DOCUMENT RESUME

ED 030 562 SE 006 884

By-Goodman, A. F.
Flow of Scientific and Technical Information: The Results of a Recent Major Investigation.
McDonnell Douglas Astronautics Co., Huntington Beach, Calif. Western Div.
Pub Date Apr 69
Note-61p.
EDRS Price MF-$0.50 HC-$3.15
Descriptors-*Engineers, *Information Needs, Information Retrieval, *Information Science, Information Storage, *Information Systems, *Scientists

Characterized were the scientific and technical information needs of 1,500 scientists and engineers from 73 companies, 8 research institutes, and 2 universities; and the flow of scientific and technical information (flow process) inherent in satisfying these needs. Interviewers asked 63 questions in the subject areas of (1) the user of scientific and technical information, (2) the user's most recent scientific or technical task, (3) the user's general utilization of the information system, and (4) the user's search and acquisition process for information used in task performance. Goals for the flow process, future analysis of the flow process, characterization of the flow process, and analysis of the flow-process data are summarized. Also provided are goals and future analysis recommendations. This investigation is the first attempt to obtain so much data on so large a position of the process, and its analysis is the first attempt to draw definitive and unifying conclusions from such data. (RS)

ED030562

DOUGLAS PAPER 4516
REVISED APRIL 1969

FLOW OF SCIENTIFIC AND TECHNICAL INFORMATION: THE RESULTS OF A RECENT MAJOR INVESTIGATION

A.F. GOODMAN

Presented to
14th International Meeting
The Institute of Management Sciences
Mexico City, Mexico
22-26 August 1967

SE 006 884

MCDONNELL DOUGLAS
CORPORATION

MCDONNELL DOUGLAS ASTRONAUTICS COMPANY
WESTERN DIVISION

ALSO INVITED FOR PRESENTATION TO:

75TH ANNUAL CONVENTION OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION, ON 1-5 SEPTEMBER 1967.

1967 ANNUAL CONVENTION OF THE AMERICAN DOCUMENTATION INSTITUTE, ON 22-26 OCTOBER 1967.

AMERICAN MANAGEMENT ASSOCIATION COURSE, "FUNDAMENTALS OF INFORMATION RETRIEVAL SYSTEMS," ON 5-7 JUNE 1968, 20-24 JANUARY 1969, AND 21-25 APRIL 1969.

FLOW OF SCIENTIFIC AND TECHNICAL INFORMATION: THE RESULTS OF A RECENT MAJOR INVESTIGATION

by

A. F. Goodman
Senior Technical Staff to Vice President
Information Systems Subdivision
McDonnell Douglas Astronautics Company — Western Division

ABSTRACT

The investigation characterized the scientific and technical information needs of 1,500 scientists and engineers from 73 companies, 8 research institutes, and 2 universities; and the flow of scientific and technical information (flow process) inherent in satisfying these needs. Interviewers asked 63 questions in the following four subject areas: (1) the USER of scientific and technical information, (2) the user's most recent scientific or technical TASK, (3) the user's general UTILIZATION of the information system, and (4) the user's SEARCH AND ACQUISITION process for information used in task performance.

Many studies have been performed, and much has been written, concerning the flow process. The tendency has been to examine only small portions of the process, or to speculate about large portions of the process in generalities. Therefore, very little of a comprehensive, definitive, and unifying nature actually has been said about the process. This investigation is the first attempt to obtain so much data on so large a portion of the process, and its analysis is the first attempt to draw definitive and unifying conclusions from such data.

Goals for the flow process, future analysis of the flow process, characterization of the flow process, and analysis of flow-process data are summarized. In addition, the goals and future analysis recommendations reflect work performed by the author after completion of the study.

DEDICATION

This paper is respectfully dedicated to the inspiring memory of Dr. Edith S. Jay, whose scientific ability was as brilliant in the unstructured area of vague abstraction as it was in the highly structured area of extreme detail. She made a unique contribution to each colleague and to each task.

ACKNOWLEDGMENTS

The investigation was sponsored by the United States Government, and was performed while the author was with North American Rockwell Corporation. Its successful completion is attributable to the efforts of many people.

Interviews were made possible by the cooperation of the participating organizations listed in the Appendix, and the sample of 1,500 scientists and engineers employed by them. The author gratefully acknowledges the guidance of Mr. Walter M. Carlson and Mr. Howard B. Lawson; and the technical contributions of Mr. John D. Hodges, Jr., Mr. Forrest G. Allen, Mr. Bruce W. Angalet, Mrs. Philotheos J. Mazzagatte, Mr. Richard B. McCord, and Mrs. Carol C. Taylor. In addition, he sincerely thanks Mr. Karl H. Meyer, Mr. Solomon L. Pollack, Mr. Hallock G. Davis, Jr., and Mr. Robert J. Mason, Jr. for their management support; and Miss Darnell Gentry, Mr. William R. Meyers, Mr. Martin Cutler, Mr. John F. Duewell, Mrs. Marian E. Farnsworth, Mr. Roland K. Jacobson, Dr. Edith S. Jay, Mr. Leonard B. Jenson, Mr. Spencer B. McCain, Dr. Franklyn J. Michaelson, Mr. William E. Nelson, Mr. Louis J. Precht, Mr. Carroll M. Shipplett, Mr. Keith V. Smith, and Mr. Hagop H. Terzagian for their operational support.

This paper summarizes the documentation of the investigation, which appears in Reference 1. Opinions expressed herein are those of the author, and do not necessarily represent the view of the United States Government.

INTRODUCTION

A major investigation to determine how scientists and engineers acquire information has been recently completed. The objective of the investigation was to characterize scientific and technical information needs, and the flow of scientific and technical information (flow process) required to satisfy these needs. The study's conclusions are as important to individual organizations as they are to the government, and as important to scientific and technical management as they are to those directly concerned with the flow process.

Data were obtained by personal interviews, with a representative sample of 1,500 from a population of approximately 120,000 scientists, engineers, and technical personnel. These personnel were employed by 73 companies, 8 research institutes, and 2 universities. The Appendix lists participating organizations, with the number of personnel interviewed from each.

To ensure high-quality data, the interviewers were thoroughly trained, and the interviews were carefully recorded and checked for accuracy and consistency. The interviewers asked 63 questions in the following four subject areas: (1) the USER of scientific and technical information, (2) the user's most recent scientific or technical TASK, (3) the user's general UTILIZATION of the information system, and (4) the user's SEARCH AND ACQUISITION process for information used in task performance.

Many studies have been performed, and much has been written, concerning the flow process. The tendency has been to examine only small portions of the process, or to speculate about large portions of the process in generalities. Therefore, very little of a comprehensive, definitive, and unifying nature actually has been said about the process. This investigation is the first attempt to obtain so much data on so large a portion of the process, and its analysis is the first attempt to draw definitive and unifying conclusions from such data. During this analysis, qualitative question responses were transformed into numerical form, a process model for relationships among questions was constructed and estimated, and numerical relationship results were transformed back to qualitative form.

Goals for the flow process, future analysis of the flow process, characterization of the flow process, and analysis of flow-process data are discussed in subsequent sections. This discussion summarizes the investigation, which is completely described in Reference 1. In addition, the goals and future analysis recommendations reflect work performed by the author after the publication of Reference 1.

Reference 2 presents the application, to a process or system in general, of the recommended program for analysis and optimization, as well as the first two portions of the analysis. Computer programs used in the analysis are documented in Reference 3.

The surveyed organizations constitute a reasonable cross-section of scientific and technical organizations in general, although they were selected on the basis of being defense contractors (see Appendix). In the absence of a comparably comprehensive and definitive investigation of the flow process in general, it is informative to view the results of the study as generally indicative, if not actually applicable.

For this reason, the terminology employed here has been selected to minimize dependence upon the defense industry. The correspondence between the terminology and that of Reference 1 is as follows:

- User's salary level replaced user's equivalent GS rating.
- Documentation Center replaced Defense Documentation Center.
- Government Information Center replaced DOD Information Center.

GOALS FOR THE FLOW PROCESS

The conclusions of the investigation provide a set of goals for the flow process, and a set of measures with which to evaluate a general information system. These goals are supported by the characterization of the flow process below, and the numerical results which appear in Volume III of Reference 1.

THE FLOW PROCESS

Figure 1 is helpful in visualizing the goals described by the remainder of this section. It represents either of the following processes:

- The flow process in task performance, when UTILIZATION represents the utilization of the information system in task performance.
- The flow process in general, when TASK represents the user's scientific or technical task in general.

FLOW PROCESS M-43395

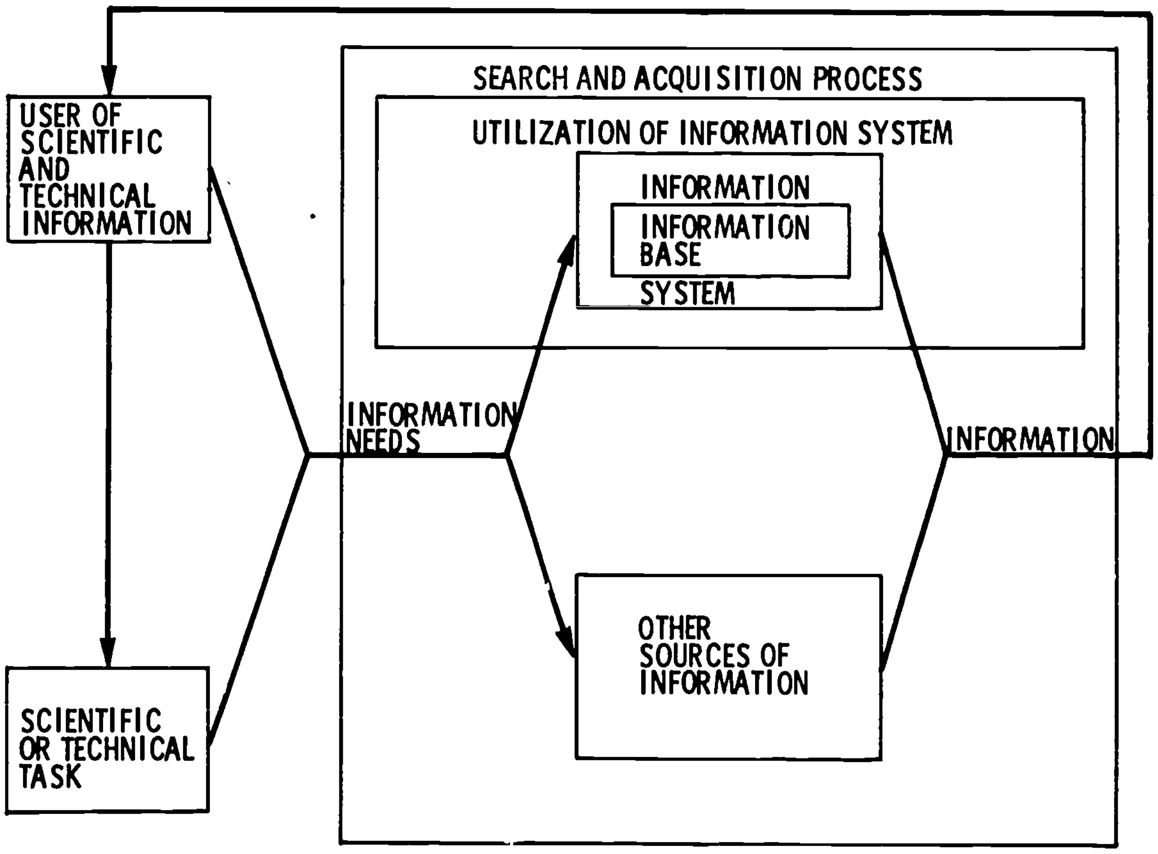


FIGURE 1

BRIDGE THE INFORMATION GAP

An information gap exists between the user of scientific and technical information, and the information system which serves his needs. This information gap must be bridged if the user is to obtain high-quality information.

REORIENT THE USER AND THE INFORMATION SYSTEM

Both the user and the information system need to be reoriented. Scientists and engineers, especially those in management or those possessing an advanced degree, must become active seekers of high-quality information services. For its part, the information system must become an active provider of high-quality information services, not merely a passive document repository.

EXPAND THE INFORMATION BASE

An information base forms the foundation of the information system. In general, it contains information which is conceptual and research-oriented. The information base has to be expanded to include design and performance information, and information which is development- and production-oriented.

RESTRUCTURE THE INFORMATION BASE

The information base is composed of information media which convey the information. For the most part, these media are written in form, formal in composition, and textual in layout. It must be restructured to include media which are oral in form, informal and semiformal in composition, and graphical in layout.

MAKE THE INFORMATION BASE FLEXIBLE

The information base should be made flexible to permit:

- Information to be indexed, abstracted, selectively organized, and selectively analyzed.

- Information to be selectively repackaged in information media of appropriate form, composition, and layout.
- Information media to be indexed and abstracted.

MAKE THE INFORMATION BASE MOBILE

The information base needs to be made mobile, so that information awareness is automatic, rapid, and selective; and information acquisition is quick and easy.

EXPAND THE INFORMATION SYSTEM

Expert personnel must be employed to expand the information system by providing both information resources and connections with the informal information system ("invisible colleges"). This expansion will add an entirely new dimension to the information system.

EXTEND THE INFORMATION SYSTEM

The information system has to be extended into the local work environment by the automatic and selective dissemination of abstracts for media in the information base, and listings of disciplinary areas with an expert's level of competence in each area.

FUTURE ANALYSIS OF THE FLOW PROCESS*

The investigation has generated a great deal of valuable data concerning the flow process. Analysis of the data, despite funding and time limitations inherent in an exploratory study, has yielded considerable insight into the flow process. This analysis also indicates that certain portions of the flow process merit additional investigation, and that certain portions of the analysis merit refinement.

A complex, but as-yet incompletely characterized, relationship exists between the flow process and the performance of scientific and technical tasks. To improve task performance (improve quality, reduce cost, or shorten time), both the government and individual organizations have made large investments in improvement of the flow process by improvement of the information system (see Figure 1). Sufficient improvement in (optimization of) the flow process is achieved, when sufficient improvement in (optimization of) task performance is achieved.

Therefore, the additional investigation should be performed in the framework of a general program for analysis and optimization of the flow process with respect to task performance.

The analysis provides the basis for this program of analysis and optimization in the following manner:

- A model of the flow process with which to plan investigations and perform analyses.
- An analytical approach to transform qualitative question responses into numerical form, to construct and estimate a process model for relationships among questions, and to transform the numerical relationship results back to qualitative form.

The application of the first two portions of the above-mentioned analytical approach to a process or system in general is contained in Reference 2.

---

*The future analysis recommendations should be assigned priorities according to the twin criteria of objectives and available resources.

ADDITIONAL INVESTIGATION

Investigation of the following areas appears promising:

- The feasibility of the conclusions, and their effect upon the flow process.
- The effect of the quality and timeliness of information acquired upon the quality, cost, and speed of task performance.
- The difficulties encountered in the utilization of the information system, with an emphasis upon separating those attributable to inside the organization from those attributable to outside the organization.
- The users who, though cognizant of certain portions of the information system, do not use them.
- The utilization of the information system in task performance.
- Those areas suggested by refined analysis of the data.

PROGRAM FOR ANALYSIS AND OPTIMIZATION

The flow process (Figure 1) is quite complex, and experimentation (investigation) regarding it is both difficult and expensive. For such a process, mathematical solution for outputs in terms of inputs is usually not feasible, and computer simulation is often an effective and efficient complement to experimentation.

When a model (mathematical representation) for the process is translated into a simulation computer program (computer representation) for the process, the process and the effects of various factors upon it may be simulated. The accuracy and precision of the computer simulation increase as the accuracy and precision of the model increase, which occurs as knowledge concerning the process increases.

Four periods occur in the evolution of a body of knowledge as it matures from an art into a science; these are description, modeling, prediction, and control and optimization. Computer simulation yields appropriate results in the modeling, prediction, and control and optimization periods. With the completion of this investigation, knowledge concerning the flow process is emerging from the description period and entering into the modeling period.

A program that provides a meaningful framework for the coordination of experimentation and computer simulation in the analysis and optimization of the flow process is illustrated by Figure 2, and is composed of the following 10 basic stages:

1. Quantitative process analysis develops a model by transforming qualitative elements of the process into numerical form, and by constructing a model for relationships among component parts of the process. The transformation of elements is accomplished by arranging the elements into an informative detailed (local) structure, and then associating a meaningful number with each element. Construction of the model is accomplished by arranging the component parts into an informative general (global) structure, and then specifying the general form for meaningful relationships among component parts.
2. Experimental trial produces experimental data.
3. Model estimation produces, from experimental and available auxiliary data, estimates of unspecified constants in the general form of relationships in the model; and a preliminary, but insufficient, evaluation and validation (positive check) of the process representation by the model.

M-43385

PROGRAM FOR ANALYSIS AND OPTIMIZATION

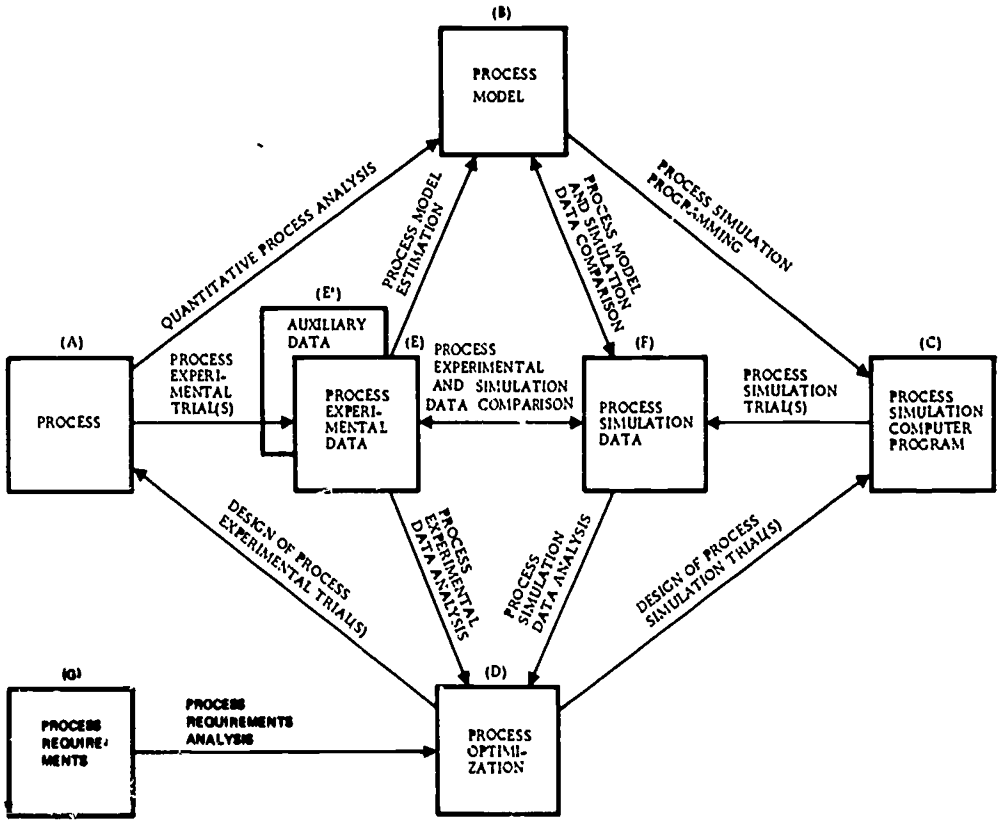


FIGURE 2

4. Simulation programming translates the model into a simulation computer program.
5. Simulation trial produces simulation data.
6. Model and simulation data comparison provides an evaluation and validation of the model's representation by the simulation computer program.
7. Experimental and simulation data comparison provides an evaluation and validation of the process' representation by the combination of model and simulation computer program. Thus, the process' representation by the model is indirectly evaluated and validated, given that the model's representation by the simulation computer program has been validated.
8. Experimental and simulation data analysis characterizes and evaluates the process, in terms of criteria and constraints, and suggests modification of the process for optimization (achievement of sufficient improvement).
9. Optimization modifies the process and applies appropriate stages of the program to the modified process, in an iterative manner until sufficient improvement is achieved.
10. Design of experimental and simulation trials aids optimization.
11. Requirements analysis provides a basis for optimization and design of experimental and simulation trials.

For the design of a new flow process, the approach of this program may be modified to yield a program for design and optimization (Figure 3). The application to a process or system in general, of both the program for analysis and optimization and the program for design and optimization, is presented in Reference 2.

REFINED ANALYSIS OF THE DATA

Because only a small fraction of the effort expended in collecting data is typically devoted to their analysis, a large amount of the information they contain generally is undiscovered and unexploited. A more profound understanding of the flow process may be achieved through more refined analysis of the data, as follows:

- Investigation of the effect of organization size, industry, and interviewer bias upon answers to questions.

PROGRAM FOR DESIGN AND OPTIMIZATION

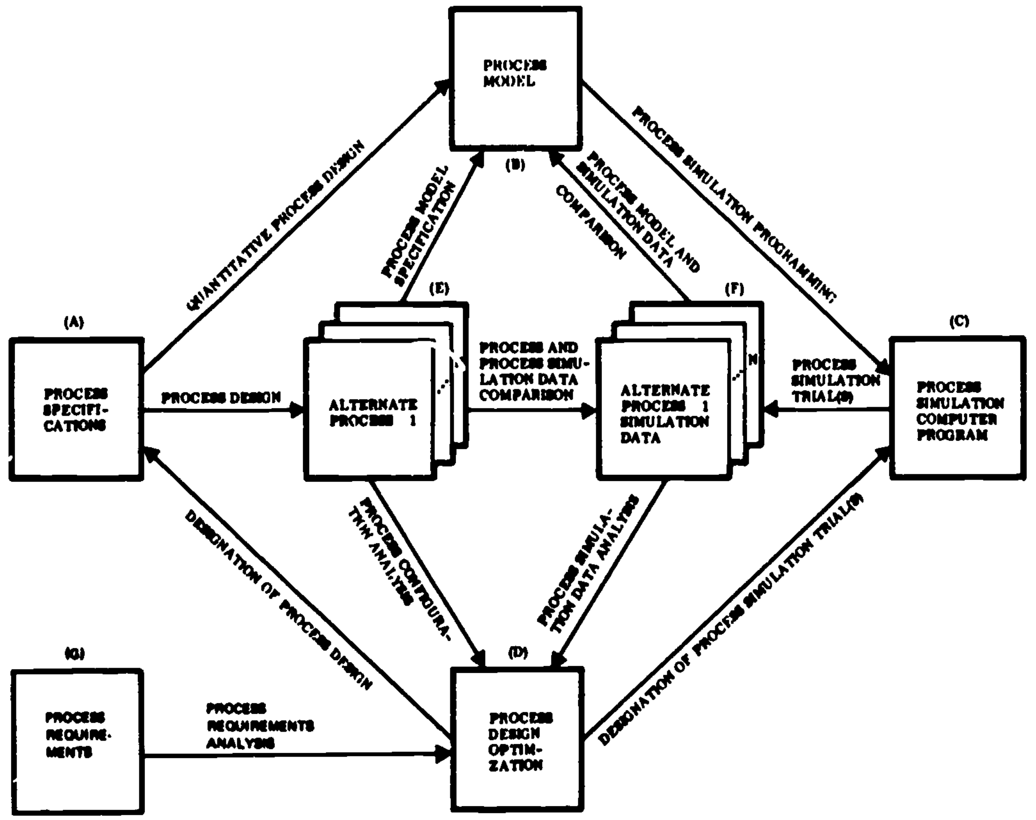


FIGURE 3

- Improvement in the arrangement of responses to a question, and the association of a numerical value with each response, with the objective of improving the linearity of relationships among questions.
- Division of the 1,500 users into appropriate groups for analysis and comparison, such as the three groups formed by those who acquired: conceptual information, design and performance information, and production information.
- Incorporation into the analysis of differences between the corresponding characteristics of the desired and actually received information, and additional special indexes.
- Reformulation and re-estimation of appropriate linear relationships among questions, to reflect the above improvements and to investigate more specific relationships which involve only single questions (rather than combinations of related questions).
- Formulation and estimation of additional linear relationships within the flow process, such as those which reverse the input/output relations of the flow process (under the characterization below)--for the investigation of the selective dissemination of information.

CHARACTERIZATION OF THE FLOW PROCESS

The findings of the investigation, which characterize the flow process, are highlighted in this section. They are illustrated by the accompanying figures, and supported by the numerical results in Volume III of Reference 1.

TYPES OF INFORMATION

Almost one-half of the information was in engineering fields, and almost two-fifths of it was in scientific fields (see Figure 4). In the conceptual-design and performance-production cycle, over 60% of the information involved design and performance (Figure 5).

MEDIA FOR CONVEYING INFORMATION

Oral information was wanted more than one out of three times, and semi-formally written information also was wanted more than one out of three

FIELD OF INFORMATION

M-43388

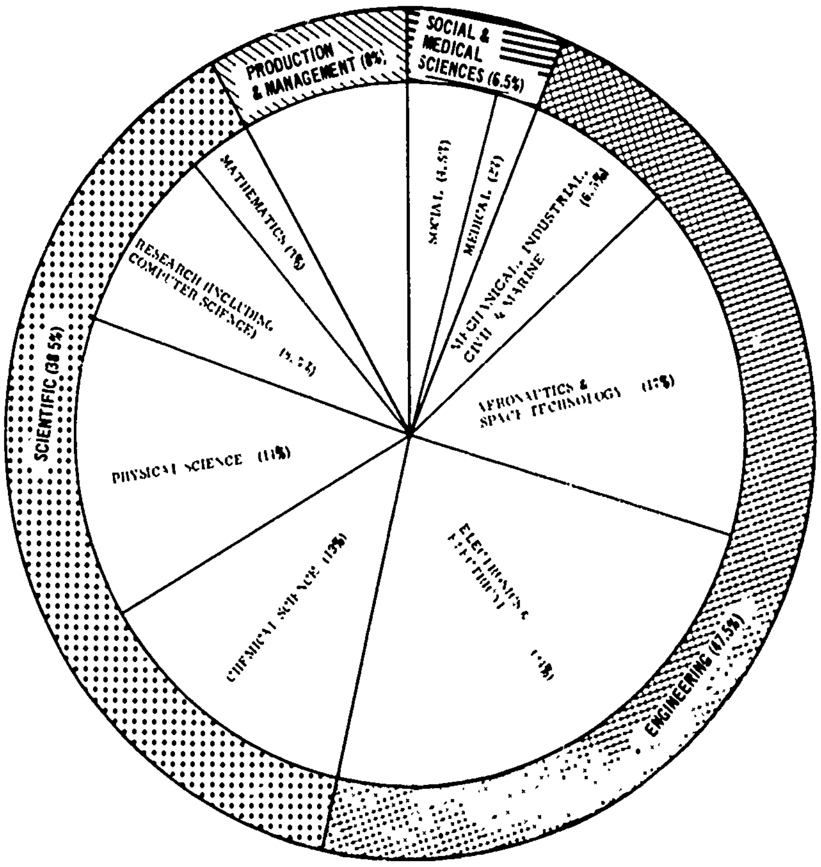


FIGURE 4

CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF INFORMATION M-43391A

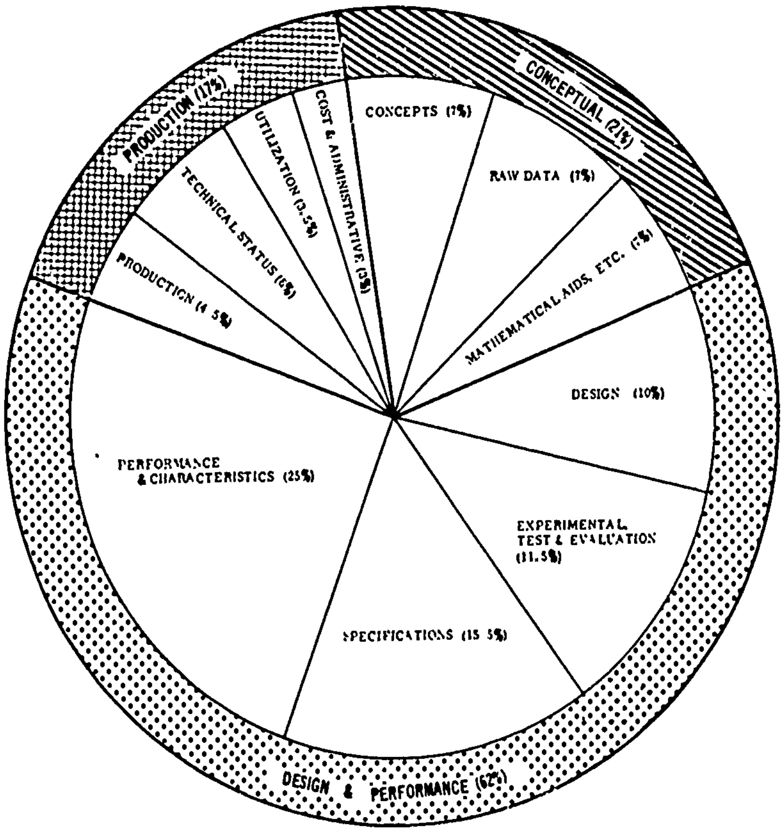


FIGURE 5

times (Figure 6). Over 60% of the information was desired in more than one document (see Figure 7). Almost three-fifths of the time, a specific answer was needed; and over one-third of the time, a detailed analysis was needed (see Figure 8).

FIRST SOURCE CONTACTED FOR INFORMATION

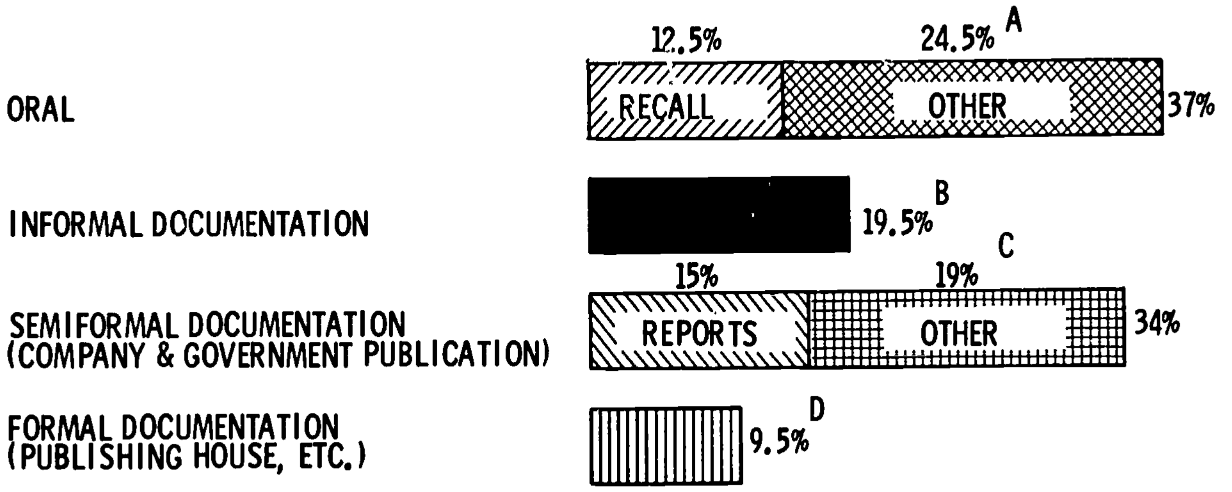
Eighty percent of the time, the users first searched for information within the local work environment (see Figure 9). The local work environment extends only as far from the user as an internal consultant. It does not extend as far from him as his organization's technical information center (library), which is his connection with the formal information system.

ACQUISITION TIME FOR INFORMATION

Almost one-half of the information was needed within 7 days, and almost three-fourths of it was needed within 30 days (Figure 10). Except for 5% of

DESIRED COMPOSITION OF INFORMATION MEDIA

M-43406



A. THOSE RESPONSES WITH OVER 3 PERCENT ARE: "ORAL CONTACTS-ALL OTHER" (18%) AND "ORAL CONTACTS WITH MANUFACTURER" (3.5%).

B. THOSE RESPONSES WITH OVER 3 PERCENT ARE: "PERSONAL NOTES, PERSONAL LOGS AND PERSONAL FILES" (3%); "CORRESPONDENCE, MEMOS AND TWX" (6%); AND "DRAWINGS AND SCHEMATICS" (5%).

C. THOSE RESPONSES WITH OVER 3 PERCENT ARE: "SYSTEM SPECIFICATION DOCUMENTS" (4.5%) AND "MANUALS" (3.5%).

D. THOSE RESPONSES WITH OVER 3 PERCENT ARE: "JOURNALS" (4.5%) AND TEXTBOOKS" (3.5%).

FIGURE 6

DESIRED VOLUME OF INFORMATION MEDIA

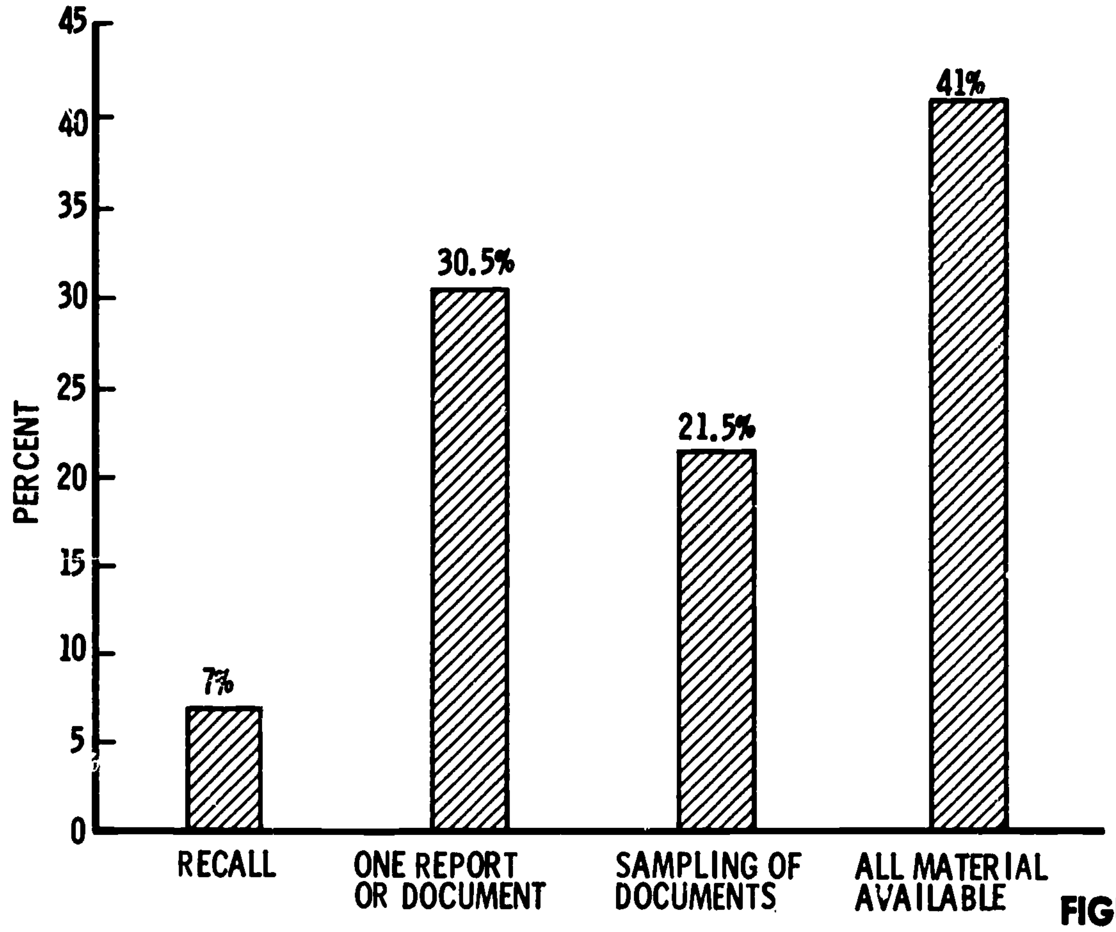


FIGURE 7

DESIRED DEPTH OF INFORMATION MEDIA

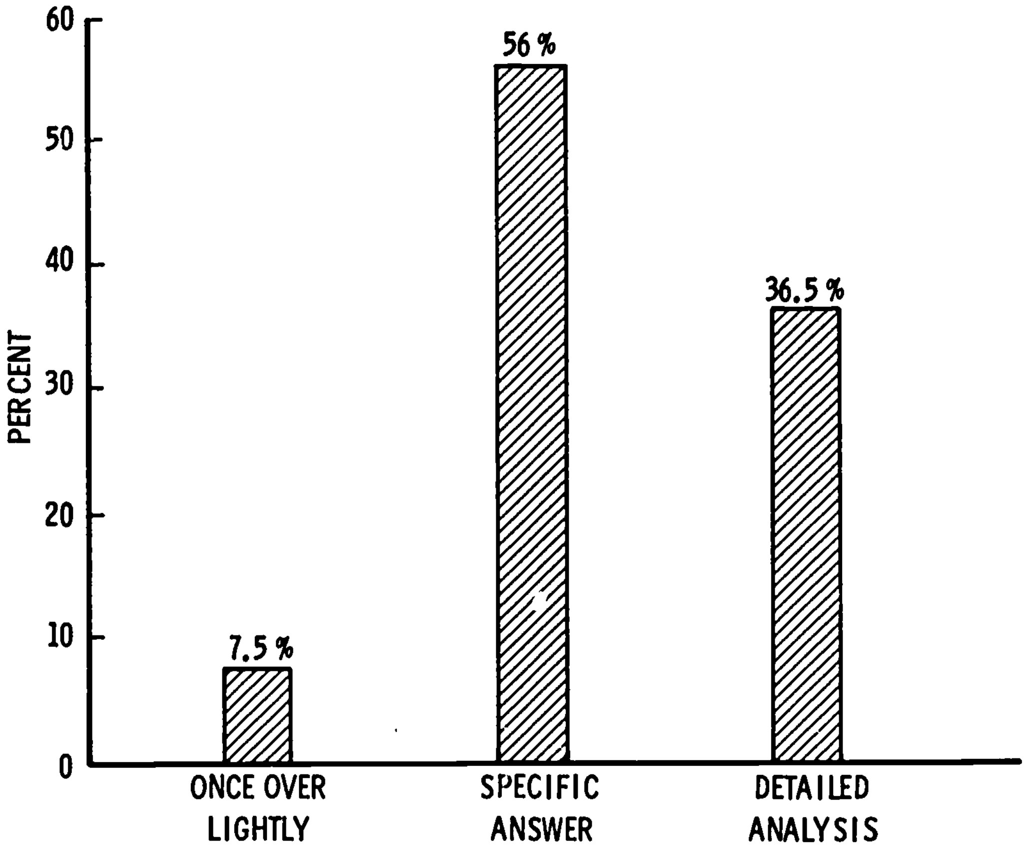


FIGURE 8

M-43389

FIRST SOURCE CONTACTED FOR INFORMATION

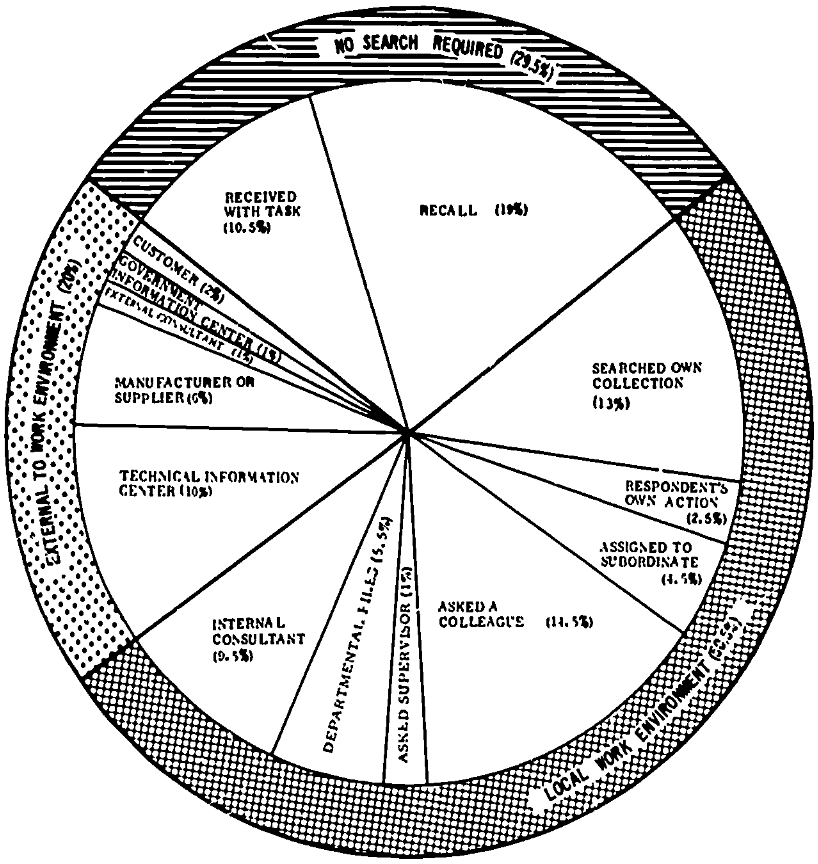


FIGURE 9

DESIRED ACQUISITION TIME FOR INFORMATION

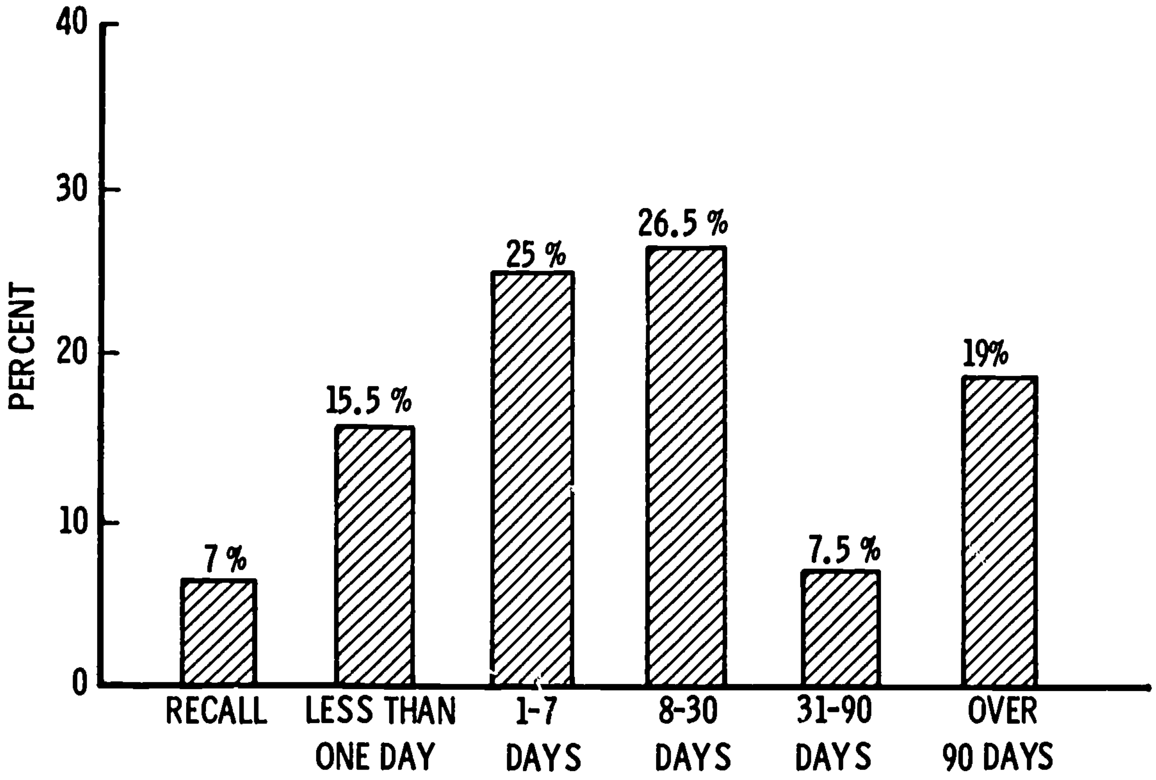


FIGURE 10

the information, the information needs were satisfied within the allowable acquisition time (see Figure 11).

UTILIZATION OF INFORMATION

Over two-fifths of the information was used throughout the entire task, and over one-third of it was used in major portions of the task (Figure 12). Almost 80% of the information was absolutely essential to the task, and over 15% of it was extremely helpful in the task (see Figure 13).

UTILIZATION OF THE INFORMATION SYSTEM

Of the users, 95% utilized their organization's technical information center (library), and over 50% utilized it twice a month or more (Figure 14).

Title listings or abstracts of information media would have been useful for finding more than two-fifths of the needed information. However, the

TIMELY ACQUISITION OF INFORMATION

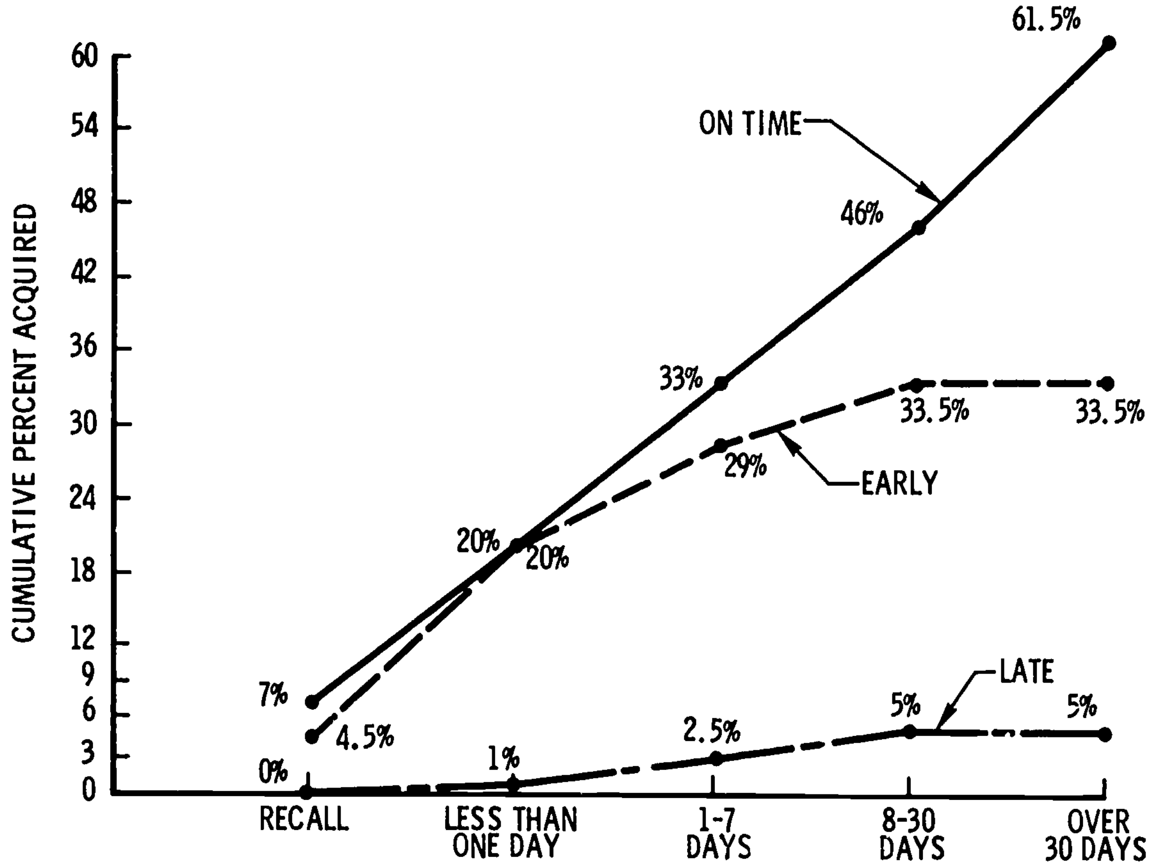


FIGURE 11

EXTENSIVENESS OF INFORMATION UTILIZATION

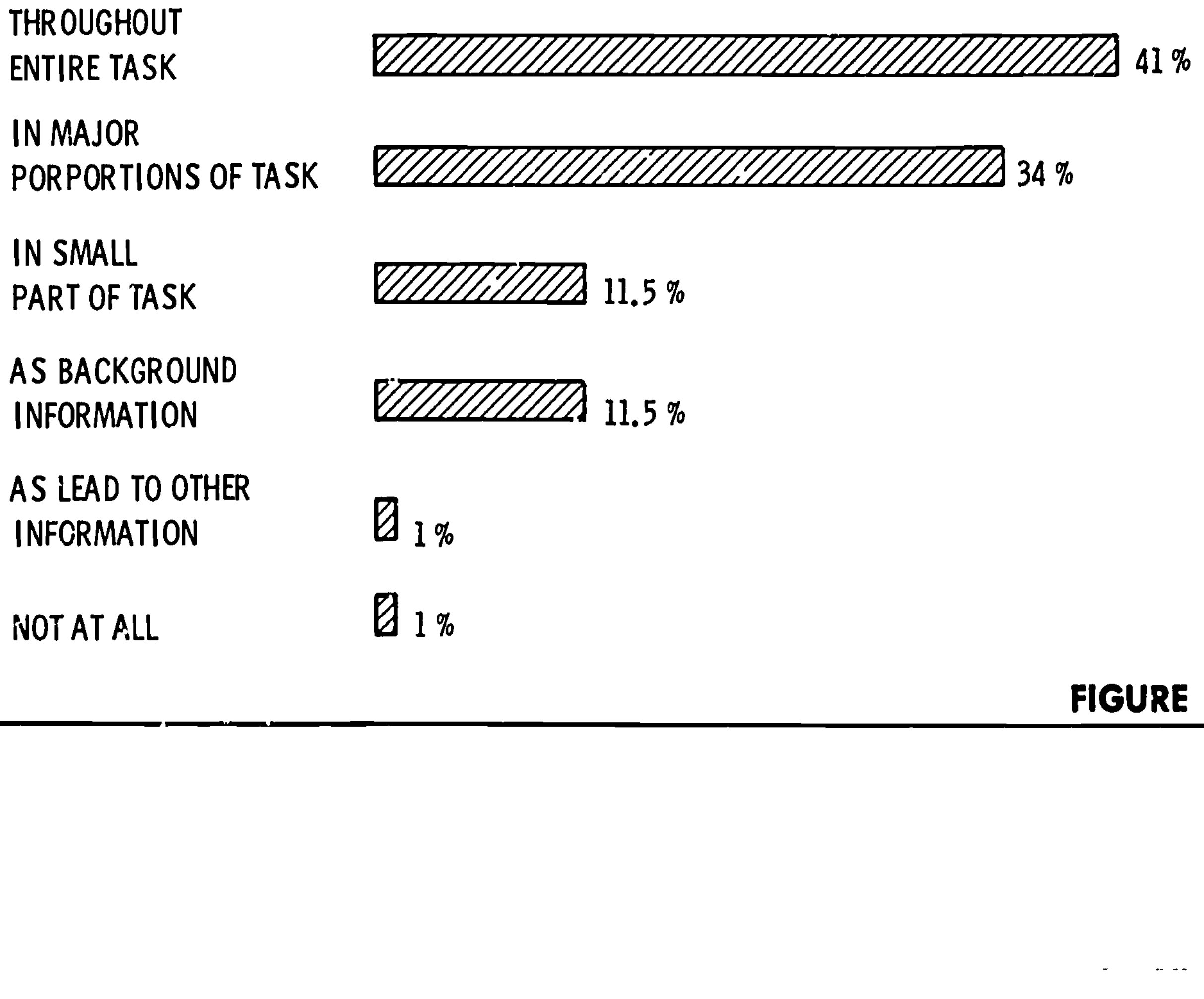


FIGURE 12

ESSENTIALITY OF INFORMATION UTILIZATION

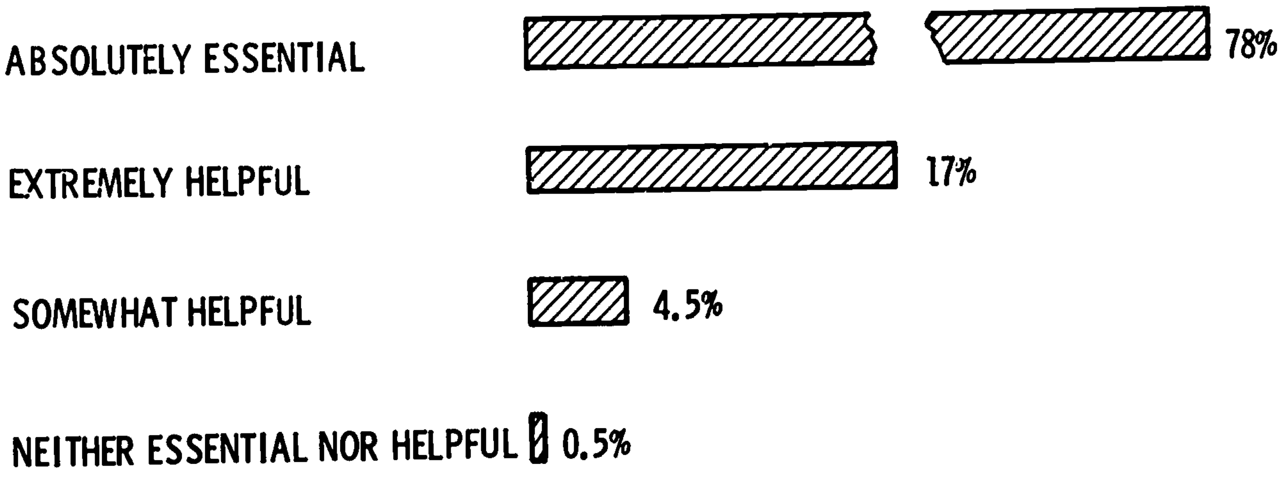


FIGURE 13

M-43399

UTILIZATION OF ORGANIZATION'S TECHNICAL INFORMATION CENTER

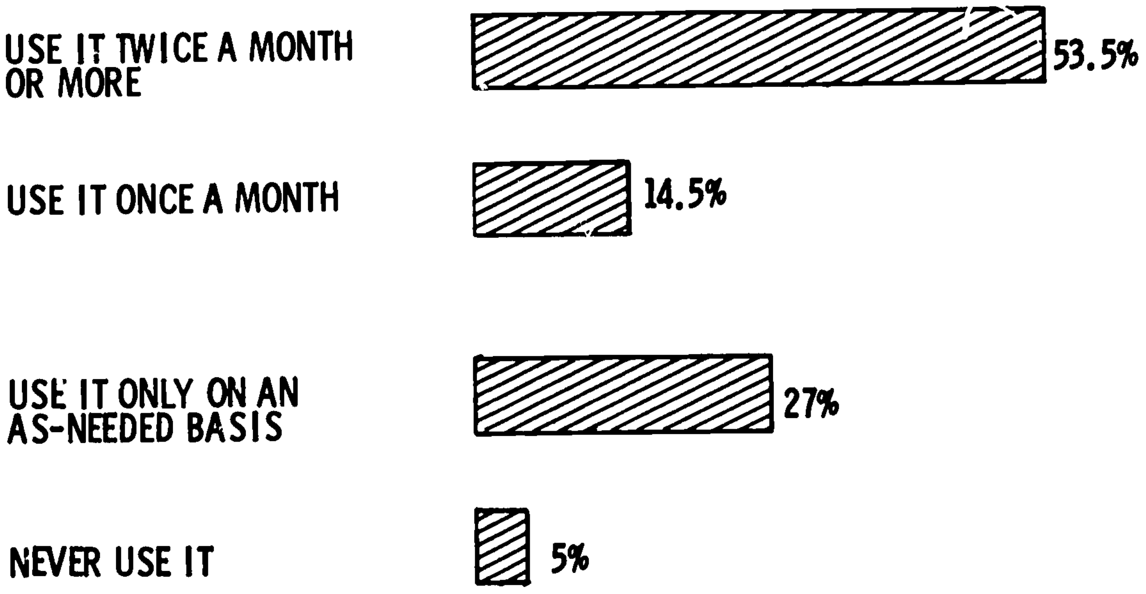


FIGURE 14

Technical Abstract Bulletin (TAB) was utilized by less than two out of five users; and it was unknown to over two out of five users (Figure 15). Less than 20% of the users utilized the Scientific and Technical Aerospace Reports (STAR), while over 60% of them did not know of it (Figure 16).

Over two out of five users encountered difficulties in the utilization of the information system. Lack of timely awareness of information accounted for almost two-fifths of these difficulties, and lack of timely acquisition of information accounted for over one-half of them (see Figure 17).

SCIENTIFIC OR TECHNICAL TASKS

More than 50% of the tasks were in engineering fields, and more than 30% of them were in scientific fields (see Figure 18). In the research-development-production cycle, almost two-thirds of the tasks were development (Figure 19). Two out of three tasks involved design and performance, within the conceptual-design and performance-production cycle (see Figure 20).

M-43405

UTILIZATION OF TECHNICAL ABSTRACT BULLETIN

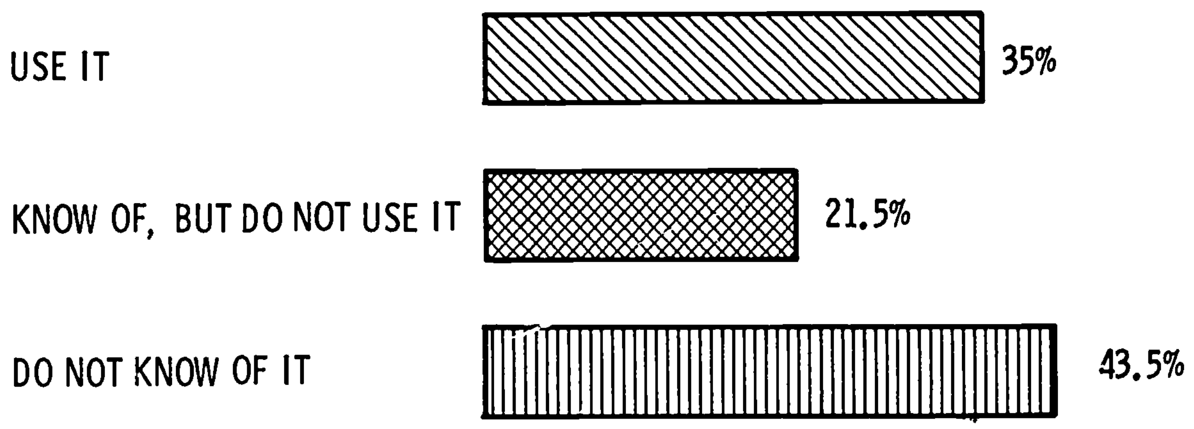


FIGURE 15

M-43404

UTILIZATION OF SCIENTIFIC AND TECHNICAL AEROSPACE REPORTS

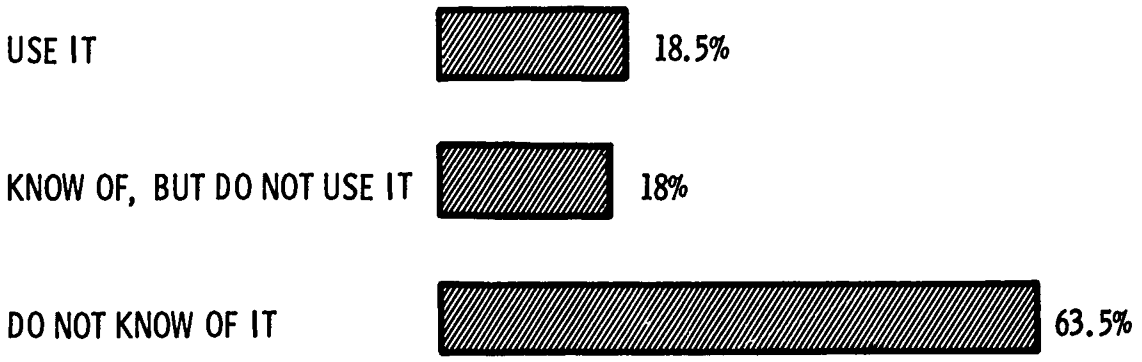
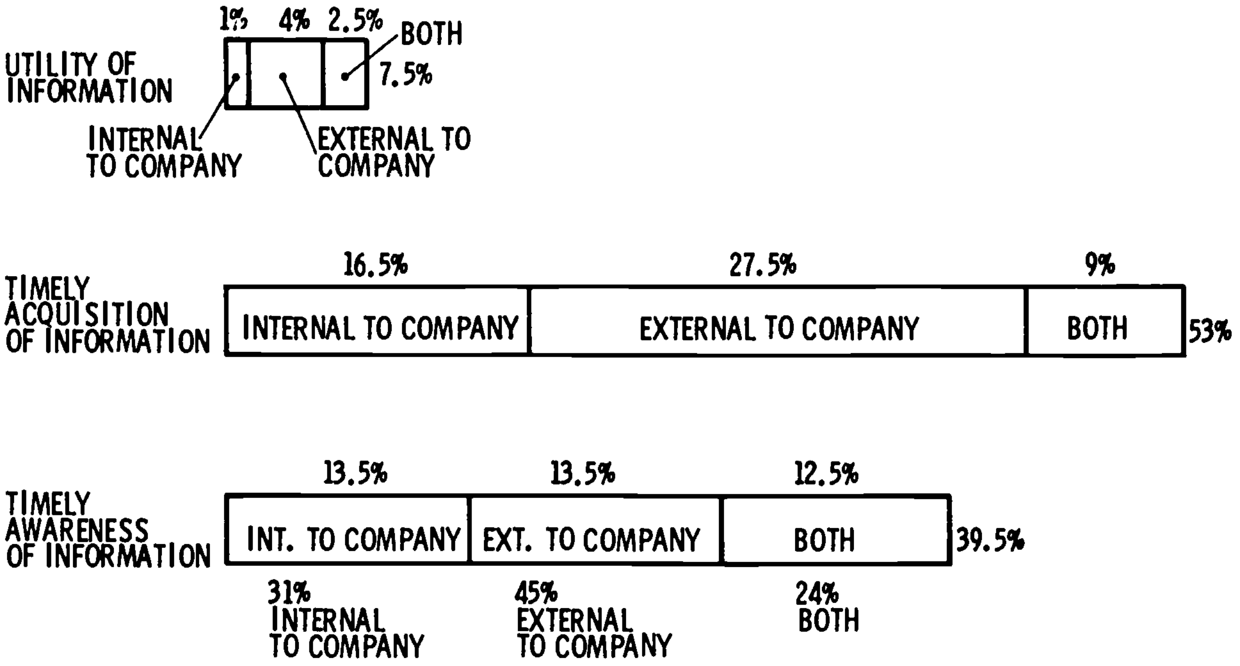


FIGURE 16

M-43403

UTILIZATION AWARENESS, ACQUISITION AND UTILITY DIFFICULTIES



*BASED ON THE CATEGORIZATION OF 628 APPROPRIATE NARRATIVE ANSWERS, OF THE 639 ANSWERS TO THE QUESTION.

FIGURE 17

FIELD OF TASK

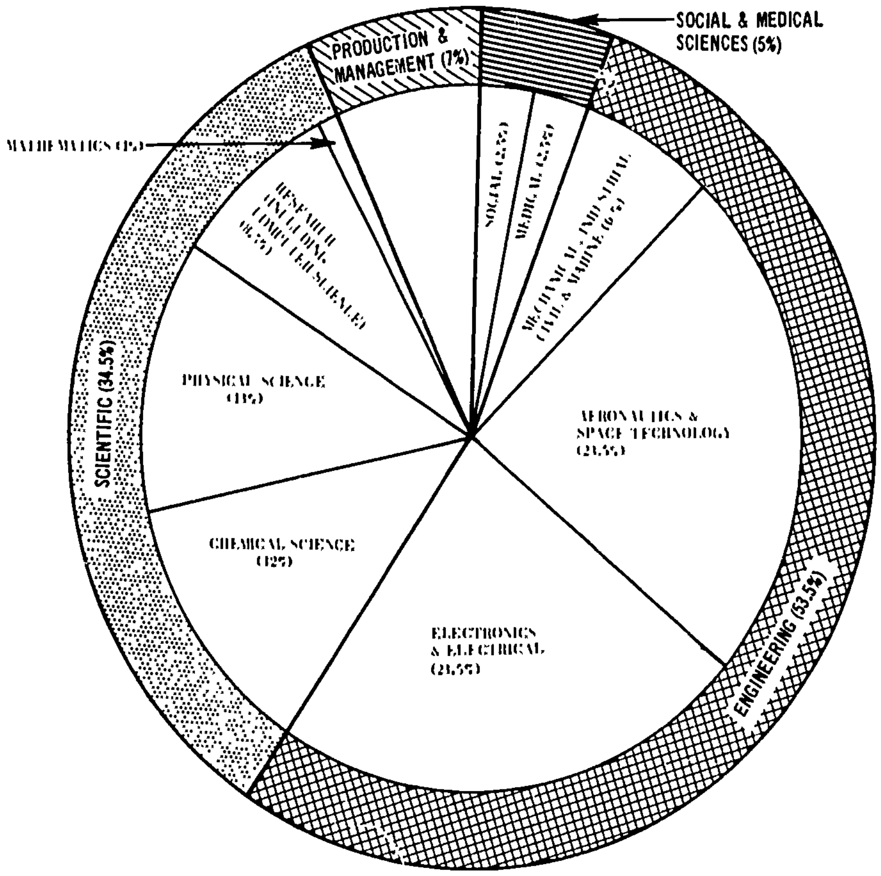


FIGURE 18

M-43386

RESEARCH - DEVELOPMENT-PRODUCTION CYCLE LOCATION OF TASK

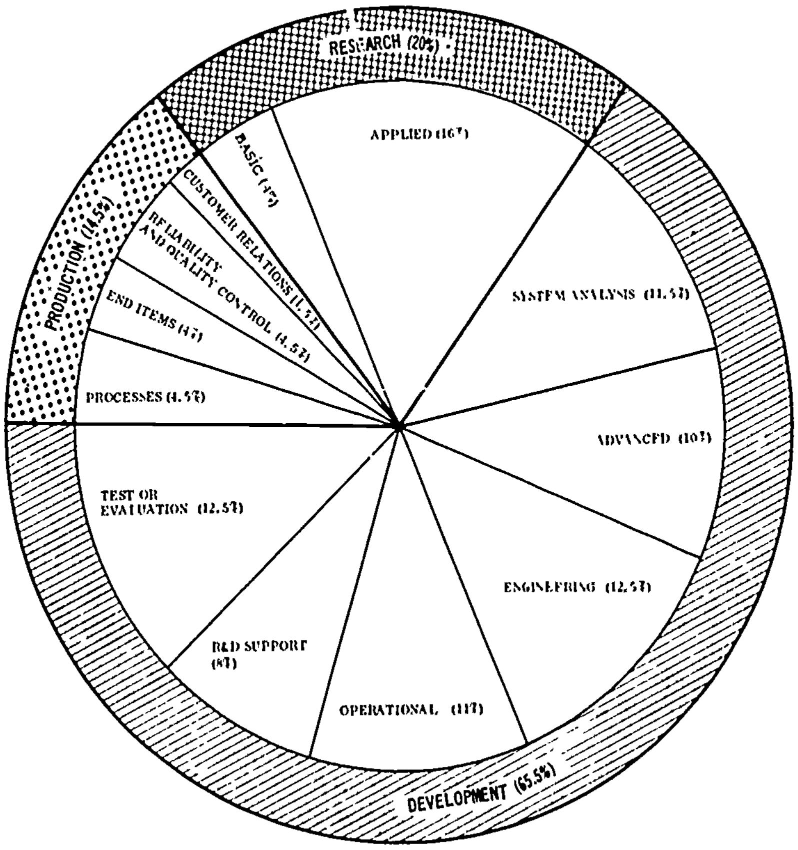


FIGURE 19

CONCEPTUAL - DESIGN AND PERFORMANCE - PRODUCTINON CYCLE LOCATION OF TASK

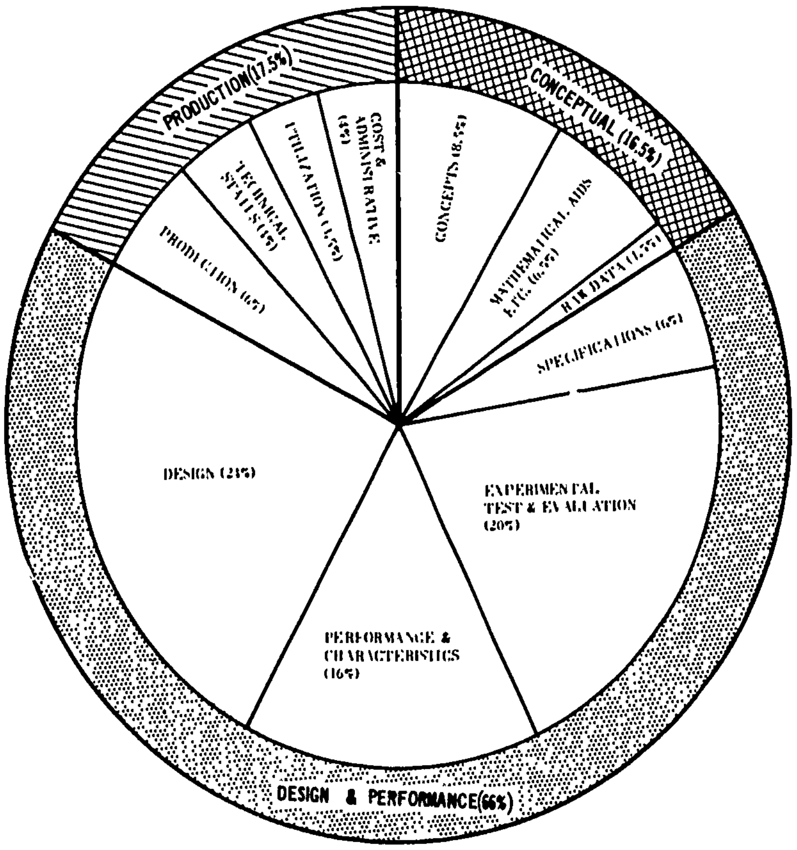


FIGURE 20

USERS OF SCIENTIFIC AND TECHNICAL INFORMATION

Over one-half of the users held engineering positions, and almost one-third of them held scientific positions (Figure 21). In the research-development-production cycle, two out of three users occupied development positions (Figure 22). Of the users, 40% were not managers, and over 30% managed from one to five persons (see Figure 23). More than one-half of the users possessed a bachelor's degree, and almost one-third of them possessed an advanced degree (see Figure 24).

In general, these significant users of scientific and technical information were the real users of the information system--and also the ones most frustrated by difficulties involving its use.

USER'S FIELD OF POSITION

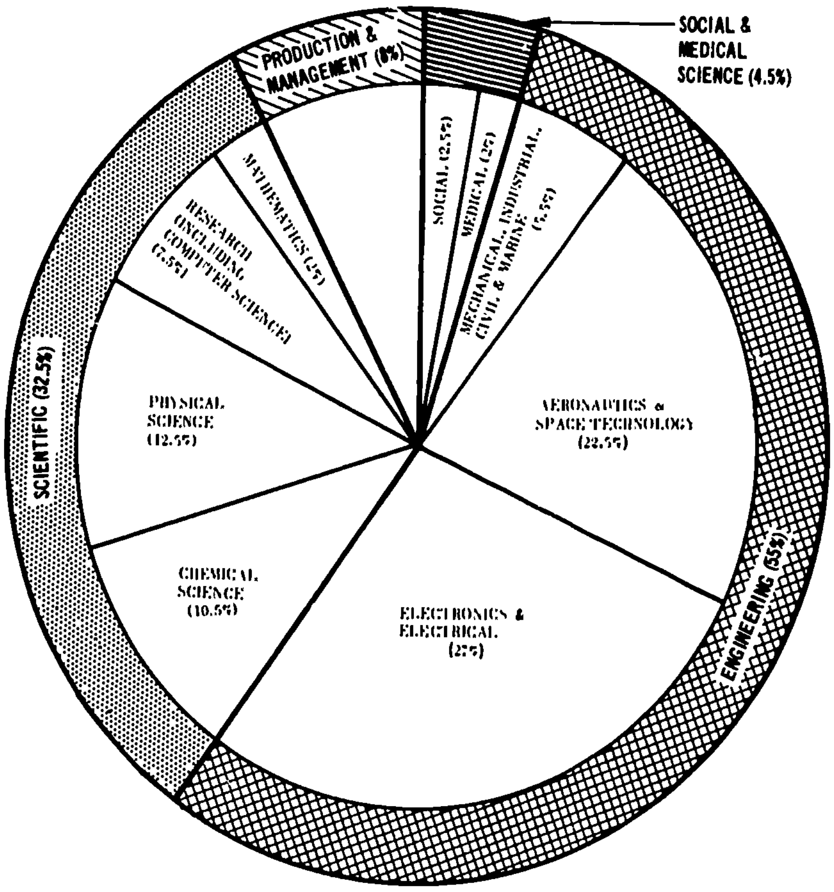


FIGURE 21

USER'S RESEARCH - DEVELOPMENT - PRODUCTION CYCLE LOCATION OF POSITION

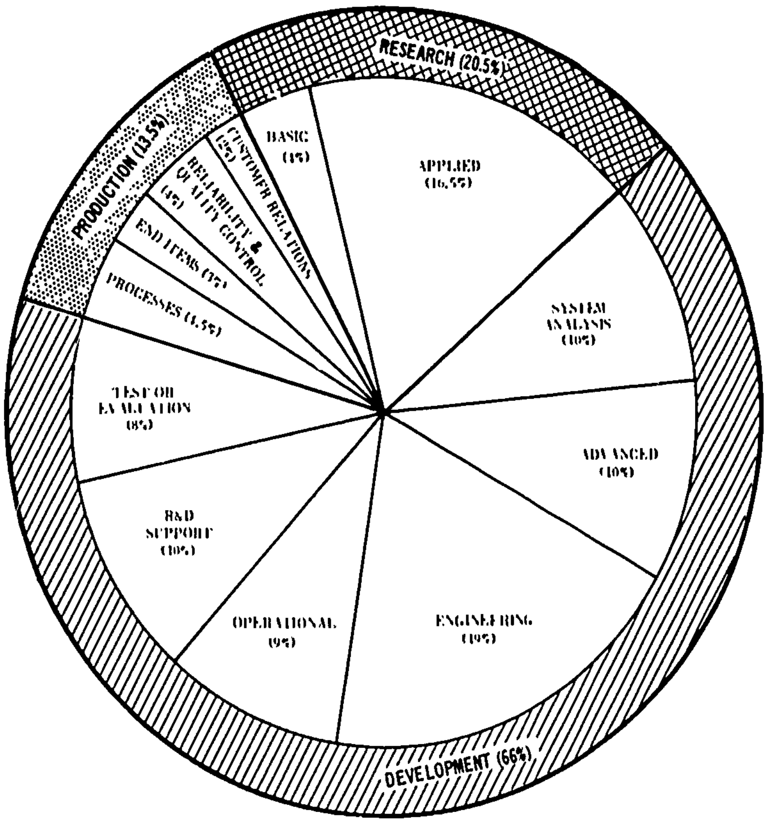


FIGURE 22

NUMBER OF PERSONNEL MANAGED BY USER

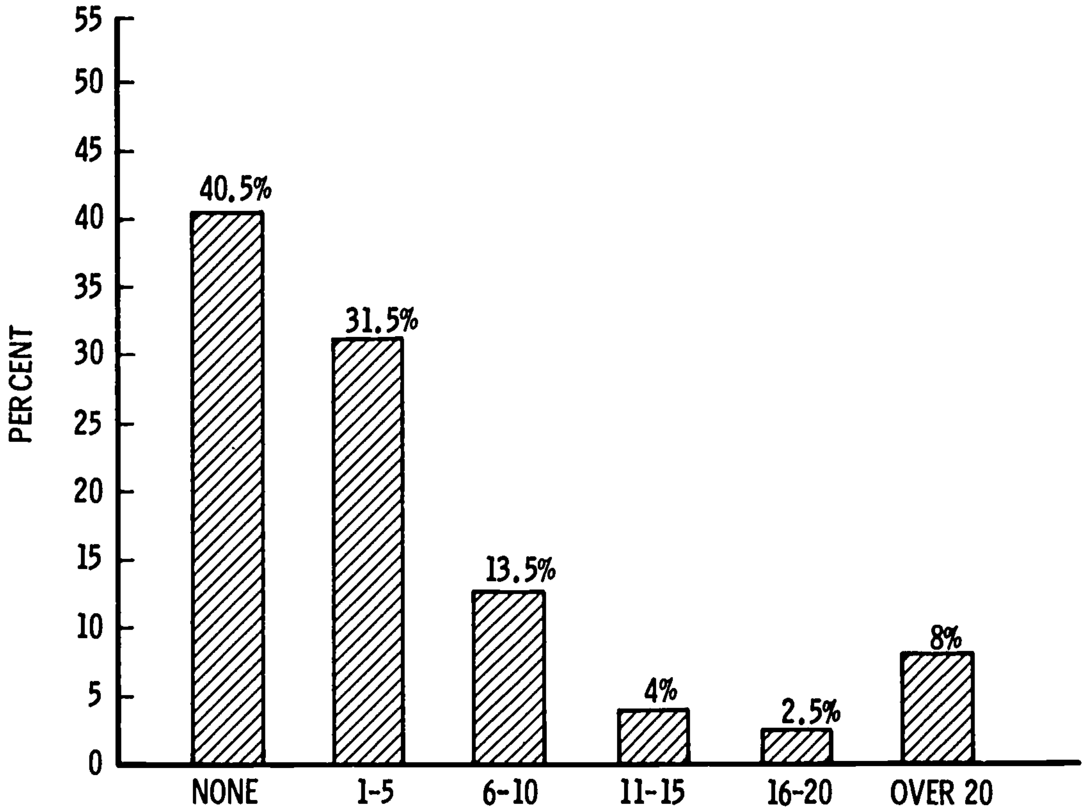


FIGURE 23

USER'S HIGHEST DEGREE

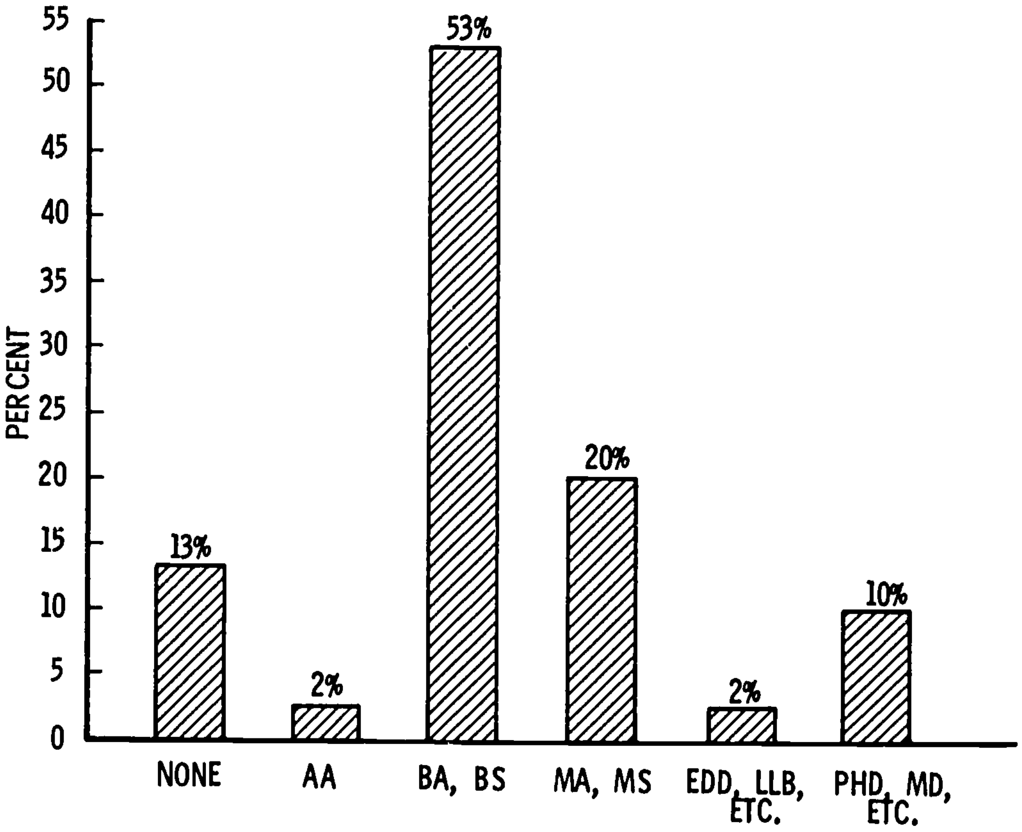


FIGURE 24

FLOW PROCESS FROM AN INPUT/OUTPUT POINT OF VIEW

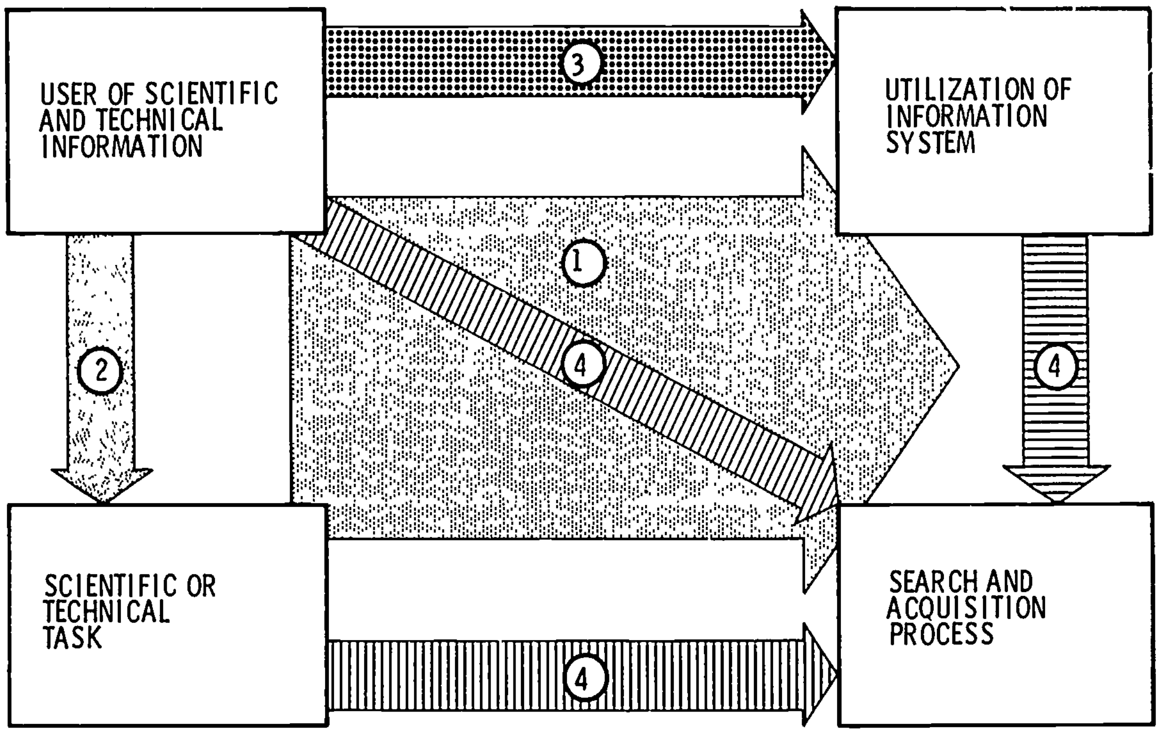
For design and analysis of the flow process, it is meaningful to consider the flow process from an input/output point of view. Input represents "tendency to influence," output represents "tendency to be influenced," and an arrow represents "the tendency of influence from input to output."

The components of the flow process are USER, TASK, UTILIZATION, and SEARCH AND ACQUISITION. For the flow process in general, USER and TASK act as input components; and UTILIZATION and SEARCH AND ACQUISITION act as output components (Arrow 1 in Figure 25). The other input/output relations among components of the flow process have the following:

- USER as input component, and TASK as output component (Arrow 2 in Figure 25).
- USER as input component, and UTILIZATION as output component (Arrow 3 in Figure 25).

M-43387

INPUT-OUTPUT RELATIONS AMONG COMPONENTS OF FLOW PROCESS



*THE ARROWS POINT FROM INPUT (TENDING TO INFLUENCE) TO OUTPUT (TENDING TO BE INFLUENCED).

FIGURE 25

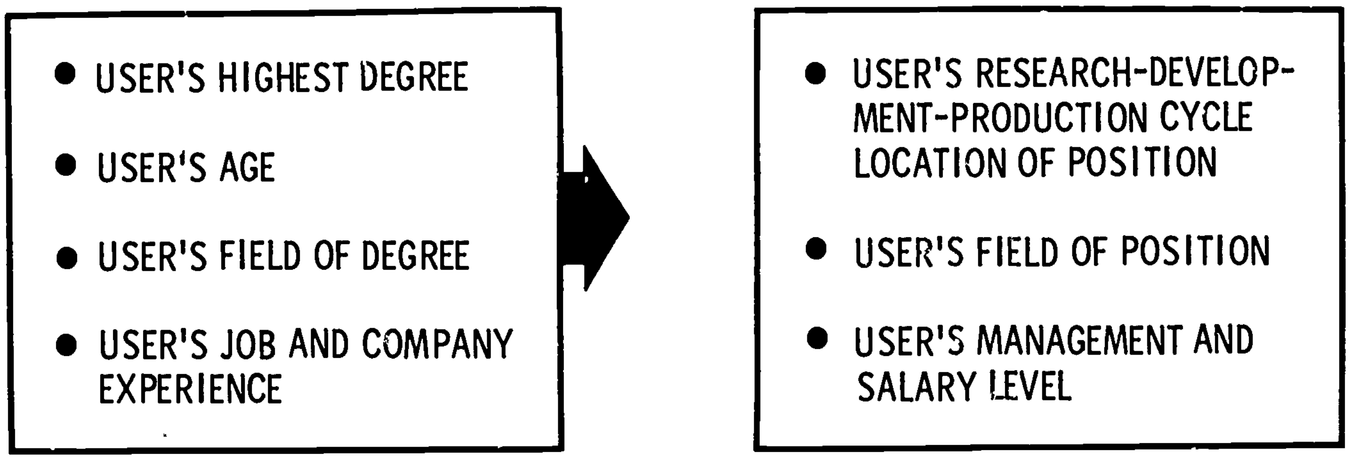
- USER, TASK, and UTILIZATION as input components, and SEARCH AND ACQUISITION as output component (Arrows marked 4 in Figure 25).

Within each component, there are input factors and output factors. Factor represents "combination of related questions." Figures 26 through 30 present input and output factors for USER, TASK, UTILIZATION, SEARCH AND ACQUISITION, and the flow process, respectively. In these figures, input factors are ranked in order of their overall contribution to the relationships within the stated component(s).

One must realize, however, that the statistical techniques of the analysis can merely characterize a relation. They cannot imply that a relation is cause-and-effect, for this can only be determined by a thorough understanding of the flow process.

M-43420

USER INPUT AND OUTPUT FACTORS



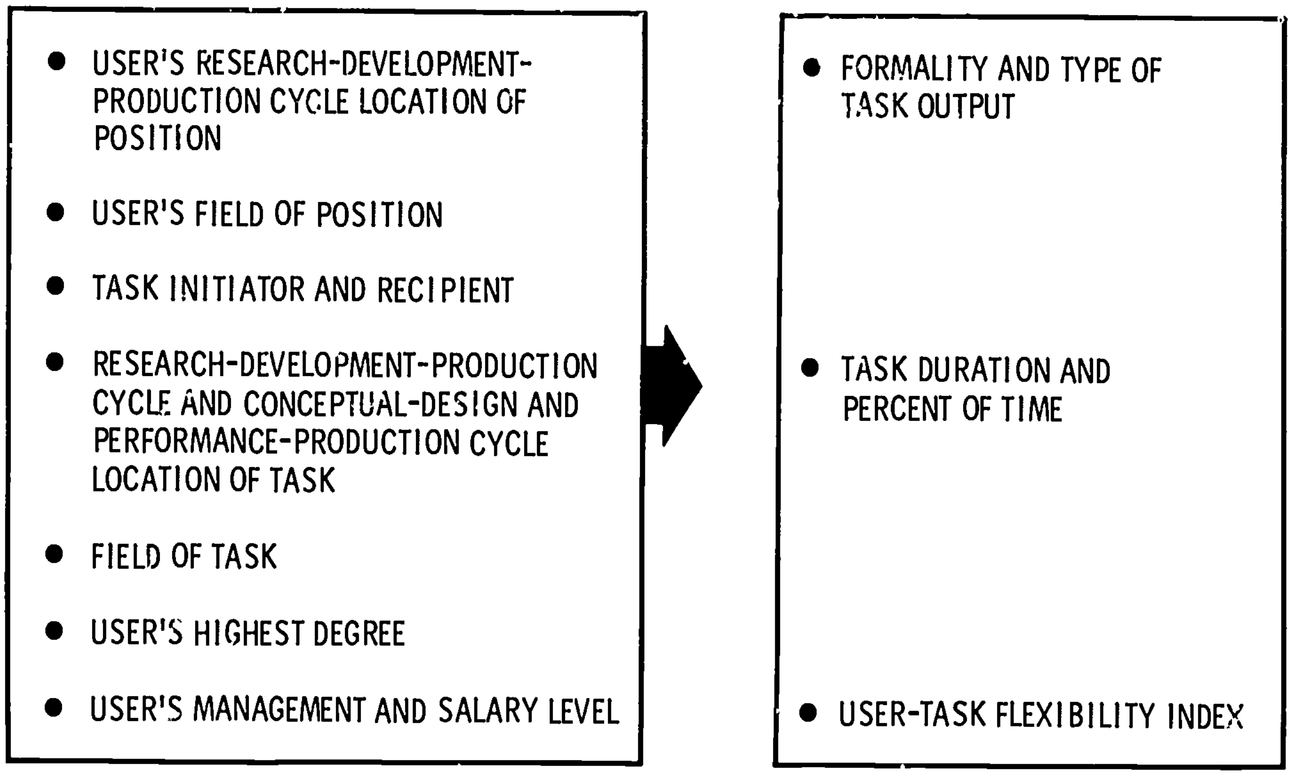
- USER'S HIGHEST DEGREE
- USER'S AGE
- USER'S FIELD OF DEGREE
- USER'S JOB AND COMPANY EXPERIENCE

- USER'S RESEARCH-DEVELOPMENT-PRODUCTION CYCLE LOCATION OF POSITION
- USER'S FIELD OF POSITION
- USER'S MANAGEMENT AND SALARY LEVEL

FIGURE 26

TASK INPUT AND OUTPUT FACTORS

M-43421



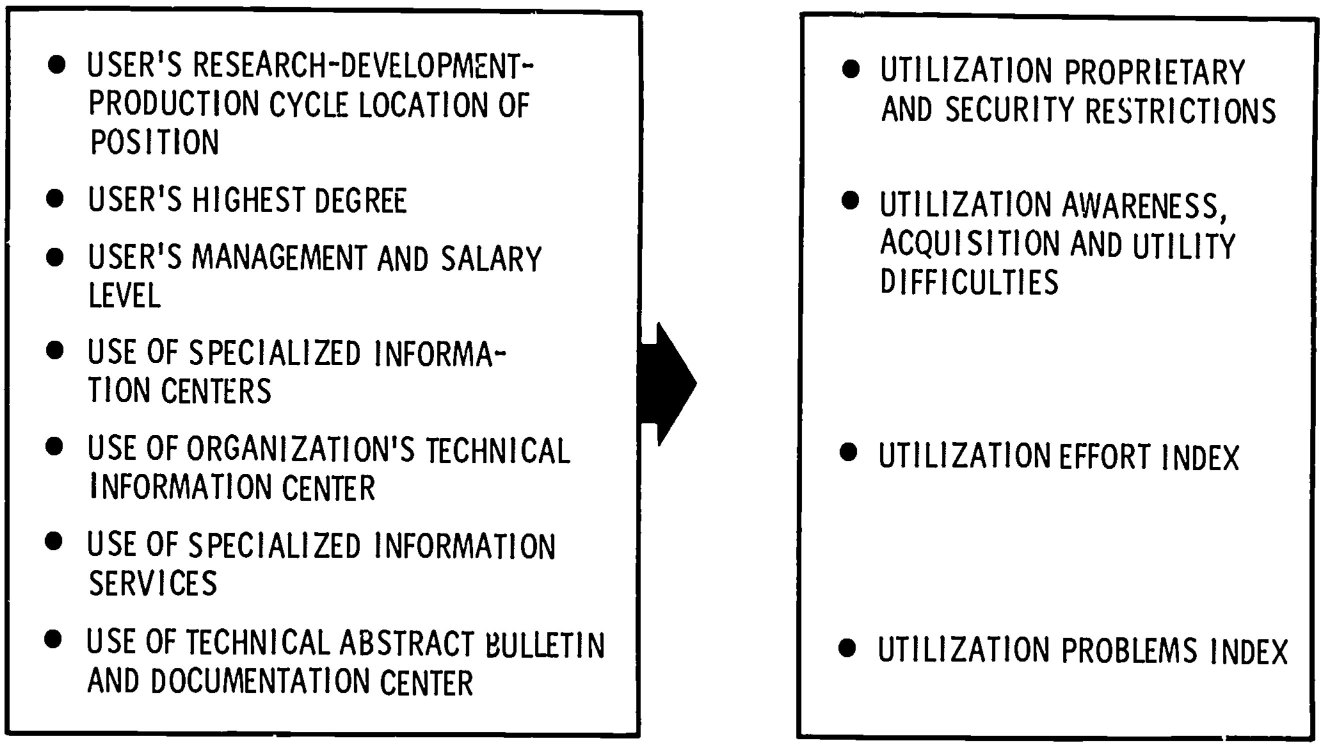
- USER'S RESEARCH-DEVELOPMENT-PRODUCTION CYCLE LOCATION OF POSITION
- USER'S FIELD OF POSITION
- TASK INITIATOR AND RECIPIENT
- RESEARCH-DEVELOPMENT-PRODUCTION CYCLE AND CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF TASK
- FIELD OF TASK
- USER'S HIGHEST DEGREE
- USER'S MANAGEMENT AND SALARY LEVEL

- FORMALITY AND TYPE OF TASK OUTPUT
- TASK DURATION AND PERCENT OF TIME
- USER-TASK FLEXIBILITY INDEX

FIGURE 27

M-43418

UTILIZATION INPUT AND OUTPUT FACTORS



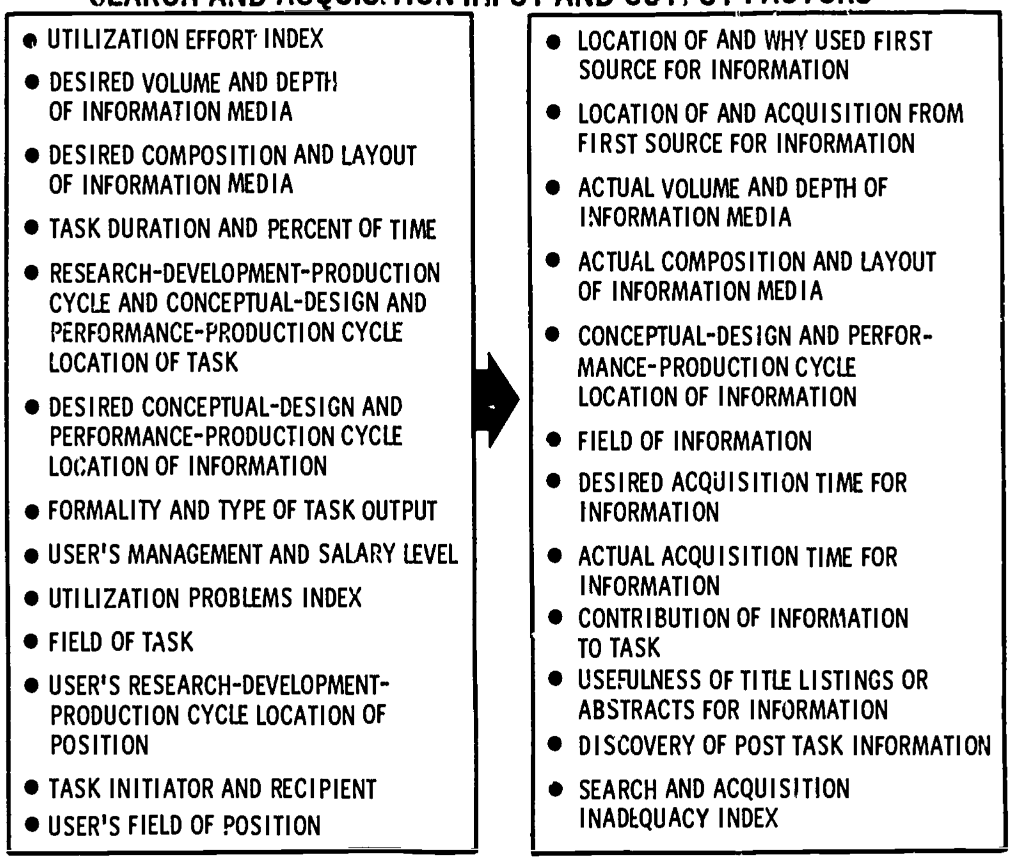
- USER'S RESEARCH-DEVELOPMENT-PRODUCTION CYCLE LOCATION OF POSITION
- USER'S HIGHEST DEGREE
- USER'S MANAGEMENT AND SALARY LEVEL
- USE OF SPECIALIZED INFORMATION CENTERS
- USE OF ORGANIZATION'S TECHNICAL INFORMATION CENTER
- USE OF SPECIALIZED INFORMATION SERVICES
- USE OF TECHNICAL ABSTRACT BULLETIN AND DOCUMENTATION CENTER

- UTILIZATION PROPRIETARY AND SECURITY RESTRICTIONS
- UTILIZATION AWARENESS, ACQUISITION AND UTILITY DIFFICULTIES
- UTILIZATION EFFORT INDEX
- UTILIZATION PROBLEMS INDEX

FIGURE 28

SEARCH AND ACQUISITION INPUT AND OUTPUT FACTORS

M-43419



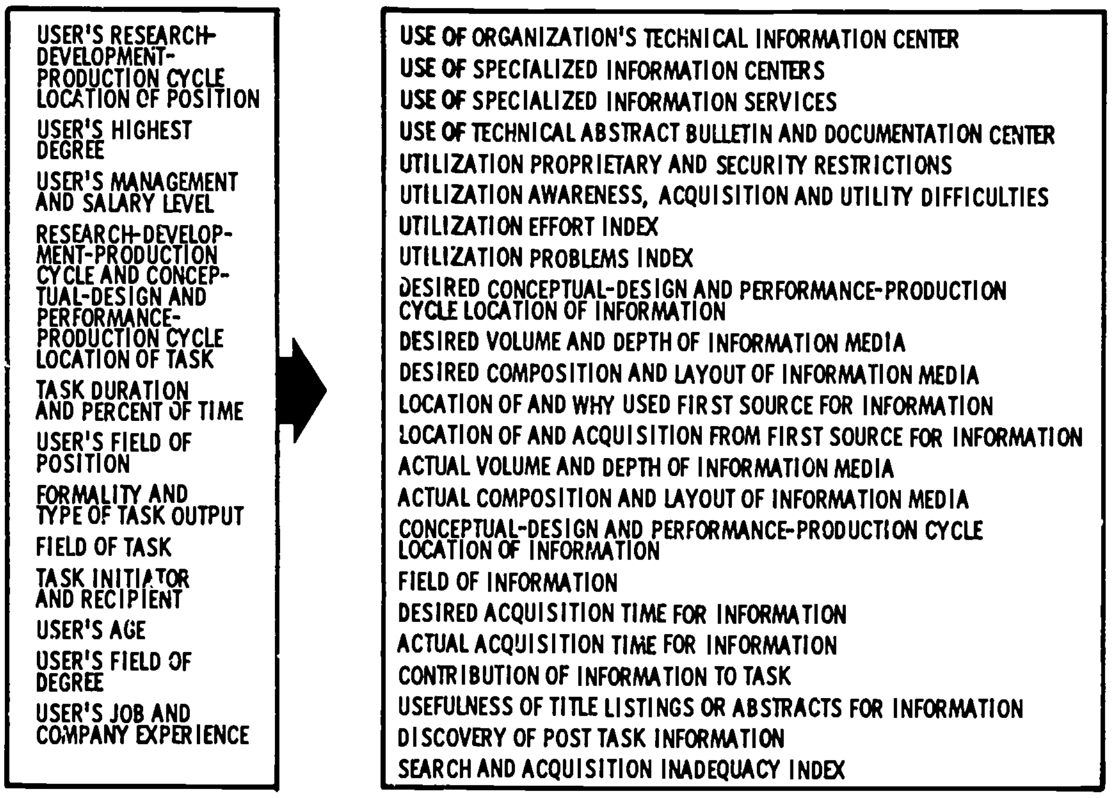
- UTILIZATION EFFORT INDEX
- DESIRED VOLUME AND DEPTH OF INFORMATION MEDIA
- DESIRED COMPOSITION AND LAYOUT OF INFORMATION MEDIA
- TASK DURATION AND PERCENT OF TIME
- RESEARCH-DEVELOPMENT-PRODUCTION CYCLE AND CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF TASK
- DESIRED CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF INFORMATION
- FORMALITY AND TYPE OF TASK OUTPUT
- USER'S MANAGEMENT AND SALARY LEVEL
- UTILIZATION PROBLEMS INDEX
- FIELD OF TASK
- USER'S RESEARCH-DEVELOPMENT-PRODUCTION CYCLE LOCATION OF POSITION
- TASK INITIATOR AND RECIPIENT
- USER'S FIELD OF POSITION

- LOCATION OF AND WHY USED FIRST SOURCE FOR INFORMATION
- LOCATION OF AND ACQUISITION FROM FIRST SOURCE FOR INFORMATION
- ACTUAL VOLUME AND DEPTH OF INFORMATION MEDIA
- ACTUAL COMPOSITION AND LAYOUT OF INFORMATION MEDIA
- CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF INFORMATION
- FIELD OF INFORMATION
- DESIRED ACQUISITION TIME FOR INFORMATION
- ACTUAL ACQUISITION TIME FOR INFORMATION
- CONTRIBUTION OF INFORMATION TO TASK
- USEFULNESS OF TITLE LISTINGS OR ABSTRACTS FOR INFORMATION
- DISCOVERY OF POST TASK INFORMATION
- SEARCH AND ACQUISITION INADEQUACY INDEX

FIGURE 29

FLOW PROCESS INPUT AND OUTPUT FACTORS

M-40659



USER'S RESEARCH-DEVELOPMENT-PRODUCTION CYCLE LOCATION OF POSITION
USER'S HIGHEST DEGREE
USER'S MANAGEMENT AND SALARY LEVEL
RESEARCH-DEVELOPMENT-PRODUCTION CYCLE AND CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF TASK
TASK DURATION AND PERCENT OF TIME
USER'S FIELD OF POSITION
FORMALITY AND TYPE OF TASK OUTPUT
FIELD OF TASK
TASK INITIATOR AND RECIPIENT
USER'S AGE
USER'S FIELD OF DEGREE
USER'S JOB AND COMPANY EXPERIENCE

USE OF ORGANIZATION'S TECHNICAL INFORMATION CENTER
USE OF SPECIALIZED INFORMATION CENTERS
USE OF SPECIALIZED INFORMATION SERVICES
USE OF TECHNICAL ABSTRACT BULLETIN AND DOCUMENTATION CENTER
UTILIZATION PROPRIETARY AND SECURITY RESTRICTIONS
UTILIZATION AWARENESS, ACQUISITION AND UTILITY DIFFICULTIES
UTILIZATION EFFORT INDEX
UTILIZATION PROBLEMS INDEX
DESIRED CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF INFORMATION
DESIRED VOLUME AND DEPTH OF INFORMATION MEDIA
DESIRED COMPOSITION AND LAYOUT OF INFORMATION MEDIA
LOCATION OF AND WHY USED FIRST SOURCE FOR INFORMATION
LOCATION OF AND ACQUISITION FROM FIRST SOURCE FOR INFORMATION
ACTUAL VOLUME AND DEPTH OF INFORMATION MEDIA
ACTUAL COMPOSITION AND LAYOUT OF INFORMATION MEDIA
CONCEPTUAL-DESIGN AND PERFORMANCE-PRODUCTION CYCLE LOCATION OF INFORMATION
FIELD OF INFORMATION
DESIRED ACQUISITION TIME FOR INFORMATION
ACTUAL ACQUISITION TIME FOR INFORMATION
CONTRIBUTION OF INFORMATION TO TASK
USEFULNESS OF TITLE LISTINGS OR ABSTRACTS FOR INFORMATION
DISCOVERY OF POST TASK INFORMATION
SEARCH AND ACQUISITION INADEQUACY INDEX

FIGURE 30

ANALYSIS OF FLOW-PROCESS DATA

OVERVIEW OF THE ANALYSIS

The data consist of 1,500 transcribed interviews, each containing answers to 55 questions having allowable responses which are qualitative, and 8 questions having allowable responses which are quantitative. Figures 4 through 24 summarize the responses to 21 significant questions, and Figures 26 through 30 summarize the subject of all but eight less-important questions. For a complete listing of questions and their responses, Appendixes 5 and 6 to Volume II of Reference 1 should be consulted.

Detailed information describing small portions of the flow process is provided by one-way and two-way frequency distributions (for example, Figures 4 through 24). A one-way frequency distribution is the distribution of the percentage of answers to a question that corresponds to each (allowable) question response, and a two-way frequency distribution is the distribution of the percentage of answers to a pair of questions that corresponds to each pair of (allowable) question responses (see Table 1).

In addition, the relationship analysis cycle yields general information describing both small and large portions of the flow process (for example, Figures 26 through 30). In this cycle, qualitative question responses are transformed into numerical form, a process model for linear relationships among questions is constructed and estimated, and numerical relationship results are transformed back to qualitative form (see Figure 31).

Transformation of qualitative question responses into numerical form is accomplished by arranging the responses into an informative detailed (local) structure, and then associating a meaningful number with each response. The construction of a process model for linear relationships among questions is accomplished by arranging the questions into an informative general (global) structure, and then specifying the general form of meaningful linear relationships among questions. Next unspecified constants, in the general form of these relationships in the process model, are estimated from the

Table 1
ONE-WAY AND TWO-WAY FREQUENCY DISTRIBUTIONS

One-Way Frequency Distribution

Question 22: Desired Volume of Information Media

Response	Frequency (%)
All from recall	7
One report or document	30
A sampling of the reports and documents available	22
All reports and documents that could be found pertinent to the question	41

Two-Way Frequency Distribution

Question 22: Desired Volume of Information Media / Question 25: Desired Depth of Information Media	A once over lightly	A specific answer	A detailed analysis
All from recall	0%	5%	2%
One report or document	2%	18%	10%
A sampling of the reports and documents available	3%	10%	9%
All reports and documents that could be found pertinent to the question	2%	23%	16%

data by employing the statistical technique called stepwise regression analysis. Finally, numerical relationship results are transformed back to qualitative form by ranking questions in the order of their contribution to each relationship, and then in the order of their overall contribution to the relationships in each component of the flow process and the flow process itself.

RELATIONSHIP ANALYSIS CYCLE

M-43392

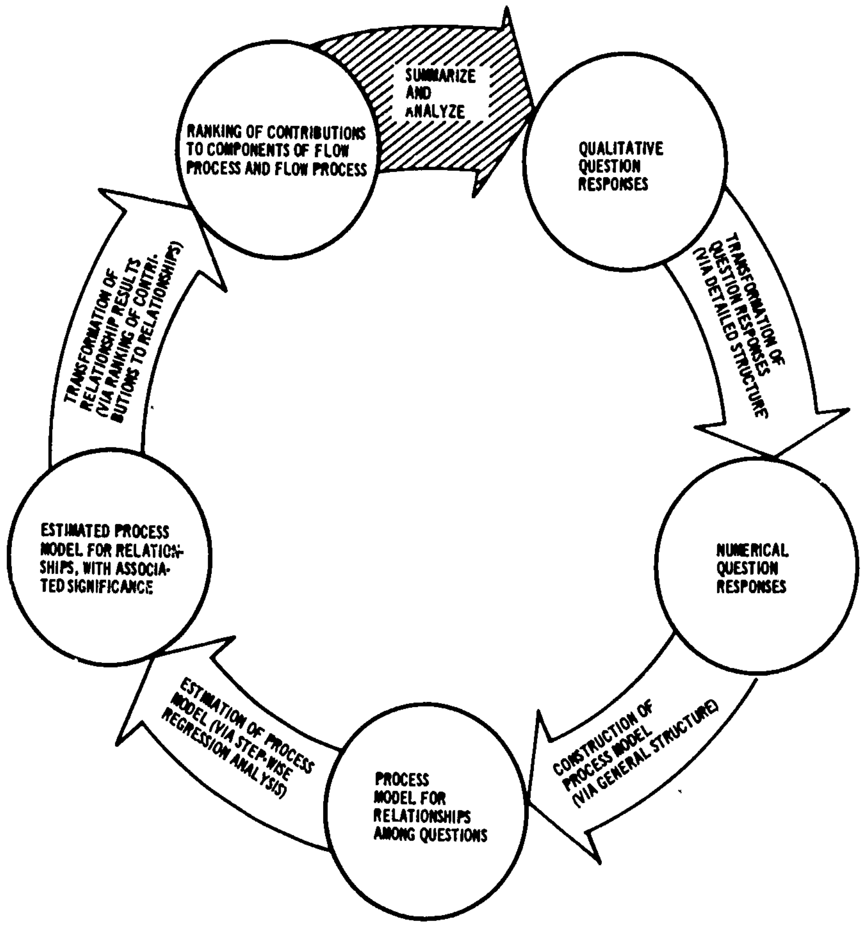


FIGURE 31

The relationship analysis cycle is believed to be novel in the field of information science. Its employment and testing in this investigation have yielded results that are encouraging, and implications for the future that are provocative.

REQUIREMENTS OF THE ANALYSIS

An analysis ought to operate upon the data in such a way, and to such an extent, that the analytical requirements are met. What an analysis ought to accomplish is determined by both the data and the analytical requirements. The weaker the data or the stronger the analytical requirements, the stronger should an analysis be.

An analysis should provide a bridge between the data, and meaningful conclusions and recommendations. It should bring the information content of the data into focus. It should transform apparent chaos into orderly findings, which readily lead to conclusions and recommendations.

To achieve this, an analysis must organize, summarize, and interpret the data. The methods of summarization employed by an analysis ought to be sufficient to bring both the detailed and general information content of the data into focus. Higher-order effects are indicated by detailed information, whereas lower-order effects are indicated by general information.

Detailed information is relatively close to the surface of the data, and requires a relatively small amount of summarization to be brought into focus. The more the detail, the less the summarization required. On the other hand, general information is buried relatively far beneath the surface of the data, and requires a relatively large amount of summarization to be brought into focus. The more the generality, the more the summarization required.

By its very nature, detailed information describing only small portions of the flow process may be comprehended at once. However, general information describing either small or large portions of the flow process may be comprehended at once. That is, only small amounts of great detail may be simultaneously digested; whereas, either small or large amounts of little detail may be simultaneously digested.

Consequently, the analysis first should summarize the data until their detailed information content, describing only small portions of the flow process at once, is brought into comprehensible focus. It then should continue to summarize the data until their general information content, describing both small and large portions of the flow process at once, is brought into comprehensible focus. Otherwise, any interested person will be forced to accept only the data's detailed information content; or to himself perform additional summarization, so that the data's general information content is brought into comprehensible focus.

STATISTICAL CONCEPTS

To aid the translation of these general analysis requirements into specific analysis objectives, pertinent statistical concepts are briefly introduced and discussed in the following paragraphs.

Frequency Distributions

One-way and two-way frequency distributions have been defined above. Higher-order frequency distributions are similarly defined. Frequency distributions necessitate the simplest operation upon the data, and contain a wealth of detailed information regarding variation in the data; however, they provide the minimal amount of summarization.

The usual procedure for summarizing a one-way frequency distribution is to combine some question responses, and/or to obtain measures of the one-way frequency distribution's location and spread. The distribution's location may be measured by its mode if the qualitative question responses are not arranged into an order, by its median if the qualitative question responses are ordered, and by its mean if the question responses are quantitative. Measures of the distribution's spread are its range if the qualitative question responses are ordered, and its standard deviation if the question responses are quantitative. More definitive information is obtained by this summarization when the qualitative question responses are ordered, and even more definitive information is obtained when the question responses are quantitative.

Summarization of two-way frequency distributions is both more necessary and more difficult to perform. The first step is to combine some responses for each question, and/or to obtain measures of the location and dispersion of each question's one-way frequency distribution. Then a measure of the association or interaction between the two questions is sought. If the qualitative responses to each question are ordered, the interaction between the two questions may be measured by the rank correlation (coefficient); and if each question's responses are quantitative, the interaction may be measured by the correlation (coefficient). An indirect approach to measuring this interaction when the question responses are qualitative is provided by Chi-square, which indicates the departure of the questions from being independent or not related.

Computation of the rank correlation automatically associates the numbers 1, 2, ..., n with the first, second, ..., nth responses to these questions. On the other hand, the computation of the correlation depends upon the quantitative responses to each question, or the numbers associated with the responses to each question.

As for one-way frequency distributions, more definitive information is obtained by this summarization when the qualitative question responses are ordered, and even more definitive information is obtained when the question responses are quantitative. Arrangement of question responses into an informative order is called development of a detailed structure, while association of a meaningful number with each response is called definition of a numerical description for the detailed structure. The development of a detailed structure, followed by the definition of a numerical description for the detailed structure, transforms the qualitative question responses into numerical form.

Higher-order frequency distributions become increasingly harder to generate, depict, and comprehend. Consequently, their summarization becomes both increasingly more necessary and more difficult. They are of relatively little analytical use, except in rare instances.

Relationships

For questions with quantitative responses, a relationship among questions is a mathematical expression of the variation in one question as a function of the variations in the other questions. It is frequently both convenient and sufficiently accurate--for example, during exploratory research such as this investigation--to represent a relationship by a linear one, which depicts the variation in one question as a linear combination of the variations in the other questions. The general form of a linear relationship is written:

$$Y = \beta_o + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon,$$

with Y being one question, X_1, X_2, ..., X_p being the other questions, β_o, β_1, ..., β_p being the unspecified constants or coefficients, and ϵ being the residual. The correlation, in reality, measures the degree of linearity for the interaction between the two questions; or the closeness of the two questions to being adequately represented by a linear relationship,

$$Y = \beta_o + \beta_1 X + \epsilon,$$

between one question Y and the other question X.

The analysis of relationships requires not only quantitative data, but also specification of the general form for meaningful linear relationships among questions. In turn, the specification of the general form for these relationships requires that the questions be arranged into an informative order. Arrangement of questions into an informative order is called development of a general structure. The development of a general structure, followed by the specification of the general form for meaningful linear relationships among questions in the general structure, accomplishes the construction of a process model. Consequently, the analysis of relationships depends upon both the transformation of qualitative question responses into numerical form, and the construction of a process model for linear relationships among questions.

Comparison

Two-way frequency distributions are easy to generate, and their concept is easy to understand. They summarize relatively little, however, and their information content is difficult to comprehend without additional summarization. On the other hand, relationships are not as easy to obtain and to understand in concept; but they do summarize a great deal, and their information content is easy to comprehend without additional summarization.

Let the responses to one question be associated with the X-axis, and the responses to another question be associated with the Y-axis. Then a two-way frequency distribution may be viewed as a geometric representation for the distribution of the answers to the two questions, in which each percentage gives the proportion of answer-pairs which are associated with the corresponding response-pair point. In addition a linear relationship,

$$Y = \beta_o + \beta_1 X + \epsilon,$$

may be viewed as a natural summarization of the two-way frequency distribution. It replaces the geometric representation of the distribution with a line through it, and with an analytic representation of the distribution and the line. The more the distribution tends to cluster closely around a line, the more appropriate is a linear relationship; and the higher is the correlation between the two questions. Figure 32 presents an example, using the two-way frequency distribution from Table 1 (for which a linear relationship is not very appropriate).

Although two-way frequency distributions may be summarized to present some general information regarding the interaction of the two questions, they are limited to describing only small portions of the flow process at once.

REPRESENTATION OF A TWO-WAY FREQUENCY DISTRIBUTION M-43398

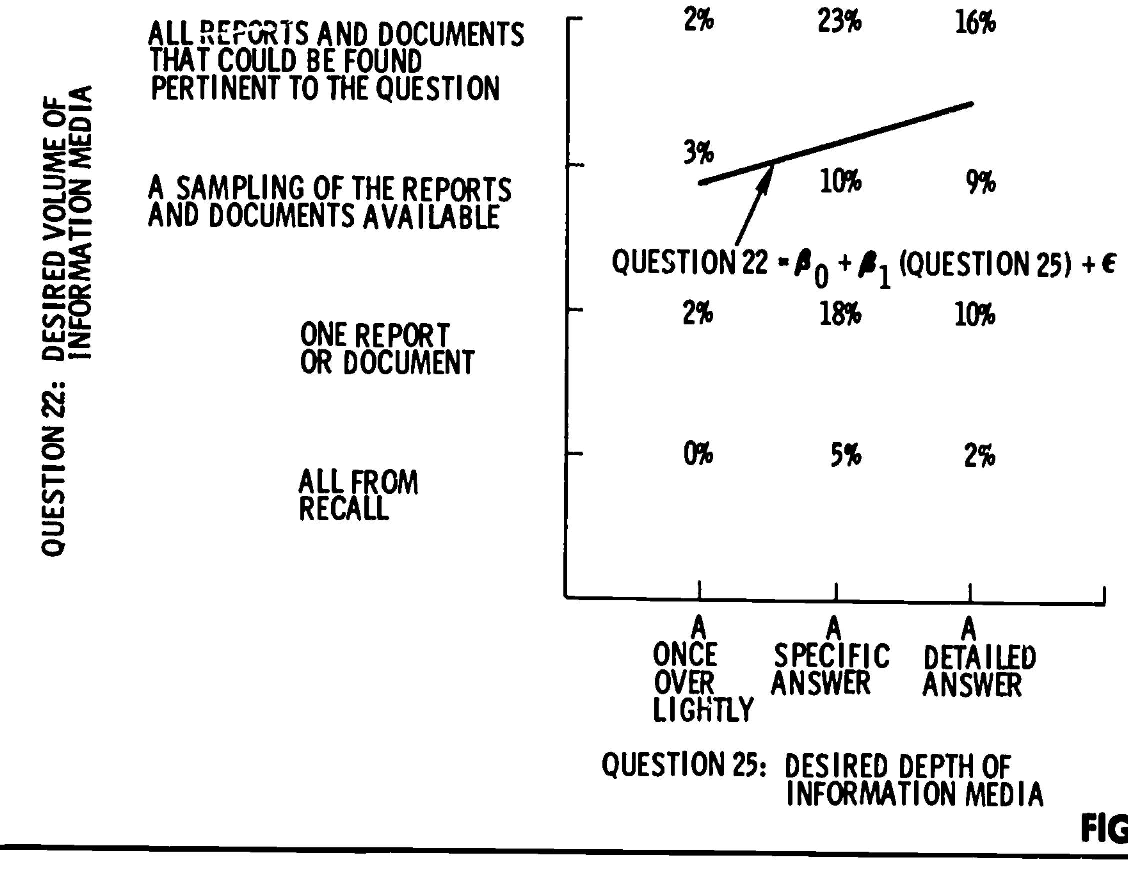


FIGURE 32

Relationships, however, are not limited at all, and may be used to describe either small or large portions of the flow process. In addition, relationships sufficiently summarize the data by an analytic representation, to bring its general information content into focus. They provide a natural summarization of not only two-way, but also higher-order, frequency distributions.

For a detailed analysis of the data, two-way frequency distributions are necessary. The analysis of relationships is required for a general analysis of the data. In addition, it is useful for such purposes as the planning and analysis of additional investigations, and the program for analysis and optimization of the process. Relationships provide a global view of large portions of the flow process, which also enables many small portions of the process to be examined simultaneously and their relative importance evaluated.

The analysis of relationships has many advantages over the generation of two-way frequency distributions. One must, however, realize that these advantages have to be paid for by the transformation of qualitative question responses into numerical form, and the construction of a process model for relationships among questions. In addition, the relationship results should be analyzed and interpreted by techniques which are relatively insensitive to changes in the transformation.

OBJECTIVES OF THE ANALYSIS

The summarization of data, to bring into focus their detailed information content describing small portions of the flow process, could be achieved by means of one-way and two-way frequency distributions for single questions and pairs of questions. An analysis of relationships among questions could accomplish the additional summarization of data, to bring into focus their general information content describing both small and large portions of the flow process.

Qualitative question responses, however, pose a problem. Although frequency distributions may be generated for qualitative question responses, they provide much more definitive information for quantitative question responses. The analysis of relationships, as noted above, requires both the transformation of qualitative question responses into numerical form, and the construction of a process model for relationships among questions.

Thus, the objectives of the analysis are to:

- Generate one-way and two-way frequency distributions for single questions and pairs of questions.
- Transform qualitative question responses into numerical form.
- Construct and estimate a process model for linear relationships among questions.
- Analyze and interpret the frequency distribution and relationship results, to provide meaningful conclusions and recommendations which are relatively insensitive to changes in the transformation.

FREQUENCY DISTRIBUTIONS

A one-way frequency distribution has been generated for 59 of the 63 questions. The remaining four questions were narrative and were not categorized.

From the large number of two-way frequency distributions that could have been generated, 196 were selected for compilation. These were supplemented by the analysis of relationships and the complete correlation matrix, which was a by-product of that analysis.

One-way frequency distributions were transcribed from the marginal distributions of the appropriate two-way frequency distributions. The computer program employed to generate two-way frequency distributions was BMD 08D (see Reference 3).

TRANSFORMATION OF QUESTION RESPONSES

As noted above, the transformation of qualitative question responses into numerical form is accomplished by the development of a detailed structure, and then the definition of a numerical description for that detailed structure.

Development of a Detailed Structure

A detailed structure for question responses is developed to serve as the basis for the transformation of these responses. In addition, the detailed structure brings the local aspects of the flow process into focus, and provides a foundation for a general structure. This detailed structure is formed by the informative arrangement of question responses.

The first step is to specify the primary unifying characteristic of each question's responses. This response characteristic should be determined from not only the responses themselves, but also the question's intent.

The next step is to collect into groups those question responses which are related by the response characteristic. According to this characteristic, an ordering is then arranged for groups and, to the extent feasible, for responses within groups. All responses to a question may be arranged into one ordering, if all responses within each group may be arranged into an ordering. According to the response characteristic, a response or a group of responses is more similar to responses or groups of responses which are closer to it in the arrangement, than to those farther away.

Depending upon the implications of the response characteristic, there are three types of detailed structure:

- Visible structure, explicitly implied by the response characteristic.
- Partially visible structure, implicitly implied by the response characteristic.
- Invisible structure, not implied at all by the response characteristic.

A visible structure is obvious, and possesses no flexibility. A partially visible structure is apparent, but possesses some flexibility. An invisible structure must be inferred, and possesses considerable flexibility. The position of responses in the arrangement is meaningful in a visible structure, and indicative in a partially visible structure, but only descriptive in an invisible structure.

Examples of visible, partially visible, and invisible structures are given in Tables 2 through 4, respectively. For the tables, the Arabic numerals in parentheses indicate the ordering in the interview, while the Roman numerals indicate the ordering in the detailed structure. The numerical description scale is included in the tables.

Definition of a Numerical Description

When the detailed structure is developed, its numerical description is appropriate. By associating a number with each ordered question response, the numerical description provides a more exact differentiation among responses, and enables estimation of the process model which is constructed for linear relationships among questions. The numerical description also represents the data in a form to which a large variety of numerical techniques may be applied.

According to the response characteristic, the base point (zero) for a numerical scale is selected. With each response, there is associated a numerical value corresponding to its relative distance from the base point.

Table 2

TRANSFORMATION OF QUESTION RESPONSES: VISIBLE STRUCTURE

Question 13: Desired Acquisition Time for Information
Response Characteristic: Days

		Informative Order	Scale
I	(01)	From recall	0.00
II	(02)	Less than 1 day	0.01
III	(03)	1 to 7 days	0.05
IV	(04)	8 to 30 days	0.20
V	(05)	31 to 90 days	0.60
VI	(06)	Over 90 days	1.00

Table 3

TRANSFORMATION OF QUESTION RESPONSES: PARTIALLY VISIBLE STRUCTURE

Question 14: First Source Contacted for Information
Response Characteristic: Distance from User

		Informative Order	Scale
I	(01)	Received with task assignment	0.00
II	(04)	Recalled it	0.05
III	(09)	Searched own collection	0.10
IV	(19)	Respondent's own action	0.15
V	(03)	Assigned subordinate to get it	0.20
VI	(05)	Asked a colleague	0.25
VII	(02)	Asked my supervisor	0.30
VIII	(08)	Requested search of departmental files	0.35
IX	(06)	Asked an internal consultant	0.45
X	(10)	Searched organization's technical information center.	* 0.50
X	(07)	Requested technical information center search.	
XI	(15)	Requested data from vendor, manufacturer, or supplier.	* 0.60
XI	(14)	Searched vendor, manufacturer, or supplier sources	
XII	(11)	Searched outside technical information center	0.70
XIII	(18)	Asked an external consultant or expert	0.80
XIV	(13)	Requested search of Government Information Center	* 0.90
XIV	(12)	Searched Government Information Center	
XV	(17)	Asked customer	1.00

---

*No distinction is made between the two responses in this group of related responses.

Table 4

TRANSFORMATION OF QUESTION RESPONSES: INVISIBLE STRUCTURE

Question 27: Desired Layout of Information Media
Response Characteristic: Formality

		Informative Order	Scale
I	(14)	Recall	0.00
II	(13)	Telephone conversation	0.06
III	(11)	Group discussion	0.12
IV	(04)	Photographs	0.19
V	(03)	Graphics (diagrams, drawings, schematics, flow charts, graphs, maps)	0.25
VI	(02)	Tables or lists	0.31
VII	(01)	Narrative text	0.37
VIII	(18)	Narrative text, and tables or lists	0.44
IX	(09)	Graphics and lists	0.50
X	(08)	Photographs and text	0.56
XI	(07)	Graphics and text	0.63
XII	(16)	Graphics, text, and oral	0.69
XIII	(17)	Graphics, text, oral, and recall	0.75
XIV	(12)	Informal briefing, with chalk or pencil drawings	0.82
XV	(05)	Microfilm or microfiche	0.88
XVI	(06)	Slides or motion pictures	0.94
XVII	(10)	Formal briefing or lecture	1.00

Except for two questions, -1, 0, or a positive integer is associated with each question response. The two exceptional questions have multiples of one-half associated with some responses for convenience. When it is meaningful to consider the response to be null, 0 is used; and when it is meaningful to consider the response as opposite in direction to the remaining responses, -1 is used. Variable spacing between the associated numbers indicates that the responses exhibit variable similarity, or distance from each other, according to the response characteristic. The same number is associated with two responses to a question if--and only if--the two responses are in the same group of related responses, and the responses within that group are not arranged into an ordering (that is, are considered to be the same distance from the base point).

The association of a number with each question response associates a scale of possible numerical values with the question. Then all numerical values in the scale are divided by the largest one, so that the scale is normalized to between -1 and 1--and usually between 0 and 1.

The value of the numerical description is meaningful for responses in a visible structure, and indicative for responses in a partially visible structure, but only descriptive for responses in an invisible structure. Examples are again provided by Tables 2 through 4.

A detailed structure suggests its own numerical description when the question responses have been properly arranged. For a more refined relationship analysis, a numerical description could be altered to improve the linearity of important relationships which involve the corresponding question.

CONSTRUCTION OF A PROCESS MODEL

Development of a general structure, and specification of the general form for meaningful linear relationships among questions in the general structure, accomplish the construction of a process model.

Development of a General Structure

A general structure now is developed to serve as the basis for the construction of a process model for linear relationships among questions, and to bring the global aspects of the flow process into focus. This general structure is formed by the informative arrangement of questions.

The first step is to identify the components of the flow process as USER, TASK, UTILIZATION, and SEARCH AND ACQUISITION. The next step is to form groups of related questions within components. Then an ordering is arranged for components, groups within components, and questions within groups. To the extent feasible, the arrangement should possess the desirable characteristic that a question tends to influence only those questions which follow it.

It is frequently both convenient and sufficiently accurate--for example, during exploratory research such as this investigation--to combine groups of related questions. The combination of related questions summarizes the general structure, and simplifies the specification and estimation of meaningful linear relationships among questions.

Two of the simplest types of combinations are averages and products. They keep the combination scales normalized to between -1 and 1. Except for the four cases in which a product of two questions is employed, all of the combinations are averages of two questions.

A special USER-TASK flexibility index F summarizes the flexibility exhibited by the difference between the user's position within the research-development-production cycle and that of his task; and the difference between the user's field of position, and that of his task. To summarize the effort expended by the user in his utilization of the information system and the problems encountered by him in this utilization, the respective special indexes, E for UTILIZATION effort and P for UTILIZATION problems, are introduced. The inadequacy of the search and acquisition process, for information used in task performance, is summarized by the special index I for SEARCH AND ACQUISITION inadequacy. The scales for F, E, P, and I are also normalized to between -1 and 1.

An example is provided by Table 5, which also includes linear relationships. In this table, Q denotes Question; and β_0, β_1, β_2, ..., β_6 symbolize general unspecified constants in the relationships. For simplicity, the same symbols, β_0, β_1, β_2, ..., β_6, are used in each relationship; although they are not meant to denote the same constants.

Table 5

CONSTRUCTION OF A PROCESS MODEL: USER COMPONENT*

1. User's age: Q48
2. User's education
   - A. User's highest degrees: Q50A $= \beta_0 + \beta_1$ (Q48)
   - B. User's field of degree: Q50C $= \beta_0 + \beta_1$ (Q48)
3. User's experience

   Combination: 1/2 (Q51 + Q52) $= \beta_0 + \beta_1$ (Q48)
   - A. User's job experience: Q51
   - B. User's company experience: Q52
4. User's position
   - A. User's position within the research-development-production cycle

     Q55 $= \beta_0 + \beta_1$ (Q48) $+ \beta_2$ (Q50A) $+ \beta_3$ (Q50C) $+ \beta_4$ (1/2 (Q51 + Q52))
   - B. User's field of position

     Q56 $= \beta_0 + \beta_1$ (Q48) $+ \beta_2$ (Q50A) $+ \beta_3$ (Q50C) $+ \beta_4$ (1/2 (Q51 + Q52))
5. User's level

   Combination:

   1/2 (Q49 + Q58) $= \beta_0 + \beta_1$ (Q48) $+ \beta_2$ (Q50A) $+ \beta_3$ (Q50C) $+ \beta_4$ (1/2 (Q51 + Q52)) $+ \beta_5$ (Q55) $+ \beta_6$ (Q56)
   - A. User's salary level: Q58
   - B. Number of personnel managed by user: Q49

---

*Q denotes Question; and $\beta_0, \beta_1, \beta_2, \ldots, \beta_6$ symbolize general unspecified constants in the relationships. For simplicity, the same symbols, $\beta_0, \beta_1, \beta_2, \ldots, \beta_6$, are used for each relationship; although they are not meant to denote the same constants.

A question combination (component) which tends to influence other combinations of questions (components) is called an input factor (component), and a combination of questions (component) which tends to be influenced by other question combinations (components) is called an output factor (component).

The terms, combination of questions and question combination, also are used to cover the degenerate case of a single question--for example, Q56 in Table 5. Arrangement of components and question combinations within components, according to an input/output point of view, facilitates the specification of the general form for meaningful linear relationships among combinations of questions. It also provides insight into the flow process.

When a more refined relationship analysis is desired, the question combinations could be separated, and more special summarizing indexes could perhaps be defined.

Specification of the General Form for Relationships

Once the general structure is developed and groups of related questions are combined, it is appropriate to specify the general form for meaningful linear relationships among combinations of questions in the general structure.

Analysis of the general structure, from an input/output point of view, yields those question combinations which are judged to be potentially related to each combination of questions in the general structure. Only the potentially related question combinations are included in the general form of the linear relationship, for that combination of questions. An example is provided by Table 5.

When the questions have been properly arranged and summarized by combination, a general structure suggests the general form for meaningful relationships. A more refined relationship analysis could specify the general form for additional relationships, particularly those necessitated by the separation of question combinations.

ESTIMATION OF RELATIONSHIPS

The unspecified constants, in the general form of meaningful linear relationships among combinations of questions, are estimated from the numerically transformed question responses by the statistical technique called stepwise regression analysis. Reference 3 presents a complete description of this technique. A brief discussion of only the pertinent aspects of stepwise regression analysis follows.

Stepwise regression analysis estimates the relationship in steps, by entering one question combination at a time. At each step, the question combination which is entered is the one that adds the greatest contribution to the relationship from the previous step. A measure of this contribution is the F to enter of that question combination. The contribution of a question combination to the relationship at each step is measured by its F to remove at that step, and the significance of the relationship at each step is measured by the multiple correlation (coefficient) at that step. Relative significance within a relationship is indicated by the former, while relative significance among relationships is indicated by the latter. In addition, the potential contribution to the relationship at each step, of some question combinations which were not included in the general form of the relationship, is measured by their potential F to enter at that step.

The computer program employed for the stepwise regression analysis is BMD 02R (Reference 3).

TRANSFORMATION OF RELATIONSHIP RESULTS

The stepwise regression computer printouts contain a wealth of numerical detail, concerning relationship results and their significance. To make the conclusions of the relationship analysis relatively insensitive to the transformation of qualitative question responses into numerical form, the numerical relationship results must be transformed back to qualitative form. In addition, summarization of the numerical detail is quite informative.

Both of these requirements are accomplished by a ranking procedure which:

- Ranks question combinations in the order of their contribution to each relationship.

- Then ranks question combinations in the order of their overall contribution to the relationships in each component of the flow process, and the flow process itself.

The former focuses upon a given combination of questions, and observes which question combinations are most significantly related to it; while the latter focuses upon the appropriate collection of combinations of questions, and observes which question combinations are most significantly related to most of them.

Ranking of Contributions to Relationships

An effective step in the stepwise regression analysis, beyond which relatively little is contributed to the relationship, is determined when the F to enter of the question combination entering at that step becomes less than some lower bound. Analysis of the stepwise regression computer printouts indicates that a reasonable value for this lower bound is 6.66 (F level of 0.01). When a question combination is included in the relationship at the effective step, it is said to be related to the given combination of questions. Those question combinations, whose potential F to enter at the effective step is at or above 6.66, are said to be candidates for the relationship.

For each combination of questions in the general structure, the question combinations which are related to it are ranked in the order of their contribution to the relationship. Table 6 contains an example.

Ranking of Contributions to Components and Flow Process

These rankings of contributions may be obtained by properly combining the rankings of contributions for the appropriate collection of relationships. To accomplish this, numerical values must be assigned to the relationship rankings. This return to numerical form is, however, an artifice and only temporary.

The procedure assigns a value to a relationship ranking as follows: 0 to the given combination of questions, 1 to the question combination making the

Table 6

USER RELATIONSHIPS

User Characteristic	Judged Potentially Related To	Related To*	Candidate for Relationship
User's highest degree	User's age		
User's field of degree	User's age	User's age	User's highest degree
User's job and company experience	User's age	User's age	User's highest degree
User's position within the research-development-production cycle	User's age, highest degree, field of degree, and job and company experience	User's highest degree	
User's field of position	User's age, highest degree, field of degree, and job and company experience	User's field of degree, highest degree, and age	User's position within the research-development-production cycle
User's management and salary level	User's age, highest degree, field of degree, job and company experience, position within the research-development-production cycle, and field of position	User's highest degree, job and company experience, age, and field of position	

---

*Ranked in order of contribution to each relationship.

largest contribution to the relationship, 2 to the question combination making the second largest contribution to the relationship, ..., m to the question combination making the smallest contribution to the relationship; m + 1 to the candidate for the relationship potentially making the largest contribution to the relationship, m + 2 to the candidate for the relationship potentially making the second largest contribution to the relationship, ..., $p \leq 11$ to the candidate for the relationship potentially making the smallest contribution to the relationship; and 12 to those question combinations which do not appear, although they might have appeared according to the general structure and the input/output view of the flow process. This value was selected because no combination of questions had more than 11 question combinations, which were either related to it or were candidates for the relationship.

Now the sum of these numerical values is computed for a question combination over each component, and over their aggregate for the flow process. Then the sums for each component and those for the flow process are ranked among themselves, in order of increasing size. Only a few ambiguities were present in computing these rankings and their sums. They involved questions which occurred in relationships, both alone and in question combinations. These questions were always associated with the appropriate question combination which contained them. Table 7 contains an example.

For a more refined ranking, the significance of the actual or potential contribution to a relationship and the significance of a relationship could be employed to compute weights for use in calculating a weighted sum, upon which to base the ranking. A question combination appears to make a significant contribution to the relationship, when its F to remove at the effective step is between 30 and 90 ($30 \leq$ F to remove < 90); and appears to make a highly significant contribution to the relationship, when its F to remove at that step is at or about 90 (F to remove ≥ 90). If the multiple correlation at the effective step is at or above 0.40 in absolute value, then the relationship seems to be significant. Those question combinations, whose potential F to enter at the effective step is at or above 30, appear to potentially make a significant contribution to the relationship.

Table 7

USER RANKS*

Combination of Questions \ Related Question Combinations	User's age	User's highest degree	User's field of degree	User's job and company experience	User's position within the research-development-production cycle	User's field of position	User's management and salary level
User's highest degree	---	0	---	---	---	---	---
User's field of degree	1	2	0	---	---	---	---
User's job and company experience	1	2	---	0	---	---	---
User's position within the research-development-production cycle	---	1	---	---	0	---	---
User's field of position	3	2	1	---	4	0	---
User's management and salary level	3	1	---	2	---	4	0
Question combination column total	32	8	49	50	52	52	60
Question combination rank	2	1	3	4	5-1/2	5-1/2	7

---

*Table entries are assigned, according to order of appearance in Table 6, as follows: 0 to combination of questions in "Characteristic" column: 1 to first question combination, 2 to second question combination, . . ., m to last question combination in "Related To" column; m + 1 to first question combination, m + 2 to second question combination, . . . , $p \leq 11$ to last question combination in "Candidate for Relationship" column; and 12, which is omitted for simplicity, to those question combinations not appearing.

It is both informative and suggestive to characterize combinations of questions as input factors and output factors, in designing and analyzing the flow process (see Figures 25 through 30). One must realize, however, that stepwise regression analysis can merely estimate and indicate the significance of a relationship. It cannot imply that the relationship is cause-and-effect, for this can only be determined by a thorough understanding of the flow process. Therefore the terms, input factor and output factor, are used in full recognition of the attendant advantages and disadvantages.

REFERENCES

1. A. F. Goodman, J. D. Hodges, Jr., et al. Final Report, DOD User-Needs Study, Phase II: Flow of Scientific and Technical Information within the Defense Industry. North American Rockwell Corporation - Autonetics Division Report No. C6-2442/030, Volumes I, II, and III (AD 647 111, AD 647 112, and AD 649 284), November 1966.

2. A. F. Goodman, L. Gainen, and C. O. Beum, Jr. Complete System Analysis: Quantitative System Analysis, Computer Simulation, and System Optimization. McDonnell Douglas Astronautics Company — Western Division Paper No. DP-4431, Revised September 1968.

3. W. J. Dixon, Editor. BMD: Biomedical Computer Programs. Health Sciences Computing Facility, School of Medicine, University of California, Los Angeles, Revised 1965.

Appendix
PARTICIPATING ORGANIZATIONS

Table A-1 lists the organizations whose personnel were interviewed for the investigation.

Table A-1 (Page 1 of 3)
ORGANIZATIONS

Organization	Number of Persons Interviewed	Population of Qualified Personnel
Aerospace Corporation	25	1,800
Allegheny Ludlum Steel Corporation	1	80
Allis-Chalmers Manufacturing Company	2	185
American Machine & Foundry Company	1	100
Ampex Corporation	10	760
Arthur D. Little, Inc.	7	800
Armstrong Cork Company	4	210
AVCO Corporation, Research and Development Division	31	3,500
The Babcock & Wilcox Company	3	250
Battelle Memorial Institute	11	775
Bechtel Corporation	1	70
Beech Aircraft Corporation	6	470
Bell Aerosystems Company	11	1,000
Bell & Howell Research Center	3	500
The Bendix Corporation	6	500
Bissett-Berman Corporation	1	65
The Boeing Company	64	6,600
Colt Industries, Inc.	8	725
Cornell Aeronautical Laboratory, Inc.	6	450
Corning Glass Works	5	450
De Laval Turbine, Inc.	2	160
Douglas Aircraft Company, Inc.	66	8,645
Dupont Company, Inc.	45	3,200
Electric Storage Battery Company	1	200
Emerson Electric Company of St. Louis	5	325
Fairchild-Hiller Corporation, Republic Aviation Division*	1	

---

*The person from Republic Aviation had just joined the company at which he was interviewed. His answers to questions reflect his position, task, and so forth, at Republic Aviation.

2

Table A-1 (Page 2 of 3)

Organization	Number of Persons Interviewed	Population of Qualified Personnel
GCA Corporation, Technology Division	3	145
General Dynamics Corporation	129	13, 155
General Precision, Inc., Link Group	8	315
Goodway Printing Company, Inc.	3	200
Hamilton Watch Company	1	110
Hazeltine Corporation	10	800
Hercules Powder Company	23	1, 350
Honeywell, Inc., Aeronautical Division	12	910
HRB Singer, Inc.	6	385
IBM, Federal Systems Division	34	3, 780
Ingersoll-Rand Company	1	55
Institute for Defense Analysis	15	400
Institute of Science & Technology	4	475
International Harvester Company, Solar Division	4	250
International Resistance Company	1	65
Johns Hopkins University, Applied Physics Laboratory	14	860
Kollsman Instrument Corporation	4	250
Lear Siegler, Inc., Power Equipment Division	9	255
Leesona Moos Laboratories	1	100
Ling-Temco-Vought, Inc.	63	3, 500
Loral Electronics Systems	4	350
Lord Corporation	2	125
Lundy Electronics & Systems, Inc.	1	60
Management Systems Corporation	1	20
Massachusetts Institute of Technology	32	2, 000
Monsanto Company	44	3, 500
Martin Company	100	7, 000
McDonnell Aircraft Corporation	27	1, 900
Melpar, Inc.	8	900
Menasco Manufacturing Company	1	65
North American Aviation, Inc., Columbus Division	21	1, 570
North American Aviation, Inc., Divisions in the Los Angeles Metropolitan Area	269	18, 590
Northrop Corporation	29	1, 730
Olin Research Center	4	300
Otis Elevator Company	1	50
Philco Corporation	26	5, 000
Pittsburgh Plate Glass Company	3	225
The Rand Corporation	11	750
Raytheon Company	49	4, 000

Table A-1 (Page 3 of 3)

Organization	Number of Persons Interviewed	Population of Qualified Personnel
Remington Arms Company, Inc.	3	135
Simmonds Precision Products, Inc.	2	190
Sparton Corporation, Electronics Division	1	35
Sperry Gyroscope Company	9	650
Sprague Electric Company	7	540
Stanford Research Institute	17	1,220
System Development Corporation	25	850
Texas Instruments, Inc.	25	1,500
Thompson Ramo-Wooldridge Inc., Equipment Laboratories	7	450
The Timkin Roller Bearing Company	5	355
United Aircraft Corporation, Norden Division	4	275
United Aircraft Corporation, Sikorsky Aircraft Division	18	1,125
United States Steel Corporation	9	700
University of Pittsburg	7	500
University of Southern California	29	1,400
Vickers, Inc.	5	380
Western Electric Company	1	120
Westinghouse Electric Corporation	22	1,730
	1,500	119,470

MCDONNELL DOUGLAS
CORPORATION

MCDONNELL DOUGLAS ASTRONAUTICS COMPANY
WESTERN DIVISION
5301 Bolsa Avenue, Huntington Beach, CA 92647 (714) 897-0311