE70] Weihrich, W. Fred, "Computer Selection," Data Management, Vol. 8, No. 2, February 1970, pp. 31-33.

[WE76] Webster, G.J. and Johnson, C.W., "The Evaluation and Selection of a Computer System for Interactive Design," Computer Aided Design, Vol. 8, No. 4, October 1976, pp. 247-251.

[WO76] Wolin, Louis, "Procedure Evaluates Computers for Scientific Applications," Computer Design, Vol. 15, No. 11, November 1976, pp. 93-100.

[--81] "Unbalanced Proposals," Procurement Systems Digest, Vol. 4, No. 7, October 1981, pp. 1-2.

APPENDIX B

ORGANIZATIONAL INFORMATION SOURCES

The study documented here used information drawn from interviewing personnel of the organizations listed below. In some cases, MITRE staff members conducted the interviews. In other cases, draft reports of interviews conducted by GAO staff for a separate, independent GAO study provided the necessary information:

Federal Organizations

Department of the Army, Computer System Selection and Acquisition Agency
Department of Commerce, Geophysical Fluid Dynamics Laboratory
Department of Energy
Department of Housing and Urban Development
Department of the Treasury:
- Bureau of Government Financial Operations
- Service Center

FEDSIM
Internal Revenue Service
Marine Corps
National Aeronautics and Space Administration
- Goddard Scientific Applications Computing Center
- Goddard Management Services Office

National Institutes of Health
Postal Service
Securities and Exchange Commission

Private Sector Organizations

Amdahl Corporation
BGS Systems
Burroughs Corporation
CBEMA
Control Data Corporation
Cray Research
Digital Equipment Corporation
Harris Corporation
Honeywell Corporation
International Business Machines Corporation
Martin Marietta
Neshaming Valley Information Processing, Inc.
Texas Instruments
Vion

APPENDIX C

GUIDANCE ON BENCHMARKING

The results of the qualitative evaluation of benchmarking and its alternatives indicates that benchmarking is a viable tool for evaluating vendors' proposed systems, especially for procurements over $2 million estimated life cycle cost. Agencies planning to use benchmarking should find the following documents useful: [NA77], [NA80], and [GS79]. The "Guideline on Constructing Benchmarks for ADP System Acquisitions" FIPS PUB 75 [NA80] describes how to construct "representative" benchmarks to the maximum extent possible. The remainder of this section is an extract from FIPS PUB 75 for emphasizing the importance [GE82] of the proper documentation of the benchmark mix(es), the Live Test Demonstration (LTD) rules, and the testing of the benchmark by running each benchmark mix on one or more systems other than the one on which it was developed.

1. Prepare the Benchmark Package

1.1 Document Each Benchmark Mix

A functional description of each benchmark problem, as well as internal documentation within each problem, should be provided in the benchmark package portion of the RFP. English-language scenarios for batch and on-line benchmark problems should be provided and, where possible, supplemented with sample scripts. Sample results of the benchmark, as well as the expected service time requirements for the benchmark problems, should be included as part of the benchmark package. A glossary of terms should also be provided to reduce any misunderstandings. A general block-diagram showing the input files and their origin should be provided. For example, "file A generated by program ABC," "provided by the Government on tape 2," "vendor provided," "generated by data generator program XYZ" may be necessary qualifiers in such a description. The destination of the output files should be depicted on such a diagram. A description of each file should include information such as record length, blocking factor, number of records in the file, access method, storage media on which the file will reside when the benchmark is executed, field definitions, data formats, etc. The data provided to the vendors should be in a machine-independent format, and the volume of data provided on magnetic tape should be kept to a minimum. All data provided should be in compliance with Federal standards for media and interchange codes. Constraints on modifications to the source code of benchmark problems must also be documented. Manual modifications beyond those necessary to interface with the vendor's system are normally not allowed. Source or object code optimization should be allowed only if the optimization mechanism will be part of the standard software delivered with the computer system (for example, the vendor's off-the-shelf optimizing compilers). The RFP should require that each vendor meet with the agency benchmark team a few weeks before the LTD so that questions (on both sides) concerning the nature of the benchmark and the LTD can be resolved. Prior to such a meeting, the vendor should furnish the following information to the benchmark team:

1. a diagram of the complete configuration that is being proposed for each augmentation point, and the configuration(s) upon which the benchmark will be run (if different than proposed);

2. complete source program and data file listings, with a complete description of any modifications to benchmark programs or scenarios (including the exact changes made and reasons for the changes);

3. compilation listings for all programs showing job control information, compilation maps, size of the object modules, main (or virtual) memory allocations, disk or drum allocations, peripheral device requirements; also, complete listings of program outputs, and any other listings which would be a direct result of compilation and execution of the benchmark (e.g., diagnostics, cross-reference lists, etc.);

4. complete hardcopy of all operator/computer communications generated during compilation, loading, and execution of each benchmark problem;

5. listing of all software packages used to process the benchmark problems, including a list of all system generation routines and other system utilities that may be required (the software should be identified by release and version);

6. a complete set of manuals describing the system generation for each proposed configuration.

1.2 Document the LTD Rules

The rules for setting up and performing the LTD must be carefully documented in the RFP in order to avoid any misunderstandings between the vendors and the procuring agency. Furthermore, if not stated elsewhere in the RFP, the rules covering the following should also be stated:

1. allowable variations in the benchmark results;

2. acceptance and evaluation criteria of the benchmark results;

3. how the benchmark will be operated and supervised;

4. the environment during the benchmark (as discussed in more detail below).

a. Timed Benchmark Tests

When practical and only when it is believed necessary, the agency may require that the full complement of components be configured during the timed benchmark test, even if only partially used by the benchmark, in order to include the effects of device tables resident in memory, operating system overhead, file placement, channel contention, etc. (It should be noted that because such a requirement usually places an undue expense on the vendors and could limit the number of responding vendors, it should be stated only when absolutely necessary.) For example, the agency might require the vendor to configure a full complement of disks on which a set of "dummy" files might be loaded. The allocation of these files to specific disks should be done in the same manner as would occur for the real workload; namely, the vendor should have the system assign the files automatically, or the vendor should assign them manually using whatever utilities and suggested practices are contained in the vendor's user manuals. Care should be taken to prevent the vendor from physically arranging the data on or across disks in order to optimize only the benchmark. When it is not feasible to benchmark the complete proposed configuration, the agency may require the offeror to perform a functional demonstration for those devices or components that were not part of the timed benchmark test (see below).

The LTD itself must be well-documented. The execution priorities of the benchmark mix problems, the allowable number and actions of operating personnel, the number of replications of benchmark problems in the benchmark mix, which programs may be resident in memory, maximum/minimum number of jobs/terminals active at any one time, and execution constraints, if any, should all be clearly stated. The LTD documentation should also specify that the benchmark demonstrations must use the same versions and releases of the software and hardware as proposed by the vendor in response to the RFP, unless waivers are granted by the Government.

Pre-execution and start-up requirements must be documented. This should include items such as preloading of programs, files, databases, etc. prior to the timed test demonstration. When modifications will be made to the benchmark data files immediately prior to the test (in order to reduce the effects of any vendor tuning to a specific set of data), the procedures for doing so should be clearly specified.

Benchmark validation data requirements must be specified. That is, data should be requested which allows the benchmark team to verify the accuracy of results, as well as the correct performance of the benchmark. Sources for such data might include accounting logs, console logs, printer listings, RTE logs, and hardware and software monitor data.

b. Functional Demonstrations

Instructions for performing functional demonstrations must also be specified, if any are to be performed. Functional demonstrations are usually designed to test certain mandatory requirements or desirable features that cannot be satisfactorily evaluated from vendor proposals or would not be appropriate for inclusion in a timed benchmark test. Examples are data file security, utility capabilities, speed and capabilities of unit record equipment, and start-up and shut-down procedures. Component parts of the functional demonstration should be keyed to specific requirements in the RFP that the functional demonstration is designed to test. Furthermore, at least the following should be explicitly described: the material to be provided by the Government or vendor, what the Government expects to observe, and the criteria used to determine the acceptability of a given functional demonstration. The reader is referred to FIPS PUB 42-1 for additional guidance on conducting functional demonstrations.

1.3 Develop Internal Agency Documentation

In addition to developing the above external documentation which goes to the responding vendors, the agency should also maintain its own internal documentation on such items as the technical and policy decisions that were made which affected the benchmark construction, the data used to develop the workload forecasts, and the sources from which benchmark problems and data files were obtained. This information may prove useful later, especially over long acquisition periods when changes to the benchmark team are likely to occur.

2. Test the Benchmark

There are several reasons for running each benchmark mix on computer systems other than the current one, especially on systems similar to those likely to be proposed by the vendors. Running the mix on other systems can provide valuable information on the transportability of the benchmark problems from one vendor's system to an another. Doing so can also determine the correctness and clarity of both the benchmark mix and the supporting documentation. For example, errors introduced into a benchmark package commonly involve incorrectly generated benchmark tapes, incompatibilities between the benchmark problems and the accompanying documentation, inconsistencies in the documentation, and even program logic errors. It is likely that these and other errors will be detected if the benchmark mix is run on one or more other systems, especially if performed by personnel other than those who designed the mix. Running the mix on other systems is also useful for determining the repeatability of the benchmark problems by comparing the execution results to the results obtained on the present system. It is likely that the numerical precision will not be identical on different vendor systems, but it should be determined if the difference in results is due to execution errors or to numerical precision differences on other vendor systems.

It should be noted that some of the same problems associated with running the benchmark on the agency's current system may exist here also, notably, the need for a separate machine to function as an RTE and the need for transaction or DBMS software. For this reason, if the complete benchmark cannot be run on another system, at least significant portions of it should be run to test its transportability.

Running the benchmark on other systems has value, although limited, for validating the benchmark timing. It also gives some insight into the size of the systems likely to be bid.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBS SP 500-113	2. Performing Organ. Report No.	3. Publication Date March 1984

4. TITLE AND SUBTITLE

Computer Science and Technology: Assessment of Techniques for Evaluating Computer Systems for Federal Agency Procurements

5. AUTHOR(S)

Helen Letmanyi

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

Same as item 6.

10. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 84-601012

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

The primary purpose of this document is the identification and qualitative assessment of computer system evaluation techniques for use during acquisition of computer systems. Also addressed is the identification of several criteria by which these alternative evaluation techniques may be compared and selected. The concepts presented in this study are applicable to all sizes of general purpose computers, from microcomputers to mainframes. Embedded or single-purpose computers, such as those used in weapon systems, have been excluded.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

acquisition; benchmarking; evaluation; instruction timing analysis; modeling; prototyping; rating charts analysis; system selection

13. AVAILABILITY

[X] Unlimited
☐ For Official Distribution. Do Not Release to NTIS
[X] Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

34

15. Price

USCOMM-DC 6043-P80

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SCIENCE & TECHNOLOGY

Superintendent of Documents,
Government Printing Office,
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Bureau of Standards Special Publication 500-.

Name ______________________________

Company ______________________________

Address ______________________________

City ______________ State ______________ Zip Code ______________

(Notification key N-503)

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent Bureau publications in both NBS and non-NBS media. Issued six times a year. Annual subscription: domestic $18; foreign $22.50. Single copy, $5.50 domestic; $6.90 foreign.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NBS under the authority of the National Standard Data Act (Public Law 90-396).

NOTE: The principal publication outlet for the foregoing data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NBS publications—FIPS and NBSIR's—from the National Technical Information Service, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.