technical note technical note tech

Air Traffic Controller Workload: An Examination of Workload Probe

Earl S. Stein

April 1985

DOT/FAA/CT-TN84/24

Document is on file at the Technical Center Library, Atlantic City Airport, N.J. 08405

U.S. Department of Transportation
Federal Aviation Administration

Technical Center
Atlantic City Airport, N.J. 08405

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN84/24		

4. Title and Subtitle	5. Report Date
AIR TRAFFIC CONTROLLER WORKLOAD: AN EXAMINATION OF WORKLOAD PROBE	April 1985
	6. Performing Organization Code
7. Author(s) Earl S. Stein	8. Performing Organization Report No. DOT/FAA/CT-TN84/24
9. Performing Organization Name and Address Federal Aviation Administration Technical Center Atlantic City Airport, New Jersey 08405	10. Work Unit No. (TRAIS) 11. Contract or Grant No. F-20-320-1
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Technical Center Atlantic City Airport, New Jersey 08405	13. Type of Report and Period Covered Technical Note 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Air traffic controller workload represents a human response to the demands or taskloads produced by the airspace system. The prediction of workload could be a useful tool for planning and staffing. The automated en route air traffic control (AERA) program proposes such a tool. Ten air traffic controllers were exposed to a series of 1-hour simulations which were designed to produce a workload range of low, moderate, and high. Controller and observer responses indicated that three levels of workload were generated and that the workload was directly related to the difficulty of the control tasks performed. Workload was particularly influenced by airspace factors of aircraft count, clustering, and restricted airspace. Overall results support the premise that workload could be predictable using measures of system activity.

17. Key Words	18. Distribution Statement
Workload Taskload Air Traffic Control Workload Prediction Workload Probe Human Factors Operator Stress	Document is on file at the Technical Center Library, Atlantic City Airport, New Jersey 08405

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	139	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (Continued)

EXECUTIVE SUMMARY

An air traffic controller operates in a complex person-machine system in which he is subject to multiple demands or taskloads over time. His workload in response to those taskloads will be a function of what he brings with him to the situation (knowledge, abilities, and skills) and what he must do in order to maintain a safe and expeditious traffic flow.

The prediction of controller workload could be a useful management tool for such decisions as sectorization and staffing. Such a tool has been proposed by the "workload probe" concept of the automated en route air traffic control (AERA) program. This concept suggested four catagories of measures drawn from system operations which might be useful in predicting the workload of sector controllers. These catagories included: (1) Aircraft Count, (2) Encounter Count — the frequency of minimum separation violations, (3) Clustering — the average number of aircraft in a small block of airspace, and (4) Planned Actions. This later catagory included a series of control actions such as altitude and heading changes.

The Federal Aviation Administration (FAA) Technical Center began a project to study the feasibility of workload probe and it's potential for use in advanced automation systems. A simulation study was designed and carried out at the Technical Center's National System Support Facility (NSSF).

Ten en route field controllers volunteered to participate in the experiment. They received training/familization and then were exposed to a series of 1-hour simulations. These simulations were designed to produce a range of possible workload from low, through moderate, to high — where some tasks might have to be left incompleted. Each controller spent about 4 days at the Technical Center, and test runs were spread out over that period.

RESULTS.

Participant responses on several measures demonstrated that three levels of workload were achieved. These were directly related to the difficulty of the control tasks produced by the research design. This meant that there was enough variability in the participants workload so that workload probe measures could receive a fair test.

A series of analyses were computed. These demonstrated that the workload probe predictors could account for about 85 percent of the variability in participants self-reported workload responses.

Two independent observers, who were controllers themselves, made real-time estimates of participant workload and effectiveness. These estimates proved to be a reliable source of information and provided confirmation of the participant controllers' self-reports of how hard they were working. The observers noted a tendency for effectiveness to decrease when the workload was very high.

Participants were interviewed after each experimental simulation. They felt the simulations were reasonably realistic. They were challenged by the tasks they had to perform, particularly at the higher levels of workload. They were able to point to several variables which they felt significantly influenced their

workload. These included aircraft count, clustering, and the impact of restricted areas within sectors. This later element influenced the number and type of planned actions the controller must initiate.

The overall results support the proposition that controller workload could be predictable using measures of system activity. Workload probe appears to be a feasible concept. However, there was considerable variability between the participating seasoned controllers in terms of their workload responses to given levels of system activity. The use of automated workload predictors should be guardedly tempered by the supervisor's knowledge of individual controller habits and abilities.

INTRODUCTION

BACKGROUND.

The purpose of air traffic control is to introduce order into the airspace and provide safe, expeditious, and economical flight for aircraft under control. Air traffic control (ATC) is a highly complex process involving the interaction of multiple person-machine subsystems. In each of these subsystems, controllers must perform under time and event pressure in order to maintain aircraft flow. The controller shares in the responsibility for insuring system operation and, in doing so, is subjected to a variety of stressors which influence each operator in his own unique way (Smoller and Schulman, 1982). Over the history of aviation, equipment has become increasingly reliable, while the human operator, although being highly adaptable, has also remained fallible (Roscoe, 1978).

Controllers must function in a multitask environment in which they not only act and react with machines, but also interact with other controllers to affect coordination. Treating controllers as only components of the overall system without considering their needs and abilities may lead to incorrect conclusions on potential system performance (Hopkin, 1970). Buckley (1969) documented the extent of variability in terms of how well controllers perform their tasks. Human performance and workload are characterized by the dynamic aspects of this variability.

One aspect of the environment which will likely influence air traffic control performance is the system's taskload. This involves the stimuli and behavioral demands placed on the system by the traffic flow, the environment, and sector specific requirements. A human response to taskload is workload. There has been a wealth of literature concerning the construct of workload, and the central theme of this material is that there has never been a generally agreed upon definition of workload (Moary, 1982; Melton, 1979; Rehmann, 1982). Workload is usually defined in terms with which the person doing the defining is comfortable. Most researchers do agree that regardless of the person-machine system, motivated operators can chose to increase their workload in response to increasing taskload from system demands (Sheridan and Simpson, 1979). However, there will always be a functional limit based on the operators internalized performance standards, motivation, skills, knowledge and abilities. Even an operator's age may influence his/her performance and the resultant system operation (Cobb, Nelson, and Mathews, 1973).

In aviation and air traffic control, there are few degrees of freedom for lowering performance, and operators must increase workload to maintain operations to standards. Controllers, who fall below standards, risk the incidence of systems' errors which could have serious consequences. It would be highly desirable to maintain workload within a range so that operators were challenged but not overloaded. What types of activity might be involved in controller workload?

Coulouris, et al. (1974), defined the essence of controller workload as the amount of time spent performing observable control actions such as communicating, making keyboard data entries, and marking flight strips. This philosophy would be in line with a strictly behavioristic view of human activity. However, there has been a growing emphasis on what psychologists refer to as cognitive behaviors.

Maintaining a smooth traffic flow requires that controllers process a considerable quantity of information and make decisions. These cognitive actions add to whatever workload is, even if they cannot be directly observed (Kirchner and Laurig, 1971). To further complicate this situation, operators will establish different types of strategies which may vary in response to taskload (Sperandio, 1971). Any systems which depend on human judgment and vigilance will be subject to error (Danaher, 1980).

More information is needed about human workload in complex person-machine systems. This is particularly true in air traffic control, so that steps can be taken to get the most out of the airspace system while maintaining safety. It could be very useful to identify which systems variables had the greatest impact on system busyness; on the taskload; and, by logical inference, on the workload of the people who operate the system.

SECTOR WORKLOAD PROBE

It is anticipated that over the next 20 years, the National Airspace System will have to handle a much larger volume of increasingly complex aircraft. To deal with this enhanced load on the airspace system, a series of automation tools have been proposed. One of these tools is the "workload probe" (Swedish, 1982; Barrer, 1982; Swedish and Neidringhaus, 1983). This is a theoretical prediction model for system taskload. According to Neidringhaus and Gisch (1983), "The sector 'workload probe' calculates measures related to workload, which is defined as any task performed by personnel to provide air traffic control services to aircraft (p 1-3)." Niedringhaus and Gish (1983) have made a discrimination between sector workload and controller workload. Sector workload referred to the tasks performed within the sector. Controller workload is the human response to the sector workload.

"Sector workload" is therefore an indicator of system activity, complexity, or taskload. It is conceived of as a management tool which would provide information to first and second level supervisors so that they could anticipate staffing and sectorization decisions. The "workload probe" was to consist of four variables, three of which were to be univariate and one of which was to be a linear weighted sum of subvariables. The first three were (1) aircraft count, (2) encounter count, and (3) complexity or clustering, which referred to variability in local aircraft density. The fourth variable was referred to as the frequency of planned actions. A weighted sum of such planned actions as altitude, route, and heading changes was to be computed. However, to date, no suggested system of weights or combinational rules has been developed.

PURPOSE

This study was designed and carried out to examine the feasibility of using variables indicative of system taskload (sector workload in Mitre's terms) to predict human operator workload. The prediction of system activity without considering the person responsible for moving the traffic has little meaning. The relationship between measures of operator workload and performance was also of interest. A number of alternative measurement techniques were tried. The study design was built around current national airspace system concepts since no simulation of advanced automation system within the automated en route air traffic control (AERA) program was currently possible. This study served another purpose also. It was the starting point for the development of measurement concepts and tools to be used later in operational suitability testing of elements of the AERA program as they become available.

METHOD

PARTICIPANTS.

The ten traffic controllers who took part in this study were all volunteers from a major eastern en route center. Selection was based on a nonrandom, as available basis. Participants were all current nonmanagement controllers who had worked radar positions continuously for the past 12 months. They ranged in experience from 4 to 18 years with a median of 15 years. Their median age was 43, with a range from 27 to 48. During a preliminary briefing, each received an entry questionnaire in which they were asked to rate their current controller skill, their agreement with their status as "volunteers," and whether or not they were looking forward to the experiment. The mean self-rating of skill was 7.8 (SD=1.55) on the scale where 10 was defined as high skill. The agreement with volunteer status was unanimous. Nine responded with a "10" indicating strong agreement, and one responded with an "8." The mean response to the statement, "I am looking forward to participating," was 9.3 (SD=.95). It could be assumed that all the participants were motivated.

Each individual was briefed concerning his rights to informed consent and privacy. Data collection was accomplished using arbitrary code numbers so that specific data could not be traced back to an individual participant. This had the advantage of not only meeting the privacy requirements, but also encouraged accurate responses to questionnaires and interviews.

RESEARCH DESIGN.

This experiment could be viewed as a three-stage process. The first stage involved the development of traffic samples/scenarios using a combination of subject matter expertise and the results of previous work accomplished at the Technical Center by Buckley, et al. (1983). These traffic samples were evolved in a very short time period and received some limited pretesting prior to the experiment. Also created at the same time were several scenarios for training and familiarization. Prior to any data collection runs, each participant received no less than 6 hours of hands-on training and familiarization with the same sector and equipment he would use during the experiment itself. During training, observers provided immediate feedback on errors and coached/encouraged the participants. The training was modeled after the Instructional Systems Design (ISD) technique. The decision to proceed from the training phase to the data collection portion of the design was based on a consensus (among the observers, the experimenter, and the participant) that the participant was comfortable with the sector, the equipment, and his ability in the environment.

The actual experimental design involved two key independent variables — taskload and replication. Three levels of taskload were defined in terms of average aircraft density, planned and unplanned constraints, as described in table 1.

TABLE 1. TASKLOAD DESIGN

Taskload Variables	Taskload: Low (A)	Medium (B)	High (C)
Avg. aircraft density	5	10	15
Planned constraints	Few constraints	Altitude changes and some vectoring	Vectors, altitude changes, hot restricted areas
Unplanned constraints	No conflicts	Few conflicts	Moderate conflicts

Taskload was defined as the complex of stimuli and requirements which imposed some level of workload. Workload would also be influenced by what the controller brought with him to the situation. This would involve skills, knowledge, abilities, and motivation. The two extremes of taskload were defined as follows: low taskload — the level and nature of events where the controller is just keeping busy; high taskload — the level and nature of events where the controller is very busy with multiple demands being made on his time. He can just maintain his picture but will probably have to leave some tasks (i.e., strip-making) incomplete. A basic repeated measures design included two replications of the three levels of taskload. The order of administration of the three levels of taskload was counter-balanced within each replication. The purpose of counter-balancing was to remove any potential order effects based on residual learning and/or experience. The two replications were to provide an estimate of measurement reliability as an element in the research design, where the variance due to the replication main effect could be determined. The amount of this variance would provide a key to the relevence of learning/familiarization in the post-training phase of the experiment. Table 2 presents the summary of the experimental design.

EQUIPMENT.

This research was accomplished using the National System Support Facility (NSSF) available at the FAA Technical Center. This simulation facility consists of three subsystems. The controller subsystem provides the sights and sounds of the ATC control room. Here the participant controller is provided the basic radar display and control input devices with which he must work. The simulator pilot subsystem provides the bridge between the controllers and a group of people who influence heading, altitude, and airspeed of simulated aircraft as they appear on the radar display. The controller maintains voice communications with the "pilots" in his piece of airspace, and he must provide them with positive radar control. He must also coordinate his actions with those of controllers in adjacent sectors. The function of these adjacent sector controllers is role-played by one or two "ghosts" who operate from another radar position in the same room but are screened from the participant's view. The processor subsystem is the final element in the simulation. The sigma 8 computer simulates and controls all aspects of the airspace system except those initiated by the controllers and simulation "pilots." The computer also monitors the status of all aircraft and the controller's keyboard entries and communications. This information is stored on tape for later processing.

TABLE 2. ATCS WORKLOAD PROJECT DESIGN MATRIX

Participant Number	Administrative Order Block 1			Block 2		
10	A	B	C	C	A	B
11	C	A	B	B	C	A
12	B	A	C	A	B	C
13	A	B	C	C	B	A
14	C	B	A	B	A	
15	B	A	C	A	B	
16	A	C	B	C	A	B
17	C	B	A	B	C	A
18	B	C	A	A	C	B
19	A	C	B	C	B	A

Taskload

A	
B	Moderate
C	

Note: The only difference between blocks are the aircraft identifications

The computer stores an internalized map of the airspace geometry to include the locations of navigational aids (NAVAIDs), fixes, and other relevant features. The sector employed in this study was one developed by Buckley, et al. (1983), for a previous study of system effectiveness measures. This sector evolved from one that was actually in use at Washington Center. It was employed because the software was already available and there was inadequate time to create another.

PROCEDURE.

Each participant arrived at the Technical Center on a Monday morning and received an initial briefing about the experiment. This briefing, which was conducted informally by the psychologist, provided an overview of what to expect during the experiment and described the participants rights to informed consent and privacy. The latter was guaranteed by using an arbitrary code number for each participant. These code numbers (along with the implied administrative order for the research design) were assigned to each participant randomly.

Upon completion of the initial briefing, the participant was asked if he still wished to continue. All responded positively. A background questionnaire was then administered (see appendix A). The results of this questionnaire were summarized in the participant description of this report.

The participant walked to the facility laboratory in another part of the Technical Center and was introduced to the laboratory supervisor. At this point, and in graduated steps, the familiarization and training began. The participant was given a tour of the facility to include the air traffic control room, simulator pilot room, and the computer facility. This was done to enhance his feeling that he was part of the program and not a passive subject of experimenter whims. The guidelines that were developed for this training/familiarization program are listed in appendix A.

The participant controller received several hours of what might be called keyboard practice in which one or two aircraft were presented on the radar display. A keyboard entry guide (see appendix A) was placed in front of the controller, and he practiced at his own speed with no pressure and no external observers. At the end of this experience, it was usually time for lunch.

Work-rest cycles were planned as part of the research program. Controllers were never asked to work the radar position for more than 4 hours per day with less than 1-hour breaks between control periods. Whenever possible, these breaks were extended to take into account interview and questionnaire periods following runs. For the preponderance of the experiment, controllers did not work simulated traffic more than 3 hours per day.

Training/familiarization continued with the first 1-hour run which usually took place on Monday afternoon. The controller was introduced to the observers who would be sitting at his side during training and data collection. They would coach him during training and make their presence part of his routine. The senior observer provided him with a briefing concerning sector procedures prior to the first run.

An outline of this briefing is available in appendix A. The lights were dimmed in the control room approximately 5 minutes prior to the onset of the control period.

Prior to the first training run, the psychologist explained the use of the Air Traffic Workload Input Technique (ATWIT) system. A briefing (see appendix A) was read to the participant. The participant was told that one purpose of the experiment was to obtain accurate estimates of his workload. He was asked to provide these estimates once each minute by pushing one button on a scale of ten buttons from 1 (low) to 10 (high) in terms of workload. These buttons were verbally anchored using an adaptation of a scale previously employed by Stein & Rosenberg (1983) for aircrew research.

This technique is based on a wholistic or Gestalt model of workload which assumes that operators experience workload as a whole and can provide accurate estimates of how hard they are working.

Once the briefings were completed, the first training run was accomplished. Actually, only one training scenario was written in two forms. The difference between the forms was the aircraft identifications. These forms, T-1 and T-2, were alternated throughout the training period. Both forms were designed as moderate difficulty scenarios with a desired average aircraft count of ten.

Training and subsequent data collection runs were all accomplished in the spirit of free-play simulation. Participants were told to do whatever they would normally do. The only experimentally induced constraint was that they had to work a one-person sector without a "D" person. Flight strips were presented in the strip bays prestacked in the correct order. The participant could, for example, develop his own strategies for moving the traffic and coordinate with adjacent sectors as he normally would. He was told that he could refuse inbound handoffs if he felt he had all the traffic he could handle. During training, participants learned by trial and error, and through some coaching, what they could realistically do. At the end of each training run, the participant and the psychologist, who had been present during the run, adjourned to another room. Here, an informal interview was conducted and a post-run workload questionnaire was administered. The purpose of this session was not so much for data collection as it was to give the participant an opportunity to express his feelings during the run and to provide assurance, particularly during the first few training runs.

The experimental design called for no less than four 1-hour training runs — not counting the Monday morning keyboard practice. At the end of the fourth run, the design called for a go - no go decision in terms of data collection or more training. In every case, the controllers felt they were ready to move into actual data collection runs and the observers/laboratory supervisor agreed. Great care was taken not to pressure anyone to make this decision prematurely, and there was more than adequate time scheduled for each participant — a 40-hour week to conduct training and 6 hours of data runs.

At the completion of training, the data collection phase of the experiment began the first block of data collection runs. The order of administration was based on the counter-balanced design and where the individual fit into the design based on his code number.

Each hour of simulation was conducted as consistently as possible with the observers sitting at either side of the participant. The controller was informed that the observers could no longer answer his questions and could not show him the results on the forms that the observers were completing (appendix A.) However, at the end of the experiment he could examine anything he wished concerning his experiences at the Technical Center. At the completion of the first, second, fourth, and fifth hours of simulation, the controller received an informal interview and he completed the air traffic control specialist workload questionnaire. At the end of the third and sixth hours, a formal interview was administered in which specific questions were asked. Copies of the interview and questionnaire are available in appendix C.

Also, at the end of the sixth hour, two additional questions were added to the formal interview for the last eight participants. These questions asked for the controller's opinion concerning whether it was feasible to predict workload and whether they would want such a prediction down at the radar position.

Data collection runs usually occurred between Tuesday and Thursday of each typical week. On Friday morning, a debriefing was held with each participant. He completed a 25 item questionnaire, which asked him to rate the workload imposed by specific elements of the air traffic control simulation on a 10-point scale. Of the 25, two items were designed to see if the participant was paying attention. These involved the workload imposed by weather and emergencies, neither of which were part of the experimental design. A copy of this post-experiment questionnaire is in appendix C.

RESULTS SUMMARY

Many readers of this report may not be interested in examining the details of each statistical analysis reported in the Results section, therefore, this summary is offered as an alternative.

Two systems variables were examined to determine how closely the actual taskloads approached those which were planned. The achieved mean instantaneous aircraft counts for the three taskloads were 4.78, 8.68, and 10.50. For the most part, aircraft counts were significantly different across the levels of taskload. Where overlaps did occur, it was between taskloads B and C, never between A and B, or A and C. The second systems variable examined was the number of flights handled. Again, for the vast majority of comparisons between taskloads, the number of flights handled was significantly different. Like aircraft count, it increased as the planned taskload increased. The analyses of these two variables demonstrated that, for the most part, three levels of taskload (demands placed on the system and the controller) were achieved.

The next series of analyses investigated differences in ATWIT workload responses across the levels of taskload. ATWIT responses correlated well ($r = .79$) against an artifical variable which simply recoded A, B, C taskloads as 1, 2, and 3 respectively. This indicated a positive relationship between workload and taskload. Workload plots (see figure 1) and analysis of variance indicated quite clearly that there were significant increases in reported workload as one progressed to higher levels of taskload. Changes in workload across time within control runs were also apparent.

Observers completed ratings of controller workload, busyness, and effectiveness (see form in appendix C). The two observers were in substantial agreement on the majority of their observations indicating a high level of interrater reliability (median r =.91). Ratings on workload and busyness scales were essentially the same. Observer ratings of workload correlated well against the taskload levels (r =.91) and moderately well against participant's ATWIT responses (r =.86). Observers workload ratings were inversely related to effectiveness estimates (r = -.55). While this was not a strong relationship, it is consistant with other findings concerning workload and performance.

A tabulation of all the bivariate relationships in this study is reported. Many systems variables were well correlated with the various indices of operator workload. These relationships appear to peak at r =.81 for NG2A (the number of ground to air communications) and r =.80 for DFLT (the duration of all flights). The significance of these bivariate relationships is that, at best, using a single predictor only 65 percent of the variability can be accounted for in the participants self-estimates of workload (WIAV). A multivariate model might work better.

A series of stepwise regression analyses were conducted to try and find the right combination and weighting of variables which could account for more ATWIT variance than the best bivariate predictor. This predictability of workload was the essence of the workload probe concept. The first analysis concentrated on an examination of only those variables suggested by the "probe." The reader will recall that there were four catagories: aircraft count, encounter count, clustering/complexity, and planned actions. The last category included a series of controller actions, such as route, speed, heading, and altitude changes, among others.

Stepwise regressions differ from standard regression techniques in that the analyses maximize the relationships between predictors and criterion variables while minimizing the intercorrelation of predictors. Those variables which do not add significantly to the regression model are discarded.

For the workload probe regression, only three variables stepped into the equation

1. Clustering/complexity
2. Frequency of handoffs outbound
3. Frequency of heading changes.

This provided a multiple correlation of R =.85 which was significant from zero and represented a respectable relationship between systems variables and controller workload responses using the ATWIT.

Inclusion of a series of other systems variables provided another significant regression. The following variables were stepped into the equation:

1. Handoff inbound delays
2. Duration of ground to air communication
3. Number of controller keyboard entries
4. Frequency of altitude changes
5. Number of controller keyboard errors
6. Clustering/complexity.

Despite an increase in the number of variables, the multiple R only increased to R =.903, a change of eight percent in accountable variability.

Regressions were also computed against observer ratings of controller workload. Using workload probe systems variables, an equation containing the following five variables was computed:

1. Clustering/complexity
2. Aircraft count
3. Frequency of heading changes
4. Frequency of handoffs inbound
5. Frequency of altitude changes.

The computed multiple correlation was R =.931, a very healthy relationship

When all variables were used in the analysis, the following were stepped into the equation:

1. Clustering/complexity
2. Number of flights handled
3. Duration of ground to air
4. Handoff inbound delays
5. Number of controller keyboard entries
6. Number of controller keyboard errors.

This analysis produced a multiple R =.955

A researcher suggested that some of the predictor variables used in these regressions were not themselves predictable (based on the system itself without operator actions). An unbiased airspace system expert was asked to choose a subset of systems variables which were predictable. He chose the following four:

1. Clustering/complexity
2. Frequency of handoffs outbound
3. Number of flights handled
4. Frequency of handoffs inbound.

The multilinear regression provided a multiple R of .85 against ATWIT, which was identical to that achieved using the workload probe variable set. The multiple R was computed against observer ratings of workload using the same four variables. The result was R =.931 which was very close to that achieved using workload probe predictors.

The results of a postrun questionnaire basically concurred with other sources of information. Increases in workload over the three levels of taskload were well documented. A question asking for indications of stress also showed an increase as taskload increased. However, there was a decrease in reported stress over the two replications in the experiment as participants became more comfortable with their environment. The stress question correlated inversely with observer evaluation of controller effectiveness (r = -.65) indicating that as stress increased, performance decreased.

A post-experiment questionnaire asked participants to evaluate a lengthy series of tasks seen in ATC operations and rate each in terms of its contribution to workload. Items which contributed highly to workload included: number of aircraft handled, number of altitude changes, housekeeping, and using the keypack. A question concerning the impact of making the ATWIT responses led to a diversity of opinion with a median response close to the middle of the scale.

Informal interviews were analyzed, and results were tabulated. The training was somewhat stressful for the majority of the participants but experience brought an increasing comfort with the simulation. The low workload under taskload A was not stressful. However, under B and C, not all tasks were completed and some participants managed to "lose the picture."

The formal interviews provided a more structured opportunity to obtain information from participants. Controllers indicated that the training they received was adequate and the simulation of ATC was reasonably realistic. The most frequently cited factors in workload were the volume of traffic, complexity, and the impact of restricted airspace. In general, those factors which they felt most influenced their everyday work and performance included weather, traffic volume, other controllers in adjacent sectors, and complexity, which they defined as the number of actions required to move aircraft through the airspace.

This concludes the results summary. The goal of this section was to highlight the basic results without going into too much detail. The reader may use this summary to help understand the detail in the Results section or to avoid it.

RESULTS

QUALIFICATIONS.

This was a simulation study using a nonrandom, small sample of volunteers drawn from one En Route Center. All simulation represents a balance between cost (time, resources, finances) and fidelity or realism. There is virtually no such thing as a perfect simulation, because each can almost always be improved in terms of stimulus or response fidelity. When asked during the interview about the realism of the air traffic control simulation, participants indicated that while not perfect, it was reasonably realistic. They assigned a median response of 7.5 on a 10-point scale where 10 represented "highly realistic."

Another qualification that should be considered was the fairly restricted range of participant age and experience. These two variables, which correlated $r = .77$, demonstrated that the sample with a median age of 43 was more representative of senior, highly experienced controllers, most of whom had been at the same Center for their entire career. Participants self-rating of current skill level on a 10-point scale ranged from a value of 6 to 10, with a mean of 7.8. Skill ratings surprisingly were not correlated with age or experience.

Participants did become both physically and emotionally involved in this study. To a person, they appeared to try to do their very best. Any generalizations drawn from these results should take into account both the study qualifications and the motivation and professionalism of the controllers who were involved. Results from this study should be viewed as indicative rather than conclusive.

SCOPE.

This experiment was an effort to do a preliminary examination of a number of systems variables in terms of human operator workload. Variables were chosen based on a number of criteria. The first goal was to use what was already available through the software constructed for the ATC simulator at the Technical Center. Fortunately, this included a majority of elements suggested under the workload probe construct. Additional variables were added to this list. These variables were developed specifically for this project and did not involve major programming effort. These included the observer scales, questionnaire and interview items, and the real-time workload measurement technique, which we are referring to as ATWIT (Air Traffic Workload Input Technique). Table 3 provides a listing of the real time variables and their abbreviation codes. These codes will be used in the text for simplicity. The reader will want to refer back to this table as the discussion proceeds.

The analyses to be discussed in this results section have been selected based on the objectives of the study. Given the number of variables, the choice of what to analyze closely and how far to take each analytical approach was practically limitless. The reader may identify additional techniques or variables he/she feels deserve further statistical treatment.

This series of analyses will be discussed in a number of steps. First, several systems variables will be reviewed to try and identify what occurred in the taskload environment. These will include instantaneous aircraft count (NIAC) and the number of flights handled (NFLT). Next, will be an analysis of the real-time ATWIT workload responses (WIAV); this will be coupled with an examination of the bivariate relationships between pairs of system and workload variables. Particular emphasis will be placed on those systems variables discussed under "Sector Workload Probe."

The analysis this far will be based on what we call the 15-minute data set. Systems variables are sampled every second. However, this produces more numbers than we can reasonably analyze. Preliminary reduction of the data has produced two data sets. The 15-minute set involves point estimators — usually the average for each variable for each 15-minute period of time. The other data set involves cumulative averages over the 60 minutes of each data run. This latter data set has less internal variability (it covers up some of it by cumulating), but it has the advantage of being comparable against both real-time and postrun measures, such as questionnaires.

A series of scatterplots and correlations will be reported. These will compare all systems variables against the workload measure collected in real time. These will be bivariate (one on one) relationships. A series of regressions will be reported which show how systems variables can be empirically combined to predict operator workload responses.

Next, postrun and post-experiment questionnaires will be reviewed and analyzed. This will be followed by an analysis of interview data collected at the end of each run. A final brief section will look at the training runs to demonstrate that they accomplished what they were designed to accomplish. Throughout this section, the first time an analysis is used, it will be explained in detail.

TABLE 3. KEY VARIABLES

(M) = MITRE Variables

Abbreviation Code	Variable Name
ALT (M)	Altitude Change
CKEN	Controller Keyboard Entries
CKER	Controller Keyboard Errors
CMAV (M)	"Complexity" (Local Density)
CMTR	Communication Transfers
DAIR	Duration of Air to Ground Contacts
DDLY (M)	Duration of Holds and Turns $\geq$ 100 Seconds
DFLT	Duration of Flights Handled
DG2G	Duration of Ground to Air Contacts
DG2G	Duration of Land Line Contacts
DINB	Duration of Flights Inside Sector
DIST	Distance Flown by all Aircraft
D5CF (M)	Duration of Conflicts
FUEL	Fuel Consumed by all Aircraft
HDG (M)	Heading Change
HOID	Handoff Delay Time
HOIN (M)	Inbound Handoff
HOLD	Hold Frequency
HOLD (M)	Hold Message Employed
HOUT (M)	Outbound Handoff
NAIR	Frequency of Air to Ground Contacts
NDLY (M)	Frequency of Holds and Turns $\geq$ 100 Seconds
NFLT	Number of Flights Handled
NG2A	Frequency of Ground to Air Contacts
NG2G	Frequency of Land Line Contacts
NIAC (M)	Instantaneous Aircraft Count
N5CF (M)	Number of Conflicts
OBQ1	Observer Rating — Workload
OBQ2	Observer Rating — Busyness
OBQ4	Observer Rating — Effectiveness
RTE (M)	Route Change
SAM	Sample: A (1) B (2) C (3)
SPD (M)	Speed Change
WDAV	Workload Input Delay
WIAV	Real-Time Workload Input

AN EXAMINATION OF TWO TRAFFIC SAMPLE VARIABLES.

It is not a simple task to define what variables might influence the taskload of air traffic controllers. Instantaneous aircraft count or the average number of aircraft over a time period has achieved a certain amount of face validity and with it an amount of credibility. It has served as a general indicator of system busyness and, by implication, workload. It has often been related to other measures of system operation, and controllers usually mention the number of aircraft as an influencing factor in their workload.

Our goal in the research design was to have three distinct levels of taskload of 5, 10, and 15 instantaneous aircraft count (NIAC) respectively. As indicated in table 4, overall means for the three taskloads actually achieved were 4.78, 8.68, and 10.50. An examination of NIAC across 15-minute time blocks and across replications is informative. There appears to be some variability across the time blocks, but little or none across replications. Variability across taskloads appears fairly strong.

In order to objectively analyze the appearances in the summary table of means, an analysis of variance was applied to the aircraft count data. Analysis of variance (ANOVA) is a means of comparing treatment variability (i.e., attempting to make the system operate at three levels of NIAC) to error variability, which is produced by numerous uncontrollable influences, such as individual controller strategy. When a treatment has had a significant effect, then differences induced by the independent variable(s) exceed those generated by error sources. We can then conclude that the treatment differences did not occur by chance alone. Figure 1 graphically describes the information presented in table 4.

TABLE 4. MEAN INSTANTANEOUS AIRCRAFT COUNT

	First Replication					Second Replication					
	Time Block					Time Block					
Task Load		2	3	4	Run Mean	1	2	3	4	Run Mean	Overall Mean
A	3.63	5.25	5.01	5.09	4.75	3.54	5.35	5.01	5.39	4.82	4.78
B	7.37	10.51	10.19	7.33	8.85	7.48	9.17	9.52	7.85	8.51	8.68
	8.99	10.77	10.76	10.45	10.24	9.19	11.72	10.37	11.72	10.75	10.50
Time Block Means											
	6.66	8.84	8.65	7.62		6.74	8.75	8.30	8.32		

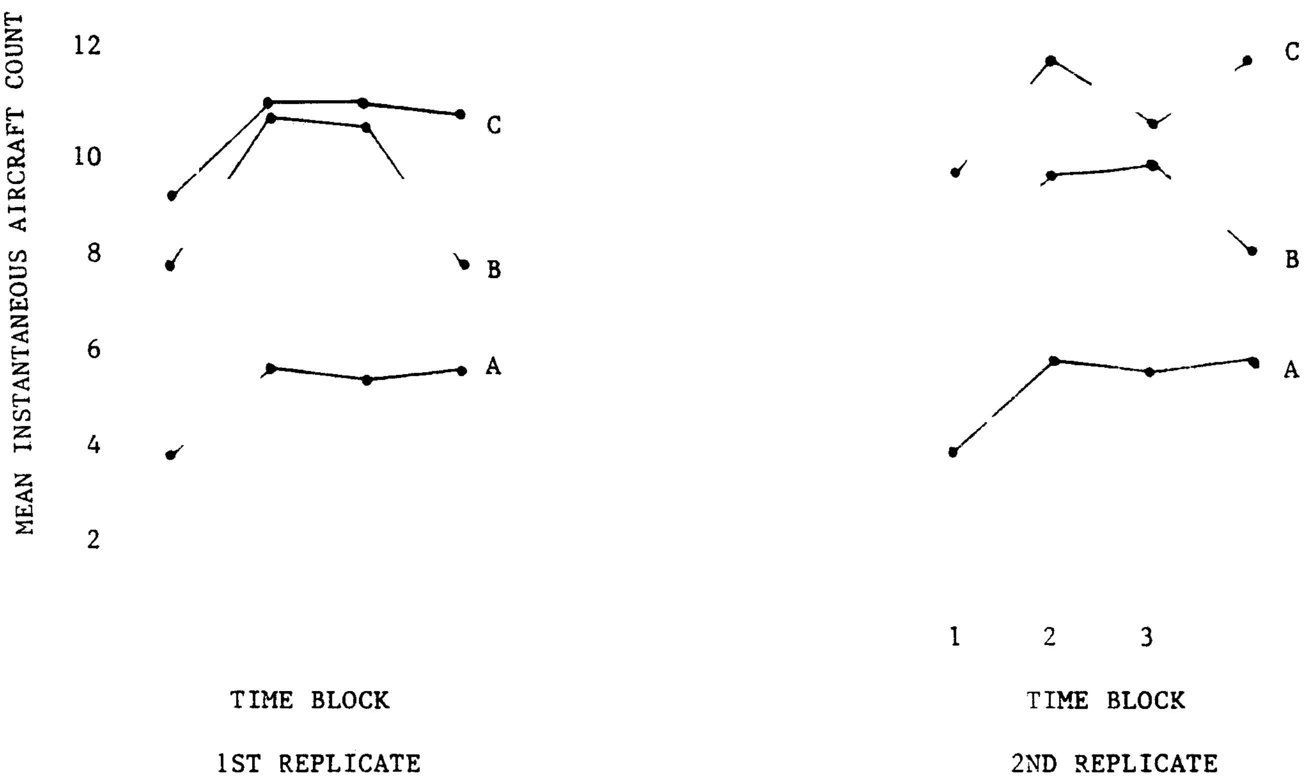


FIGURE 1. PLOTTED MEAN INSTANTANEOUS AIRCRAFT COUNTS

Table 5 summarizes a three-way analysis of variance in which three independent variables are involved. These include the taskload (A, B, C), the time block (1, 2, 3, 4), and the replication (1 or 2). The dependent variable analyzed was the instantaneous aircraft count for each 15-minute time block. Aircraft counts were not frozen, but were influenced by the dynamic flow of the simulation. Table 5 reports the sources of variance, the degrees of freedom upon which the source was based, the F ratio, and the correlation ratio. The F ratio is the relationship of treatment variance to error variance. The correlation ratio is a measure of what is called the strength of association. It provides an estimate of the proportion of total variance which is accountable from a given source. A given effect may be significant (unlikely to have occurred by chance) but can account for such a small segment of the total variance that its meaningfulness is suspect. A rule of thumb, proposed by Linton and Gallow (1975) suggests that effects which account for less than 10 percent of the variability in a design should lead to guarded conclusions. In table 5, the only significant effect which exceeded the 10 percent criterion was the taskload main effect. Effects which account for 78 percent of the variability in an ANOVA are rare. Table 5 also demonstrates that there was no significant shift in NIAC between the first group of three data runs and the second group. Had there been such a shift, it would have represented a design deficiency. It appears that the experimental design was successful in producing three levels of NIAC. This can be seen in figure 2 which shows the plotted means of NIAC. In figure 2, it appears that there were two time blocks (2, 3) in the first replication and one time block (3) in the second replication in which there was no meaningful difference between taskloads B and C. This brings out a deficiency in the ANOVA procedure.

TABLE 5. SUMMARY OF ANOVA ON INSTANTANEOUS AIRCRAFT COUNT

Source of Variance	Degrees of Freedom	F Ratio	Correlation Ratio in %
Task Load (A)	2,18	241.05**	78.57
Time Block (B)	3,27	53.60**	8.81
A X B Interaction	6,54	14.79**	3.11
Replication (C)	1,9	1.06	0.02
A X C Interaction	2,18	13.07**	0.41
B X C Interaction	3,27	12.93**	0.52
A B C Interaction	6,54	6.33**	0.59

**P $\leq .01$

NOTE: Correlation between NIAC and WIAV (ATWIT) was =.776

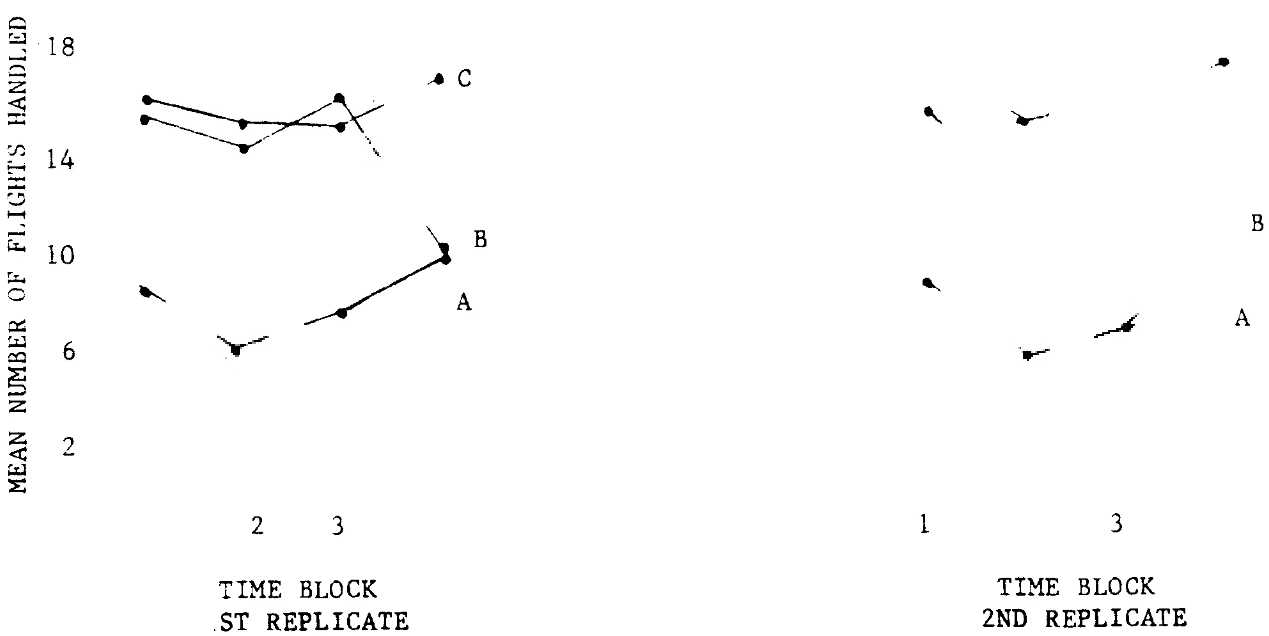


FIGURE 2. PLOTTED MEAN NUMBER OF FLIGHTS HANDLED

When you find that you have a significant main effect, such as taskload in table 4, it tells you only that something is significantly different from something else. It does not tell you where the differences lie. This requires what has been referred to as post-hoc testing. The first step is to break the experimental design down into those components that matter to the researcher. These are usually referred to as simple effects. Of primary concern in this experiment was whether or not NIAC differences were achieved across the three taskloads. The design was broken down so that paired comparisons could be made in NIAC across the taskloads for each time block at each replication. An analysis of simple main effects reported in table 6 indicated that at every time block computed, F ratios were significant. This was not surprising, since simple main effects are basically mini ANOVAs and like ANOVAs, they only tell you that significance exists, not where it exists.

TABLE 6. SIMPLE MAIN EFFECTS — NIAC

	Time Block	Mean Square	F Ratio
First Replication			
	1	75.41	26.64**
	2	97.14	34.33**
	3	100.47	35.50**
	4	72.44	25.60**
Second Replication			
	1	83.94	29.66**
	2	102.70	36.29**
	3	83.02	29.33**
	4	101.97	36.03**

**$P \leq .01$

The next analysis reported in table 7 is based on the Newman-Keuls technique. This is a method of making multiple paired comparisons and testing each pair against a significance criterion which takes into account the number of ordered steps between the two means being compared. Table 7 shows the comparisons for each time block and replication. The means for NIAC appear on the top and left side of each box. Within the box are the differences between these means, which are compared against the critical values at the bottom of the table. Of all the paired comparisons, only three were not significant — the same three as would have been assumed from figure 2. These included NIAC between taskloads B and C for two time blocks the first replication and one block in the second. If workload differences occurred between these taskloads, then they would have to be driven by more than differences in NIAC alone. A workload analysis will be presented later in this section of the report.

.ED AC)AD: ME BLOC

<u>irst Replication</u>

1			2			3			4		
A	B	C	A	B	C	A	B	C	A	B	C

<u>Second Replication</u>

1			2		3			4		
A	B	C	B	C	A	B	C	A	B	C

<u>Ordered Steps</u>

2 3

A second system variable which was chosen for further analysis was the number of flights handled per unit of time (NFLT). Table 8 summarizes the means for NFLT by taskload, time block and replication. This table coupled with figure 3, a plot of the means, provides a picture of a fairly complex variable which behaved in some unusual ways. The correlation between NFLT and NIAC was moderately high, r =.80, and this can be seen by the similarity of their plots in figures 2 and 3 respectively.

TABLE 8. MEAN NUMBER OF FLIGHTS HANDLED

	First Replication					Second Replication					
	Time Block					Time Block					
Task Load	1	2	3	4	Run Mean	1	2	3	4	Run Mean	Overall Mean
A	7.90	5.2	6.3	8.9	7.07	8.0	5.0	6.1	10.0	7.28	7.18
B	15.0	13.7	15.8	9.3	13.45	14.9	11.9	13.9	8.2	12.23	12.84
C	15.8	14.8	14.6	16.3	15.38	17.1	14.6	15.5	17.2	16.10	15.74
Time Block Means											
	12.90	11.20	12.20	11.50		13.30	10.50	11.83	11.80		

An ANOVA was computed on NFLT, and a summary is presented in table 9. This ANOVA summary indicates significant main effects for both taskload and time block. These effects, however, are complicated by the interactions between variables. Again post-hoc analysis was required. It was decided to focus on the differences in NFLT across taskloads at specific levels of time and replication. Anticipating that all the simple main effects would be significant as they had been in the previous analysis of NIAC, we proceeded directly to the paired comparisons, using the Newman-Keuls technique. Results are presented in table 10. For the most part, statistical results confirm what could be observed directly from figure 3. In the first replication, the separation between taskloads B and C for the first two time blocks was not significant. Also between A and B in the fourth block, the mean NFLT's were not significantly different. The reader will note a reversal between B and C in terms of NFLT for the third time block. In the second replication, all paired comparisons were significantly different including the reversal between B and A in the fourth block. The reader may want to remember the idiosyncracies of the NFLT variable — especially the reversals when examining the workload (ATWIT) data which follow. Also note that the correlation between the systems variables NIAC and workload responses WIAV was r =.776, while the correlation between NFLT and workload (WIAV) was r =.649. Note that correlation measures the degree to which variables covary in relation to their internal variance. Correlations range from -1 to 1, with the measured relationship becoming stronger as it approaches either end of the continuum.

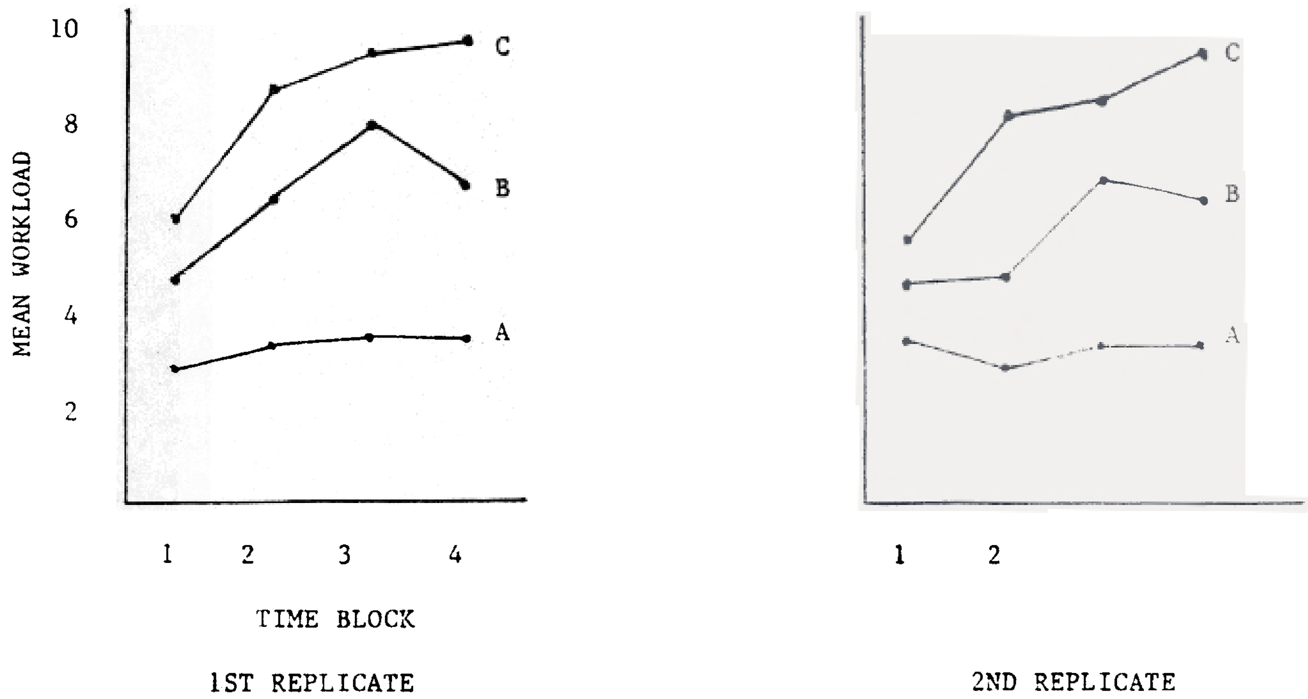


FIGURE 3. PLOTTED MEAN WORKLOAD (ATWIT) RESPONSES

TABLE 9 SUMMARY OF ANALYSIS OF VARIANCE ON NUMBER OF FLIGHTS HANDLED

Source of Variance	Degrees of Freedom	F Ratio	Correlation Ratio in %
Task Load (A)	2,18	401.56**	70.16
Time Block (B)	3,27	43.50**	3.64
A X B Interaction	6,54	58.88**	14.55
Replication (C)	1,9	1.30	0.01
A X C Interaction	2,18	20.91**	0.94
B X C Interaction	3,27	3.33**	0.32
A B C Interaction	6,54	0.88**	0.19

**P ≤.01

NOTE: Correlation between NFLT and WIAV was .649

TABLE 10 PAIRED COMPARISONS OF NFLT ACROSS TASKLOADS BY TIME BLOCK

First Replication

Time Block:															
Taskload:	A	B	C		A	B	C		A	B	C		A	B	
Mean:	7.9	15.0	15.8		5.2	13.7	14.8		6.3	14.6	15.8		8.9	9.3	16.3
7.9		7.1	7.9	5.2		8.5	9.6	6.3		8.3	9.5	8.9		0.4#	7.4
15.0			0.8#	13.7			1.1#	14.6			1.2	9.3			7.0
15.8				14.8				15.8				16.3			

Second Replication

	Time Block:		1				2				3				4	
	Taskload:	A	B	C		A	B	C		A	B	C		A	B	C
	Mean:	8.0	14.9	17.1		5.0	11.9	14.6		6.1	3.9	15.5			10.0	17.2
	8.0		6.9	9.1	5.0		6.9	9.6	6.1			9.4	8.2		8	9.0
B	14.9			2.2	11.9			2.7	13.9			1.6	10.0			
C	17.1				14.6				15.5				17.2			

$P \leq .01$

Differences between pairs are not significant

Critical Values

Ordered Steps

1	2
.172	1.330

ANALYSIS OF AIR TRAFFIC WORKLOAD INPUT TECHNIQUE.

Participant controllers were asked to report how hard they were working every minute during each control run. They were instructed to press (as soon as possible) one switch button out of a choice of ten in a specially constructed response box when the clock at the bottom of their radar display began blinking. The buttons were numbered from left to right with the numerals 1 to 10. These numbers were anchored verbally during the briefing so that the controllers response would indicate workload which ranged from 1 — very easy (all tasks easily completed) to 10 — very hard (some tasks left incompleted). The instructions are presented in appendix A.

While controllers showed some initial reluctance to use the ATWIT system, they all managed to integrate it into their behavioral repetoire during training. The task did add somewhat to their workload; but during interviews, they indicated that ATWIT responses became easier to accomplish with training. Most indicated that if they missed a response, it was because they were completely occupied. In light of such comments and previous experience with a similar system for aircrew research, missed responses were recorded as maximum workload responses (Rosenberg, Rehmann, and Stein, 1982; Stein and Rosenberg, 1983).

Since the primary purpose of this research study was to investigate the relationship between what was happening in the air traffic system and human operator workload, timely and accurate estimates of workload were essential. The fact that controllers were willing and able to use the ATWIT system is an important result in and of itself. Better still are the facts which follow.

ATWIT responses were coded as WIAV in the 15-minute data set. This WIAV variable was correlated against an artificial variable created by recoding taskloads A, B and C as quantitative values 1, 2, and 3. The resultant correlation of r =.79 indicated a moderate positive relationship with higher workload response corresponding to higher taskloads. The reader should recall from the design section that controllers received taskloads in what probably appeared to them as a jumbled order based on the counter-balance design. They did not know what to expect, and yet it appears that there was a relationship between workload and taskload.

Statistical purists might criticize the application of ANOVA techniques to subjective scale responses because of the nature of the data. However, there has been a growing precedence for such applications. ANOVA is robust with respect to many of its technical assumptions. The applications which will follow should be considered as indicative rather than conclusive. They will be confirmed by other multivariate techniques as we proceed. Whenever possible ANOVA will be accomplished using Greenhouse-Geisser probabilities, which make it a considerably more conservative test.

Table 11 is a summary of the mean real-time workload responses. An examination of this table indicates that, despite some overlaps previously noted in NIAC and NFLT, there appear to have been three levels of workload achieved. This is even clearer if one looks again at figure 1 where the mean responses are plotted. Comparison of figures 1 with 2 and 3 is interesting. The pattern of figure 1 is more similar to figure 2 than to figure 3, which explains in part the difference in the two reported correlations (.776 versus .649).

TABLE 11. MEAN WORKLOAD RESPONSES (ATWIT)

	First Replication					Second Replication					
	Time Block					Time Block					
Task Load	1	2	3	4	Run Mean	1	2	3	4	Run Mean	Overall Mean
A	2.82	3.07	3.31	3.17	3.10	3.31	2.96	3.22	3.24	3.18	3.14
B	4.40	6.13	7.73	6.48	6.19	4.53	4.81	6.73	6.23	5.58	5.88
C	5.81	8.43	9.19	9.50	8.23	5.49	8.00	8.29	9.33	7.78	8.01
Time Block Means											
	4.35	5.88	6.74	6.38		4.45	5.26	6.08	6.27		

An ANOVA was applied to the ATWIT data summarized in table 11. The results of this analysis appear in table 12. Both taskload and time block produced significant main effects ($P \leq .01$). These were complicated by several interactions (taskload x time) and (time x replication), neither of which represented much accountable variance as indicated by low correlation ratios.

An analysis of simple main effects of taskload on workload responses at specific levels of time is summarized in table 13. Again, all the simple main effects were significant. Additional post-hoc analysis was required to determine if paired comparisons were also significant (table 14). Results indicated that in the first replicate all pairs of means were significantly different from each other. This finding was also true for all time blocks in the second replicate except the first block where only taskload C had significantly more reported workload than taskload A. Note that the only paired differences in workload which were not significant did not correspond to overlaps in either NIAC or NFLT. This indicated that there is more to the workload response than these two systems variables. We could tediously analyze all systems variables as was done for NIAC and NFLT, and probably we would find more than one with overlaps in the first time block of the second replicate. That would not clearly explain human response to taskload. We will attempt to do it another way in a later section using multilinear regression.

TABLE 12. ANOVA SUMMARY FOR ATWIT RESPONSES

Source of Variance	Degrees of Freedom	F Ratio	Correlation Ratio in %
Task Load (A)	2,18	138.50**	68.11
Time Block (B)	3,27	107.90**	11.22
A x B Interaction	6,54	22.88**	6.47
Replication (C)	1,9	7.39	0.45
A x C Interaction	2,18	1.51**	0.38
B x C Interaction	3,27	5.11**	0.45
A x B x C Interaction	6,54	1.63	0.2

**P $\leq$.01

TABLE 13. SIMPLE MAIN EFFECTS ON WIAV OF TASKLOAD AT LEVELS OF TIME BLOCK AND REPLICATION

	Time Block	Mean Square	F
First Replicate	1	22.32	15.80**
	2	72.32	51.20**
	3	93.57	66.07**
	4	100.14	70.89**
Second Replicate	1	11.94	8.45**
	2	64.98	46.00**
	3	67.53	47.81**
	4	92.83	65.72**

**F critical (2,60) (P $\leq$.01 = 4.98

Error term MS within cell 1.412

'ON. RO: AD BLOCI

F rst Rep cation

1			2		3					
A	B	C	B	C	A	B	C	A	B	C
										50

Second Replication

			2		3			4		
A	B	C	B	C	A	B	C	A	B	C

Ordered Steps

2

The differences in workload across time blocks were interesting as indicated in figure 1. An analysis of the simple main effects of time block at specific taskloads was computed (table 15). In both replications, the variability of the workload responses across time for taskload A were not significant (P >.01). For taskloads B and C, the influence of time was quite strong. Post-hoc testing for paired comparisons was required for only the latter two taskloads (table 16). Note that taskload B had a reversal in workload responses between the third and fourth time blocks. In taskload B, there was a steady increase in workload in the first replication from the first through the third time block, then a significant decrease in the fourth. In the second replication, workload was steady for two blocks, then increased in the third, but did not drop off significantly in the fourth block. In taskload C, first replication, there was a significant increase in reported workload from the first to second time blocks, and there were nonsignificant changes from the second to the third and from the third to the fourth blocks. In the second replication, the pattern was similar except that workload did increase from the third to the fourth blocks.

TABLE 15. SIMPLE MAIN EFFECTS OF TIME AT LEVELS OF TASKLOAD AND REPLICATION

	Taskload	Mean Square	F
First Replication	A	0.422	0.938
	B	18.9108	42.060**
	C	28.0345	62.350**
Second Replication	A	0.23778	0.529
	B	11.46767	25.506**
		26.51199	58.967**

F critical (d,F = 3,81)

**(P .01) = 4.07

These ebbs, plateaus, and flows of reported workload would have been completely missed by traditional post-run questionnaire techniques. Participant controllers demonstrated a willingness and an ability to make workload ratings during primary task performance without any apparent decrement in the control of air traffic. Another approach for estimating controller workload in real-time is to ask someone else to observe the controllers activities and provide a rating.

TABLE 16. PAIRED COMPARISONS OF MEAN WORKLOAD RESPONSES ACROSS TIME BLOCKS

First Replication

Time Block:	1	2	4	3
Mean:	4.40	6.13	6.48	7.73
4.40		1.73**	2.08**	3.33**
6.13			0.35	1.60**
6.48				1.25**
7.73				

Time Block:	1	2	3	4
Mean:	5.81	8.43	9.19	9.50
5.81		2.62**	3.38**	3.69**
8.43			0.76	1.07**
9.19				0.31
9.50				

Second Replication

Time Block:	1	2	4	3
Mean:	4.53	4.81	6.23	6.73
4.53		0.28	1.70**	2.20**
4.81			1.42**	1.92**
6.23				0.50
6.73				

Time Block:	1	2	3	4
Mean:	5.49	8.00	8.29	9.33
5.49		2.51**	2.80**	3.83**
8.00			0.29	1.33**
8.29				1.04
9.33				

**$P \leq .01$

Critical Values

Ordered Steps:	2	3	4
	0.793	0.903	0.962

OBSERVER RATING.

Throughout the course of this project, two air traffic control specialists from the FAA Technical Center acted as observer/evaluators during all simulation. They participated from the ground level up, even to the point of helping to develop the measurement scales which they were to use to form their ratings. The observers sat on either side of the controller and rated every 15 minutes on three 10-point scales: workload, busyness, and effectiveness. They had to maintain constant attention to controller behavior and to the radar display. This was done for over 100 hours of simulation during a 10-week period.

The first question that should be asked about any measurement system is, is it reliable? By reliability, we generally mean internal consistency and the tendency to measure the same kinds of things over time and space. In this experiment, the observers and their abilities to assign numbers based on what occurred were our measurement system. A principal way of estimating reliability in the design cited here is to examine the degree to which the two raters tended to agree with each other or in statistical terms, the degree to which they covaried. A way of measuring this is the correlation coefficient.

The median correlation between the observer ratings for the ten participants was r =.91 with a range from .72 to .97 (table 17). This indicated that both observers were tuned to the same kinds of behavior when rating, and that the rating system was reasonably reliable.

Observer ratings on the three scales of workload, busyness, and effectiveness across the 60 data runs were subjected to a factor analysis. This is a statistical technique which looks at the relationship between variables and tries to explain it in simpler terms. The results of this analysis indicated four factors.

TABLE 17. INTERRATER RELIABILITY

Participant Code	Interrater Correlation
13	0.82
18	0.94
14	0.73
15	0.93
12	0.94
16	0.97
11	0.90
20	0.72
19	0.90
17	0.95

On three of these factors the workload and busyness scales loaded together for each of the three taskloads A, B, and C. Effectiveness formed a factor of its own with ratings from all three taskloads. This meant that the observers viewed workload and busyness as virtually identical measures but took a unique view at each taskload. Their view of effectiveness was more consistent across the three taskloads, which may have been partially a function of their reluctance to down-rate the effectiveness of brother controllers.

Despite this reluctance, there was still enough variability in the effectiveness ratings to produce a significant inverse correlation with the observers ratings of workload (r = -.55). These same ratings of workload correlated very well and positively with the taskload r =.91. A summary of the mean workload and effectiveness ratings is presented in table 18. Note that busyness is not presented since it behaved so similarly to the workload scale (correlation r =.97).

Table 18 represents the means of the summed ratings for the two observers. The ranges of these summated scores were actually from 2 to 20 rather than 1 to 10. A quick examination of this table makes evident the strong positive relationship between workload and taskload and the somewhat weaker inverse relationship between effectiveness and workload.

TABLE 18. MEAN OBSERVER RATINGS

Taskload	Workload	Effectiveness
1. A	4.31 (1.41)	19.20 (1.08)
2. B	12.08 (2.37)	18.41 (1.29)
3. C	16.19 (1.67)	17.21 (1.85)

Note: Standard deviations are shown in parentheses

Observer workload estimates were strongly related to the self-ratings made by participant controllers themselves. Workload responses made during actual control of simulated traffic and cumulated over the control period were correlated against observer ratings, r =.86. Observer ratings also correlated well (r =.93) against participants post-run questionnaire workload responses to be discussed in more detail later.

What kinds of system activity or controller behavior might observers have been noting and integrating into their responses? The Sigma 8 computer sampled and stored a wide variety of systems variables. Those which correlated best with observer ratings are presented in table 19.

TABLE 19 SYSTEM/CONTROLLER VARIABLES WHICH WERE RELATED TO OBSERVER RATINGS

Variable	Correlation
Instantaneous Aircraft Count	0.87
Number of Flights Handled	
Duration of Flights	0.84
Total Distance Travelled	0.86
Number of Ground-to-Air Communications	0.85
Number of Air-to-Ground Communications	0.85
Average Density of Aircraft in the Controlled Airspace	

Most seasoned controllers would state that the impact of such variables on workload is only common sense. However, much of research involves documenting what appears to be common sense. When results confirm what user groups already think they know, then the results have face validity.

BIVARIATE RELATIONSHIPS

A bivariate relationship is a description of how two variables covary together taking into consideration the variability within each. This is, of course, a definition of correlation. Computing all the relationships between all the variables in the data set can be a handy way of taking a first look at what is available. It also provides an indication of which systems variables might be the best univariate (one variable as compared to another variable) predictors of controller workload. If a system variable such as NIAC correlates at one point in time with workload measures, it might very well serve to predict workload at the next point in time. It should be remembered that workload here is viewed as an operator response to taskload and not as a synonym for taskload. However, we have already demonstrated that NIAC, a systems variable is related to taskload and also to workload. There must be overlap between taskload and workload statistically, or the whole purpose of this research design becomes meaningless.

Table 20 tells the story of all bivariate relationships. It does, in fact, tell more stories than many will ever want to hear. Taking a very conservative view of correlations and their significance from zero, those greater than or equal to r =.632 (8 degrees of freedom) should be considered significant.

The correlations in table 20 were computed using the 15-minute data set. This should be remembered since correlations based on the 60-minute cumulative data set tend to be inflated, due to less internal variance within the variables. A similar table of the 60-minute correlations is available in the appendix. Table 20, however, is probably a better estimate of the relationships between these variables.

WORKLOAD PROBE - 15 MINUTE DATA SET. BMDP8D

COMPLETE CORRELATION MATRIX

		SAMPLE 3	REPLICAT 5	ALT 6	HDG 7	SPD 8	RTE 9	HOLD 10	CMTR 11	HOIN 13	HOID
SAMPLE	3	1.0000									
REPLICAT	5	.0000	1.0000								
ALT	6	.5522	-.0411	1.0000							
HDG	7	.6459	-.0626	.4389	1.0000						
SPD	8	.0835	-.0511	.1446	.0729	1.0000					
RTE	9	.5049	-.0921	.3475	.8314	.0770	1.0000				
HOLD	10	.0964	-.0962	-.0128	.1387	.0308	.0690	1.0000			
CMTR	11	.5797	-.0060	.4477	.3857	.0753	.2874	.0325	.0000		
HOIN	13	.7962	.0203	.5196	.5130	-.0297	.4011	-.0488	.4302	1.0000	
HOID	14	.6627	-.1022	.3092	.5095	.0324	.3397	.2519	.6571	.4566	1.0000
HOUT	15	.6014	-.0059	.4776	.4507	.0940	.3693	.0237	.9188	.4511	.6194
NIAC	16	.8565	.0148	.5500	.5873	.0845	.4783	.0220	.7572	.7511	.5716
NFLT	17	.8291	-.0119	.5401	.5901	.0574	.4673	.0040	.4961	.9441	.5299
DFLT	18	.7907	.0205	.5487	.5316	.0805	.4256	-.0099	.8589	.6717	.6125
DIST	19	.8118	-.0159	.5760	.5623	.0806	.4530	-.0118	.8265	.7450	.6071
FUEL	20	.7640	.0348	.6641	.5084	.1407	.4165	-.0408	.7892	.7055	.5582
DINB	21	.7377	.0442	.5086	.4671	.0630	.3541	-.0280	.8652	.6242	.5728
NG2G	22	.3046	-.1558	.3382	.3643	.2390	.2950	.2088	.2917	.1176	.5007
DG2G	23	.1716	-.1909	.1785	.2770	.0755	.2036	.3023	.1448	.0212	.4013
NG2A	24	.7344	-.0754	.7278	.6913	.1591	.5428	.1049	.7530	.6663	.6421
DG2A	25	.7032	-.1073	.7026	.6655	.1695	.5033	.1592	.6685	.5522	.5970
NAIR	26	.7976	-.0121	.7487	.7109	.1559	.5690	.0189	.7260	.7604	.5432
DAIR	27	.8161	-.0190	.7183	.6509	.1170	.5033	-.0038	.6580	.8262	.5169
NDLY	28	.3357	-.0727	.1449	.4288	.0159	.2962	.2329	.2991	.1213	.4963
DDLY	29	.3387	-.1291	.0998	.3937	.0307	.3670	.3337	.2474	.1459	.3976
N5CF	30	.3770	-.0690	.3391	.3346	.1872	.2576	-.0193	.3044	.3852	.2896
D5CF	31	.2805	-.1297	.2906	.1804	.1671	.1628	-.0427	.2686	.2976	.1304
CKEN	36	.6125	.0475	.5632	.5049	-.0088	.4155	-.0564	.6060	.5556	.4205
CKER	37	.4947	-.0991	.2672	.2712	-.0182	.1997	.0383	.4096	.3445	.5372
WIAV	39	.7912	-.0645	.5002	.6044	.1180	.4823	.1099	.7316	.5610	.7460
WDAV	42	.5967	-.0936	.2273	.3047	.0991	.2223	.1230	.5227	.4172	.6400
CMAV	44	.8826	.0067	.5406	.5207	.1002	.4083	.0592	.6563	.7083	.6293
OBQ1	46	.9219	-.0840	.6155	.6416	.1193	.4925	.0774	.6787	.7718	.7091
OBQ2	47	.8938	-.0878	.6378	.6062	.1289	.4330	.0979	.6426	.7587	.6666
OBQ4	48	-.4908	.2420	-.3028	-.4123	-.0052	-.3643	-.1535	.3023	-.2833	-.5362

TABLE 20. BIVARIATE CORRELATIONS — 15 MINUTE DATA (of 3)

WORKLOAD PROBE 15 MINUTE DATA SET. BMDPRD

		HOUT 15	NIAC 16	NFLT 17	DFLT 18	DIST 19	FUEL 20	DIN3 21	NG2G 22	DG2G	NG2A 24
HOUT	15	.0000									
NIAC	16	.8013	1.0000								
NFLT	17	.5347	.7988	1.0000							
DFLT	18	.8966	.9525	.7259	1.0000						
DIST	19	.8721	.9536	.7979	.9828	1.0000					
FUEL	20	.8236	.8834	.7453	.9317	.9311	1.0000				
DINB	21	.8767	.9236	.6658	.9766	.9485	.8944	1.0000			
NG2G	22	.2987	.2602	.2341	.2757	.2703	.3002	.2395	1.0000		
DG2G	23	.1410	.1244	.1276	.0998	.0976	.1073	.0678	.7012	1.0000	
NG2A	24	.7750	.8207	.7445	.8298	.8512	.8362	.7727	.5355	.3439	1.0000
DG2A	25	.6860	.7227	.6368	.7249	.7396	.7173	.6718	.5894	.4515	.9171
NAIR	26	.7465	.8502	.7988	.8444	.8718	.8553	.7922	.3295	.1501	.9298
DAIR	27	.6672	.8336	.8502	.8053	.8501	.8246	.7507	.2751	.1333	.8909
NDLY	28	.3037	.2672	.2252	.2915	.2802	.2445	.2311	.3286	.2538	.3610
DDLY	29	.2693	.2637	.2147	.2706	.2650	.2157	.1932	.3060	.1938	.3454
N5CF	30	.3470	.4317	.4248	.4225	.4402	.4894	.4121	.1624	.0660	.4004
D5CF	31	.2944	.3903	.3130	.3563	.3620	.3765	.3654	.0368	-.0207	.3264
CKEN	36	.6338	.6638	.5717	.6895	.6799	.7316	.6766	.1189	.0036	.6224
CKER	37	.3702	.4137	.4153	.4372	.4523	.4050	.4150	.2547	.1813	.4502
WI4V	39	.7459	.7756	.6492	.7977	.7905	.7578	.7510	.4245	.2771	.8096
WDAV	42	.5281	.5417	.4966	.5562	.5558	.4979	.5090	.2612	.2359	.5028
CMAV	44	.6803	.8632	.7316	.8608	.8581	.8377	.8394	.3060	.1397	.7702
OBQ1	46	.6936	.8690	.8327	.8369	.8563	.8356	.7924	.4060	.2483	.8494
OBQ2	47	.6535	.8449	.8131	.8061	.8229	.8182	.7682	.4034	.2389	.8285
OBQ4	48	.3385	-.3639	-.3859	-.3414	-.3729	-.3275	-.2623	-.4021	-.4086	-.4439

TABLE 20 BIVARIATE CORRELATIONS 5 MINUTE DATA (2 of 3)

WORKLOAD PROBE - 15 MINUTE DATA SET. BMDP8D

		DG2A 25	NAIR 26	DAIR 27	NDLY 28	DDLY	N5CF 30	D5CF 31	KEN 36	CKER 37	WIAV 39
DG2A	25	1.0000									
NAIR	26	.8403	1.0000								
DAIR	27	.8047	.9574	1.0000							
NDLY	28	.3476	.2666	.2116	1.0000						
DDLY	29	.3512	.2683	.2251	.8018	1.0000					
N5CF	30	.3369	.4055	.4074	.0361	.0731	1.0000				
D5CF	31	.2984	.3445	.3574	-.0140	-.0025	.5437	1.0000			
CKEN	36	.4890	.6937	.6497	.1269	.0977	.3515	.2706	1.0000		
CKER	37	.3989	.3990	.3974	.3295	.2990	.2006	.0992	.1805	1.0000	
WIAV	39	.7817	.7536	.7233	.4027	.3891	.3838	.2550	.6158	.5100	1.0000
WDAV	42	.5144	.4634	.4685	.3032	.3319	.2402	.1481	.3256	.4401	.7259
CMAV	44	.6807	.7533	.7495	.3093	.2385	.4642	.3929	.6021	.4730	.7769
OBQ1	46	.7814	.8433	.8390	.3593	.3469	.4438	.3329	.6561	.5425	.8616
OBQ2	47	.7798	.8236	.8291	.3317	.3304	.4348	.3269	.6449	.5043	.8326
OBQ4	48	-.4958	-.3544	-.3321	-.4466	-.4235	-.1746	-.0761	-.2183	-.4325	-.5129

		WDAV 42	CMAV 44	OBQ1 46	OBQ2 47	OBQ4 48
WDAV	42	1.0000				
CMAV	44	.5584	1.0000			
OBQ1	46	.6305	.8815	1.0000		
OBQ2	47	.6048	.8580	.9736	1.0000	
OBQ4	48	-.4318	-.3946	-.5496	-.4993	1.0000

NUMBER OF INTEGER WORDS OF STORAGE USED IN PRECEDING PROBLEM 2126
CPU TIME USED 74.808 SECONDS

TABLE 20 BIVARIATE CORRELATIONS 15 MINUTE DATA (3 of 3)

A great deal of information can be drawn from this table. An examination of the correlations between all variables and the "sample" column indicates which variables were most closely related to taskload. Sample was an artifically developed numerical variable created by recoding taskloads A, B, and C as 1, 2, and 3, respectively. It was based on the assumption that three taskloads were produced by the manipulation of the independent variables (NIAC, Restricted Areas, etc.).

Two variables, SPD and HOLD, were literally unrelated to sample, so little could be expected of them as systems drivers of workload. Neither one of them is related to either WIAV or OBQ1. The relationship between taskload (if you accept three trichotomized levels) and the magnitude of the systems variables was so pervasive that the computer correlations could be used as estimators of the relationships between systems variables and workload estimates. "Sample" was related to a sizable number of systems and workload variables. The real-time workload variables included WIAV, WDAV, OBQ1, and OBQ3. Most of the post-run questionnaire items were also closely related to both taskload (sample) and the real-time workload variables (see 60-minute correlation matrix in the appendix B).

Another column in the matrix, which calls for special mention, is the second one labeled replicate. Here were reported the correlations of all variables against another artifical variable: a dichotomy representing the first (1) and second (2) replications of three data collection runs under taskload A, B and C. If the training procedure had not been successful, then skill-learning should have taken place during the first three data runs and would have shown up as a shift in the second three runs. This would have been roughly indicated by a sizable correlation between replication order and the systems variable. This would mirror a significant main effect of replication if ANOVA had been computed. The small, nonsignificant correlations in the replicate column speak for themselves. The reader will recall that there was no significant replication effect for the two representative systems variables (NIAC and NFLT) and the workload variable (WIAV) which were analyzed in detail using ANOVA.

The correlations between systems variables and workload are highly relevant to the purposes of this study. Those which were the strongest (i.e., exceed r =.70) are reproduced in table 21. The fact that the relationships between these systems variables and workload reported by controllers are strong is no great surprise. It makes sense that the more aircraft controlled and the more control activity required for each aircraft, the harder a motivated operator will have to work to maintain some standard of performance.

Reporting the tabulated values of the bivariate relationships is only a "first cut," however. Looking back at the goals of this experiment, what is important is the possible existence of a workable multivariate workload prediction system.

While all the variables in table 21 correlate at a moderate level with workload (WIAV), there is also a great deal of redundancy between them. They correlate more or less with each other, which means that a simple linear combination would not necessarily be an effective way to build a prediction model. To do this, multivariate regression methods may be used to reduce predictor redundancy and improve prediction efficiency.

TABLE 21. STRONGEST CORRELATIONS WITH WORKLOAD

System Variable	Correlation $\geq$ 0.70
CMTR	0.73
HOID	0.75
HOUT	0.75
NIAC	0.78
DFLT	0.80
DIST	0.79
FUEL	0.76
DINB	0.75
	0.81
	0.78
	0.75
	0.75

Before reviewing the efforts to build the best prediction model, the reader may wish to examine the scatterplots located in appendix (B). These plots were based on the 60-minute cumulative data and describe all systems and other workload variables plotted against the real-time workload responses (WIAV) of the controllers. The letters A, B, and C represent the three taskloads respectively. The plot routine will only place one letter at a single point. If the data runs with the same taskload should be called for at one point, only one point would appear. If they had come from the runs of two different taskloads, then an asterisk (*) would appear.

These plots show that considerable variability existed on both the workload and systems variables. Since all participants are plotted, the scatter is a result of inter-and-intra person variations in perceived workload coupled with the dynamic nature of the simulation. The variance generated by people was a function of perceived workload (generated by manipulation of the independent variables) influenced by what controllers brought to the situation in terms of skills, knowledge, abilities, and motivation.

We have examined the bivariate relationships with a focus on workload. We now turn to a multivariate view of workload closing with one last analysis. The question was posed: Which of the workload variables, both real-time and post-run, would best descriminate across the three levels of taskload — A, B, and C? Using the sample variable of 1, 2, and 3 in place of taskload, a descriminate function analysis (60-minute data base) was computed using the following variables as predictors" (1) WIAV, (2) WDAV, (3) OBQ1, (4) OBQ2, (5) OBQ4 and (6) the five post-run questionnaire items. Descriminate function analysis is a special case of multilinear regression which will be defined/explained in the next section. The results were clear. The best predictor which separates the three taskloads most clearly was OBQ1 — the pooled observer estimate of controller workload. Of the 60 data runs, 58 were correctly classified as taskloads A, B, and C. It should not be forgotten, however, the observers knew which missions were being run and when; and also they could see participant ATWIT responses — so there was more than a little confounding in the result. It should be viewed as interesting, perhaps indicative, but certainly not conclusive.

REGRESSING ON WORKLOAD.

Multiple regression techniques attempt to define the best linear fit between a series of predictor variables and a criterion or dependent variable. Stepwise regression examines not only the relationship of predictors to criterion, but also how predictors are related to each other. The stepping algorithms only select those predictors which account best for criterion variance and overlap each other the least. The stepwise regression therefore discards predictors like old shoes and just keeps the best.

We already know that we can account for a finite portion of controller reported workload (60 to 65 percent for NIAC and NG2A), using bivariate relationships. The question is whether or not a multivariate prediction system could do any better. In this section, we will examine regressions against both controller self-ratings of workload (WIAV) and the observers estimates (OBQ1).

This project was initiated to examine the potential of a select number of variables suggested by Mitre Corporation under their construct of Workload Probe. These variables included four categories of systems activities: (1) controller actions, (2) instantaneous aircraft count, (3) number of unresolved conflicts, and (4) complexity (clustering). The second and third categories were straightforward and univariate. The first category called for the combination of a number of systems variables, but the combinational model was never made clear, so we elected to treat the components of "controller actions" as individual predictors. The fourth category, complexity, was not defined mathematically until well after this project was underway. The objective of complexity appears to be a measure of aircraft local density. We developed our own definition which follows. It differs somewhat from Mitre's mathematically but not in terms of intent.

$a_i(t)$ = number of aircraft within a 10-nautical-mile radius (3D) of aircraft i at time t $\quad i=1,\ldots,n.$

$$t=1,\ldots,m.$$

(CMAV) complexity = $\bar{a}$ (t)

$$\bar{a} = \sum_{i=1}^{n} \sum_{t=1}^{m} \frac{ai(t)}{nm}$$

$$\sigma^2 = \sum_{i=1}^{n} \sum_{t=1}^{m} \frac{(ai(t)-\bar{a})^2}{mn-m-n+1}$$

The list of variables to be examined with regression analysis coming from workload probe was as follows:

ALT	Altitude Change
CMAV	"Complexity" (Local Density)
DDLY	Duration of Holds and Turns $\geq$ 100 Seconds
D5CF	Duration of Conflicts
HDG	Heading Change
HOIN	Inbound Handoff
HOLD	Hold Frequency
HOUT	Outbound Handoff
NDLY	Frequency of Holds and Turns $\geq$ 100 Seconds
NIAC	Instantaneous Aircraft Count
N5CF	Number of Conflicts
RTE	Route Change
SPD	Speed Change

The regressions reported in this section are presented in table form. Each table shows the variables in the regression in the order in which they were stepped in. It does not show the variables which were discarded. The tables provide the squared multiple correlation R^2 achieved at each step, with the highest R^2 at the bottom of each table representing the results of the complete process. R^2 has been referred to as the coefficient of determination. It provides an estimate of the proportion of variability in the workload criterion variable which is accounted for by the weighted combination of the selected predictors. Regression weights and the results of the ANOVA on the regression are also described. The ANOVA on the regression tests whether the predictors can account for a significant proportion of variance in the criterion. As will be seen shortly, all the regressions reported here were significant.

The first regression against controller reported workload (WIAV) was accomplished using Mitre suggested variables (table 22). Three variables entered the regression equation, CMAV, HOUT, and HDG with a multiple R of .85, which was a slight improvement over the best bivariate relationship. Of the three variables included, two came under the category of controller actions and the third was our computation of complexity (CMAV). NIAC, the instantaneous aircraft count, was conspicuous by its absense. Stepwise regression does not weight variables based on previous popularity or face validity. NIAC is well correlated with other systems variables, so it often does not appear as a predictor in these equations.

A second analysis was completed on the Mitre predictors using only the second and third time blocks of all the data runs (table 23). It was thought that this might eliminate noise from the traffic buildup in block 1 and the fatique/letdown from block 4. The multiple R^2 did not change appreciably, increasing only 1 percent. A feature of the software package, BMDP1R, was employed to test for differences in the regressions for time blocks two and three (to see if pooling them was valid). The computed ANOVA failed to indicate that they were different, a desirable result.

TABLE 22. STEPWISE REGRESSION OF MITRE VARIABLES ON WIAV

Variables Selected	Multiple R	R^2	Regression Coefficents	F On Regression
CMAV	.78	.60	2.91	F(3,236)=210*
HOUT	.83	.69	.215	
HDG	.85	.73	.097	

*P .05

TABLE 23 STEPWISE REGRESSION OF MITRE VARIABLES ON WIAV (TIME BLOCKS 2 & 3 ONLY

Variables Selected	Multiple R	R^2	Regression Coefficents	F On Agression
CMAV	.82	.68	4.620	F(3,116)=110.70*
HDG	.85	.73	.091	
HOUT	.86	.74	.140	

*P $\leq$.05

NOTE: Multilinear regressions were computed independently on time blocks 2 and 3 using the 3 variables produced by the stepwise regression. An ANOVA on regression coefficients across the two time blocks produced an F(4,112)=0.929 failing to reject the hypothesis that the slopes and/or intercepts did not differ between the time blocks. This justified the pooled regression above. An attempt to force this regression to step further failed to increase multiple R^2.

The two regressions reported so far involve only the Mitre variables. There were a sizable number of variables available which were not included in this workload probe construct. The next two analyses will examine regressions using the entire pool of variables available (see table 3).

A stepwise regression analysis on WIAV was computed using all available system variables. Some of which are directly related to controller behavior. The results are displayed in table 24. Six variables were stepped into the regression equation with a final multiple R^2 of .815, which represented an 8 percent increase in accountable variance over the results generated using Mitre suggested variables. Table 25 provides the results of the regression produced using time blocks 2 and 3 only. This provided a 4 percent increase in accountable variance over the regression computed using all data. Both of these analyses have produced multiple R and R^2 which are relatively high. Based on the ANOVA on the regressions, a significant amount of criterion variance (workload responses) is being accounted for using systems variables as "predictors."

TABLE 24. STEPWISE REGRESSION OF ALL VARIABLES ON WIAV (ALL DATA)

Variables Selected	Multiple R	R^2	Regression Coefficents	F On Regression
HOID	.862	.743	.001	F(6,223)=171.31**
DG2A	.887	.788	.019	
CKEN	.893	.797	.050	
ALT	.896	.804	-.106	
CKER	.901	.811	.094	
CMAV	.903	.815	1.558	

**P $\leq$.01

TABLE 25. STEPWISE REGRESSION OF ALL VARIABLES ON WIAV (TIME BLOCKS 2 and 3 ONLY)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
HOID	.883	.780	.008	F(6,113)=109.60
CMAV	.896	.802	2.519	
DG2A	.901	.812	.020	
CKEN	.905	.819	.048	
ALT	.916	.839	-.162	
CKER	.923	.853	.144	

NOTE: An ANOVA on the regression coefficients for time blocks 2 and 3 respectively produced an F(7,106)=1.11 failing to reject the hypothesis that the slopes and/or intercepts did not differ between the time blocks.

Now we shitt focus to an examination of regression equations developed using OBQ1 (observer estimates of workload), as the criterion or dependent variable.

Table 26 describes the results of Mitre suggested variables which were regressed on OBQ1 data collected during all four time blocks. This produced a very creditable fit with 86.7 percent accountable variability. The instantaneous aircraft count (NIAC) made it into this regression. This was very likely a variable to which the observers, being controllers themselves, paid close attention. Recalling table 19, NIAC correlated r =.87 with observer ratings. Table 27 provides the results for the regression on time blocks 2 and 3 only. The analysis shows a slight increase in R^2. However, the ANOVA on the regression coefficients (when their regressions were computed separately) rejected the hypothesis that their slopes and intercepts were the same. Their multiple R^2 were close, but it is possible that data trends were somewhat different across these blocks for OBQ1.

TABLE 26. STEPWISE REGRESSION OF MITRE VARIABLES ON OBQ1 (ALL DATA)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.881	.777	6.546	F(5,234)=305.54*
NIAC	.907	.823	.421	
HDG	.920	.847	.139	
HOIN	.928	.860	.218	
ALT	.931	.867	.133	

*P 05

TABLE 27. STEPWISE REGRESSION OF MITRE VARIABLES ON OBQ1 (TIME BLOCKS 2 and 3 ONLY)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.899	.809	7.846	F(4,115)=229.95**
HOUT	.930	.865	.324	
HDG	.937	.879	.138	
HOIN	.943	.889	.316	

**P <.01

NOTE: An ANOVA on the regression coefficients for time blocks 2 and 3 produced an F(5,110)=4.05** which rejected the hypothesis that the slopes and/or intercepts did not differ between the time blocks.

Tables 28 and 29 represent the results of regressions using all available systems variables on OBQ1. Using all time blocks, six variables stepped into the equation with a resultant multiple R^2 of .911. NIAC did not appear in this regression. This was probably a function of its relatively high relationship to other system activity variables; and again, the cold eye of the regression model could not consider entering a variable based on face validity. Table 29 shows the five variables which were stepped in against OBQ1 for time blocks 2 and 3 only. The multiple R^2 of .936 was to be the highest of any multivariate relationship analysed for the controller workload project.

TABLE 28. STEPWISE REGRESSION OF ALL VARIABLES ON OBQ1 (ALL DATA)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.881	.777	5.157	F(6,223)=398.12**
NFLT	.924	.853	.347	
DG2A	.941	.885	.017	
HOID	.948	.899	.001	
CKEN	.951	.905	.053	
CKER	.955	.911	.185	

**P $\leq$.01

TABLE 29. STEPWISE REGRESSION OF ALL VARIABLES ON OBQ1 (TIME BLOCKS 2 & 3 ONLY)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
NFLT	.924	.854	.450	F(5,114)=335.16**
HOID	.949	.901	.001	
NG2A	.959	.920	.090	
CMAV	.965	.931	4.143	
CKER	.968	.936	.182	

**P $\leq$.01

NOTE: An ANOVA on the regression coefficients for time blocks 2 and 3, respectively, produced an F(6,108)=.853 failing to reject the hypothesis that the slopes and/or intercepts did not differ between the time blocks.

So far we have covered regressions concerning the workload probe variables as conceived in our simulation system and all variables available. A question was posed as to what would happen if predictor variables were limited to only those which could themselves be predictable. What was meant was system variables which could be predicted based on flight plan information without consideration for controller interaction with the aerospace system. Four variables were chosen somewhat arbitrarily, but they appear to meet the criteria of the original question. These variables were (1) complexity (CMAV), (2) handoffs inbound (HOIN), (3) handoffs outbound (HOUT), and (4) number of flights handled (NFLT).

The results of the regressions on real-time participant workload estimates (WIAV) are presented in tables 30 and 31. There was virtually no difference between the multiple R's when either four or two time blocks were employed. However, there was some decline in multiple R^2 using only the four "predictable" variables as compared to the best regressions previously reported. The results using these variables are closest to those reported using the variables from the "workload probe." The regressions were significant. Tables 32 and 33 provide the output of the "predictable" variables when regressed against OBQ1. The obtained multiple Rs were very close to those seen in tables 28 and 29 when all variables were employed. Closer inspection indicates that two of the variables stepped in for those equations (CMAV and NFLT) overlap two of the predictable variable set. It should be noted that the results of the all-variable regressions were not used to select the predictables. The latter selection was done independently by a disinterested individual knowledgeable in airspace operations.

An overview of this section of the results is informative concerning air traffic controller workload. Our wholistic model of workload accepts that the determiners of human workload are multidimensional, but that operators experience workload as an entity (Gestalt). The regressions demonstrate that there are multivariate (multidimensional) explanations which provide a better fit for an individual's or external observer's estimates of workload than any bivariate relationship. Both observers and controllers were able to provide univariate workload estimates based on a multiplicity of experiences. Although workload is experienced as a whole, there are complex interactions of stressors which drive workload in the dynamic person-machine environment of air traffic control. Regression techniques help identify those system elements which may serve as the best predictors.

TABLE 30. STEPWISE REGRESSION OF "PREDICTABLE" SYSTEMS VARIABLES ON WIAV

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.78	.60	3.158	F(4,235)=147.52**
HOUT	.83	.69	.206	
NFLT	.84	.70	.293	
HOIN	.85	.72	-.243	

**P $\leq$.01

TABLE 31. STEPWISE REGRESSION OF "PREDICTABLE" SYSTEMS VARIABLES ON WIAV (TIME BLOCKS 2 & 3 ONLY)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.82	.68	4.890	143.57**
HOUT	.84	.71	.211	

**P $\leq$.01

TABLE 32. STEPWISE REGRESSION OF "PREDICTABLE" SYSTEMS VARIABLES ON OBQ1

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
CMAV	.882	.780	7.567	382.54**
NFLT	.923	.850	.664	
HOUT	.929	.860	.160	
HOIN	.931	.870	-.204	

**P $\leq$.01

TABLE 33. STEPWISE REGRESSION OF "PREDICTABLE" SYSTEMS VARIABLES ON OBQ1 (TIME BLOCKS 2 & 3 ONLY)

Variables Selected	Multiple R	R^2	Regression Coefficients	F On Regression
NFLT	.924	.855	.798	258.65**
CMAV	.944	.892	7.245	
HOUT	.947	.896	.120	
HOIN	.948	.900	-.275	

**P $\leq$.01

POST-RUN QUESTIONNAIRE.

At the end of each period of air traffic control, the participant proceeded into an adjacent room and completed a five-item questionnaire. Questions focused on workload, busyness, degree of thinking required during the control period, and feelings of stress and fatigue (see appendix C). While post-run questionnaires are the traditional way of assessing operator workload, limitations based on memory and the primary/recency of salient events should not be forgotten.

As with the real-time measurement of ATWIT, it was hypothesized that respondents would rate their workload and other variables in relationship to the taskload which was induced by the experimental design. It was also anticipated that the five questions would not be consistently independent. In other words, it seemed likely that responses to some of the questions would overlap.

In the following series of tables (34 through 38), the mean responses for each question are reported along with the results of ANOVA's applied to the data. The first question asked participants to evaluate the workload which they had experienced during the previous 1-hour simulation. This question was based on a 12-point scale which was verbally authored in four blocks of 3-scale points. The bottom block referred to very low workload while the top block stated "very high workload — it was not possible to accomplish all tasks properly." No reference was made at the high end of the scale to losing control or losing the picture because of the emotional loading of such concepts for controllers.

In line with the instructions, controllers did use the entire scale, and their responses were in line with the three levels of taskload (table 34). The ANOVA indicated that there was no significant change in perceived workload between the replications demonstrating that increased experience over the data collection runs did not reduce perceived workload. This was a positive confirmation that the training and familiarization phase of the project functioned as it was designed to function.

The second question asked participants what fraction of the time they were busy during the control period. The influence of taskload was evident in the table of means and was supported by a strong main effect in the ANOVA (table 35). There was also a weak but significant main effect for the replication indicating a slight tendency to be less busy in the second replicate. However, this accounted for such a small proportion of total variance that it should be discounted.

The results for the question concerning how much thinking was required during the control period were straightforward (table 36). As taskload increased, so did the amount of thinking required to meet the requirements of the control situation. There was no replication effect.

When asked if they found the control period stressful, controllers indicated that as taskload increased so did the stress level (table 37). This was a strong main effect which was in line with what was hypothesized. The stress question also demonstrated a replication effect. The second replicate of three data runs was experienced as somewhat less stressful than the first. This result was tempered by the fact that the effect only accounted for slightly less than 4 percent of the variability in responses to this question.

TABLE 34. MEANS ON THE WORKLOAD QUESTION (12-POINT SCALE)

Taskload	Replication 1	Replication 2
A	2.67	2.33
B	8.56	6.56
C	10.56	10.22

ANOVA

Source	F Ratio	Correlation Ratio in %
Taskload	102.04**	84.44
Replication	5.22	1.57
Interaction	2.60	1.22

**P ≤.01

TABLE 35. MEANS FOR BUSYNESS QUESTION

Taskload	Replication 1	Replication 2
A	3.11	2.78
B	7.67	6.67
C	9.00	8.89

ANOVA

Source	F Ratio	Correlation Ratio in %
Taskload	96.29**	88.39
Replication	6.76*	0.81
Interaction	1.99	0.50

*P ≤.05

**P ≤.01

TABLE 36. MEANS ON THINKING QUESTION

Taskload	Replication 1	2
A	3.33	3.78
B	7.89	7.22
C	9.44	8.67

ANOVA

Source	F Ratio	Correlation Ratio in %
Taskload	42.07**	67.39
Replication	0.49	0.35
Interaction	0.59	0.95

**P <.01

TABLE 37. MEANS ON THE STRESS QUESTION

Taskload	Replication	2
A	3.22	2.22
B	6.33	5.22
C	8.00	7.44

ANOVA

Source	F Ratio	Correlation Ratio in %
Taskload	93.21**	80.43
Replication	15.06**	3.76
Interaction	.33	.27

**P <.01

The last question asked participants how tired they felt at the end of each control period. The results, as described in table 38, indicated a definite relationship between perceived fatigue and taskload. However, judging from the size of the F and correlation ratios, this effect may not have been as powerful as those seen on the other questions. There was no significant replication effect on perceived fatique.

TABLE 38. MEANS ON THE FATIGUE QUESTION

	Replication	
Taskload	1	2
A	3.11	2.67
B	5.44	4.44
C	6.22	6.00

ANOVA

Source	F Ratio	Correlation Ratio in %
Taskload	17.35**	47.55
Replication	5.00	2.07
Interaction	.23	.72

**P $\leq$.01

Results for the post-run questionnaire lend additional support to the belief that multiple levels of workload were generated by varying the taskload in this experiment. For the most part, participants did not identify any shifts from the first to the second replications of the three taskloads. The exception to this was a small perceived decrease in stress level across the replications. This may have been a result of building confidence as the actual data runs progressed.

A factor analysis was computed on post-run questionnaire responses. Packaged software (BMDP4-M) was employed. The results of the analysis are presented in table 39. For taskload A, where imposed workload was lowest, participant responses loaded all on one factor indicating that whatever the perceived workload, the other four questions were answered in a similar manner. However, as the taskload increased, the nature of controller responses on the five items became more complex. Taskload B was spread out over two factors, with the thinking, stress and fatigue items loading together on one factor and busyness and workload questions on a another factor. Taskload C was spread out over three factors as indicated in the table. It was interesting to note that questionnaire items 4 and 5, stress and fatigue, appeared in three of the four factors representing the three respective taskloads A, B, and C. The structure of participant responses across the five questions appears to have been related to the taskload. Participants made more differential responses to the five items as taskload increased. This finding is similar to one which appeared in previous aircrew workload studies (Stein, 1984) where a similar post-run questionnaire was used.

TABLE 39. FACTOR ANALYSIS OF POST-RUN QUESTIONNAIRE

Question	Taskload	1	2	3	4
2	A	.952			
	A	.941			
3	A	.875			
4	A	.726			
5	A	.612			
5	B		.859		
4	B		.822		
3	B		.745		
1	C		-.598		
5	C			.851	
4	C			.829	
2	B			-.659	
	B			-.633	
3	C				832
2	C				749

Cumulative proportion of accountable variance 79 percent

The post-run questionnaire could only be related to the cumulative data collected over the 60-minute control periods since it was completed at the end of each run. Some correlations with a selected group of the systems and workload variables are reported in table 40, these correlations were computed by considering all data in the 60 data collection runs. In other words, they represent the relationships between variables across all taskloads and replications. The factor analysis of the questionnaire, alone, was a finer grained approach in that it treated responses to each question on each taskload as a separate variable. It demonstrated that the 15 variables developed by this approach could be explained by four factors. In contrast, the correlations in table 40 ignore taskload membership. They demonstrated that the post-run questionnaire was a reasonably good estimator of cumulative workload information drawn from real-time variables. It was interesting to note that the questionnaire item which was most strongly related to OBQ4, observer estimate of effectiveness, was stress. The more stress the controller felt, the lower were his observer ratings for effectiveness. Not surprisingly, OBQ4 was inversely related to all the questionnaire items, which shared a considerable redundancy among themselves.

POST-EXPERIMENT WORKLOAD QUESTIONNAIRE.

This questionnaire (appendix C) was completed at the end of the experiment by 9 of the 10 participants. It asked a great deal of each respondent, requesting that he search his memory and integrate his experiences for the whole week. The purpose of this questionnaire was to obtain estimates of the workload impact of the simulation itself, and some of the decisions that were made in designing it.

TABLE 40. CORRELATIONS OF THE POST-RUN QUESTIONNAIRE WITH SELECTED OTHER VARIABLES (EXTRACTED FROM TABLE IN APPENDIX B)

System Variables	1 Workload	2 Busyness	3 Thinking	4 Stress	5 Fatigue
	.88	.88	.78	.80	.51
	.89	.90	.81	81	.54
	.92	.90	.76	.77	.48
NG2A	.89	.89	.82	86	.60
Workload Variables					
Sample (Taskload)	89	.89	.78	.83	55
WIAV	.87	.88	.81	.87	.61
	.93	.94	.81	.85	55
	-.54	-.54	-.53	-.65	-.54
Post-Run Questionnaire (Inter Item Correlations)					
1 Workload	1.00	.95	.83	.85	59
2 Busyness		1.00	.85	84	.56
3 Thinking			1.00	80	.61
4 Stress				1.00	77
5 Fatique					1.00

Participants were asked to provide their estimates on a 10-point scale where 1 represented a low workload impact and 10 represented a high impact. Table 41 provides descriptive statistics of the responses. Figure 4 provides a scatterplot of the responses to each item. A quick glance at the scatterplot and the standard deviation column of the table confirms the suspicion that there was a diversity of opinion on many of the items. The most meaningful items will obviously be those on which there was the highest level of agreement and the smallest standard deviation.

For the purposes of this discussion, we will first examine those items that have a standard deviation of 2.1 or less. Two items — (7) Emergencies, and (8) Weather — were included to see if participants were reading the questions. They apparently were, since the highest response given to either item was a 3 and the median for both was a 1. No emergencies or weather had been programmed into the traffic samples. Items which contributed to perceived workload and had a fairly high level of agreement included: (1) Number of Aircraft Handled, (4) Number of Altitude Changes, (13) Housekeeping, and (15) Using the Keypack. None of these were surprising based on participant behavior and comments during interview. Workload imposed by the keypack may be viewed as simulation-specific because the keyboard arrangement was dissimilar from those normally found in an En Route Center.

Items which did not contribute markedly to perceived workload and on which there was agreement included: (9) Pilot Verbal Responses/Errors/Delays, (10) Pilot Route/Altitude Errors, (16) Unfamiliarity with Airplanes, (19) Simulation Glitches, and (23) Aircraft/Pilot Procedural Violations. These items focus heavily on the simulation itself. The fact that they did not add to perceived workload supports the information provided by respondents during the formal interviews. When asked to rate the realism of the simulation, their median response was 7.5. There were a number of items in which there was somewhat less agreement (standard deviation between 2.2 and 3.0). Those which appeared to contribute to workload were: (3) Number of Vectors Given, (6) Using Strips Without a D-Man, (11) Accepting Handoffs, (12) Giving Handoffs, and (24) Responding to the Workload Response Box.

Items 3, 11, and 12, tend to confirm other information provided by the controller participants during interviews. Workload is increased in proportion to the number of tasks required to move each aircraft through the sector. Item 6 "Using Strips Without a D-Man," was simulation specific under high taskload conditions in which controllers would ordinarily call for assistance. However, the one-man sector appears to be an en route fact of life at lower taskload levels. In Item 24, we asked the participants to estimate the contribution of the ATWIT (Air Traffic Workload Input Technique) to their overall workload. Their median response of 6 was slightly above the midpoint of the scale (5.5). An examination of the scatterplot indicates quite a spread of opinion. It was anticipated that the response task would add to workload slightly. This was enhanced somewhat by a design artifact in the ATWIT system. This artifact required the controller to hold the response button down for 1.5 seconds in order for the input to be recorded by the computer. This was supposed to be upgraded so that an instantaneous push would record the input, but it never was. Controllers were extremely cooperative, and it was rather surprising that the median response was not higher than it was.

TABLE 41. POST-EXPERIMENT WORKLOAD QUESTIONNAIRE RESULTS (N=9)

Item Number	Workload Contributor	X	SD	Median
	Number of airplanes handled	8.44	1.42	8
2.	Number of conflicts	5.33	2.06	6
3	Number of vectors given	6.78		
4	Number of altitude changes	8.00	1.94	8
5	Number of airspeed reductions	1.67	1.10	1
6	Using strips without D-man	8.33	2.40	9
	Emergencies	1.22	.67	
8	Weather	1.11	30	
9	Pilot verbal response errors/delays	3.22	97	3
10	Pilot route/altitude errors	2.77	97	
11	Accepting handoffs	7.22	2.27	
12	Giving handoffs	6.44	2.30	6
13	Housekeeping (moving data blocks, removing strips)	8.44	1.90	
14	Using trackball	4.44	3.13	
15	Using keypack	8.11	1.83	
16	Unfamiliarity with airplanes	2.11	2.09	
17	Unfamiliar sector geometry	3.78	2.70	
18.	Area Restrictions	4.67	2.65	
19	Simulation glitches (failures/anomalies)	4.89	1.69	5
20	Lack of foot pedal comm. switch		0	
21.	Aircraft flight characteristics (climb, descend, airspeed, turn)	4.78	3.15	
22.	Coordination with other sectors	3.00	2.50	3
23	Aircraft/pilot procedural violations	2.75	1.39	
24.	Responding to the workload response box	6.11	2.42	
25.	Console layout	4.33	2.45	

WORKLOAD FACTORS	[illegible]									
		2	3		5	6			9	10
Number of airplanes handled						//	/	///	/	///
Number of conflicts		//			//	/	////			
umber of vectors given			/	/		/		//		//
Number of altitude changes				/			//	//		///
Number of airspeed reductions	HHT /	/	/	/						
sing strips without D-man			/			/			//	////
mergencies	HHT ///									
Weather										
t verbal response rrors/delays		//	////							
ot route/alt errors		//	////	//						
Accepting handoffs			/		//			//	////	
Giving handoffs				/	//	/			///	
'locks					/	/	/		//	////
sing trackball	//	//		/		/	/	/	/	
Jsing keypack				/			/	///	//	//
'nfam' with airplanes	HHT /									
liar sector geometry		//	//		/		//	/		
Area Restrictions	//		/	//		///		/		
			///				//			
	HHT ////									
haracteristics airspeed, turn					/	/		//		
Coordination with other sectors	///		///	/					/	
procedural violations	//	/	///	/	/					
Responding to the workload response box		/		/	//	/	/	//		/
Console Layout	//		/		//		/	/	/	

* Each tally mark represents one response to a questionnaire item.

FIGURE 4. POST-EXPERIMENT QUESTIONNAIRE SCATTERPLOT

Items which appeared to not be major contributers to workload and which had standard deviations of 2.2 to 3 were: (17) Unfamiliar Sector Geometry, and (18) Area Restrictions. The low median response to the geometry question indicates that controllers were comforable with the airspace used in the simulation. This was based on either their previous experience or the training/familiarization they received at the FAA Technical Center or both. Area restrictions were only employed in the high taskload portion of the design, and their impact may have been masked by other factors, such as traffic volume.

Items with a standard deviation greater than 3 should be interpreted with caution. Both of these items — (14) Using the Trackball and (21) Aircraft Flight Characteristics — tend to hug the middle of the scale and little could be definitively said about them.

INFORMAL INTERVIEWS.

There were two primary purposes for conducting these informal interviews. The first was to provide an unstructured opportunity for participants to express themselves in their own way about the project and what they had just experienced. This was an emotional safety valve for all training runs and four out of the six data collection periods. The second purpose was, of course, to collect information in a nonthreatening manner.

The interview form (see appendix C) contained little more than participant code, date, and run identification. There were no predefined questions as in the formal interview to be discussed later. The interviews were conducted immediately after the control run was completed in a room adjacent to the ATC simulation. The interviewer began the session with as nondescript an opening as possible, such as "How was that run?" or "How did that run feel?" Then as long as the participant kept talking, the interviewer took notes and did not interrupt. The informal interviews were collected on all 40 training runs and on 40 of the 60 data runs. A taxonomy or category system was developed for the purpose of summarizing the information available in the interviews. It consisted of three major categories (affect/emotion, fidelity/realism, and workload), each of which was composed of subcategories. This taxonomy evolved out of the structure of the responses and was not designed to do other than help describe the contents of these interviews.

Table 42 is the result of categorizing and tallying the frequency of participant responses. The tally procedure was accomplished by one individual so no estimate of rater reliability is offered. Several rules were applied to the tally process. If a participant said the same thing more than once in a given interview, only one response was counted. However, if he made the same response over a series of interviews, it was counted once per interview. The number of participants making a particular comment (in the table) was counted independently of the number of comments made. However, these two variables were well correlated (r =.93).

TABLE 42. INFORMAL INTERVIEW SUMMARY OF CATEGORIZED RESPONSES

Major Category	Subordinate Category	Train	A	B	C
	Frustrated	3(3		1(1)	2(1)
	Bored	1(1)	7(5		
Affect/Emotion	Stressed	7(6)		3(3)	7(5)
	More comfortable	21(10)	5(4)	4(4)	2(2
	Less comfortable	4(4		2(2)	
	Realistic	5(4)	1(1)	2(2)	
	Unrealistic	3(2)		2(2)	1(1)
Fidelity/Realism	Simop problems	4(3)		1(1)	2(2)
	Aircraft performance	3(3)	2(2)	1(1)	1(1
	Keyboard/display problems	21(9)		1(1	4(4)
	Low		8(7)		
	Mod			2(2)	
	High	3(3)		2(2)	2(2)
Workload	Needed help	3(3)		4(4)	4(4
	Lost picture	1(1)		1(1)	5(4)
	Failed to complete all tasks	14(8	1(1)	5(5)	7(5)
	Busy	6(4)		2(2)	7(6)

Number in parenthesis refers to number of controllers providing a given response

Descriptively, one can see a number of patterns in the comments made to the informal interviews. There were a sizable number of expressions that could be categorized as emotion or affect. This was particularly true during the training runs where participants were placed in a trial and error learning situation in an environment very similar to their normal working conditions, where errors are not acceptable behavior. Six controllers stated that they felt stressed and all participants expressed increasing comfort with experience in the simulation (suggesting discomfort in earlier training runs). During data runs there was an increasing tendency to report stress across the taskloads A, B, and C, respectively. This appears to have been inversely related to expressions of increasing comfort as taskload increased. Expressions of boredom appeared only under taskload A and nowhere else.

Most comments regarding the realism of the simulation appeared during the training runs and, to a lesser extent, during data collection. Becoming used to the keyboard, which was somewhat different from the traditional en route model, took practice. It appears to have also influenced four of the participants under the highest taskload, C. Spontaneous comments about simulation realism were not frequent and those favoring realism had a slight majority. The formal interview does more justice to the question of simulation fidelity and will be discussed shortly.

Participants' comments on their workload also formed a pattern of sorts. The low workload in taskload A stimulated more verbiage than either of the other taskloads when confining the tally to direct expressions of workload. Forty percent of the participants indicated that they really needed help in taskloads B and C, and four controllers admitted loosing the picture one or more times in taskload C, in comparison to one during B and none during A. It was very difficult for these professionals to admit that they had lost the picture. During training, eight controllers failed to complete all their duties during one or more control periods. Half the participants suffered similar experiences during taskloads B and C.

Judging from the frequencies of responses, it appeared that learning a new sector under moderately high taskload was an intense but accomplishable experience. Expressions of stress, task completion failure, and busyness increased across the taskloads despite the fact that the taskloads were counter-balanced in presentation order (table 42).

FORMAL INTERVIEWS.

The formal interviews were completed at the end of each block of three data collection runs which involved all three taskloads. These were structured interviews in which specific questions were asked. A blank copy of the protocol is available in appendix C. Once the question was posed, the interviewer allowed the participant to talk freely and only interrupted if the response strayed from the topic.

A summary of the responses to the ten interview questions follows. Response frequencies, by type, are included under each question as appropriate. Information from the two adminstrations of this interview has been pooled, and it reflects one or more responses by an individual controller. In no case will the frequency of responses exceed the total of ten controllers.

QUESTION 1. HOW REALISTIC WAS THE SIMULATION? For this question, nine out of the 10 subjects responded that the simulation was fairly, pretty, reasonably, or very realistic. In a binary choice situation, this result is significant at the a <.05 level (using a sign test). The main unrealistic aspects which were mentioned by controllers were as follows:

	Number of Controllers Commenting
Unrealistic Airspeeds	5
Need D or L Man	4
Incoming A/C Need Better Spacing	4
Incoming A/C Should be Lower Altitude	3
Airspace Seems Small	3
Unreal Climb Rates	3
Unreal Descent Rates	3
Unreal Turn Rates	2
Data Block Function is Better at Facility	2

The median rating on a 10-point scale of realism was, as indicated elsewhere a 7.5.

Regarding the simulation environment, it was mentioned that the simulator operators made some mistakes in interpreting clearances. However, controllers mentioned that in real life, these same mistakes, misinterpretations, or missed messages, occur at an equal or greater frequency. Communications in general were clearer in simulation with less noise and misunderstanding than standard en route, but did not seem to detract from realism once participants adapted.

QUESTION 2. RANK THE WORKLOAD OF THE THREE RUNS. WHAT MADE THE RUNS DIFFERENT? Each of the 10 controllers experienced two sets of runs; therefore, there were 20 responses to this question. Results showed that for every run controllers answered "yes," they did perceive a difference in workload across the runs. The reasons given are as follows:

	Number of Controllers Commenting
Volume of Traffic	9
Complexity	8
Restricted Area (In Taskload C)	5

In ranking the runs, the controllers were correct 18 out of 20 times. However, two controllers confused the mid and heavy samples (B and C). Controllers perceptions of differences in difficulty and subsequent workload were largely accurate and appear to have been based on those systems elements controlled by the experimental design.

QUESTION 3. HOW ADEQUATE WAS THE TRAINING/FAMILIARIZATION? Answer to this question for the 10 subjects ranged from "sufficient" to "excellent." Two controllers indicated that training could have been better. One said he would have liked a demonstration of the operational sector beforehand and more time to examine the sector. The other said he would have liked to practice the restricted area situation. The breakdown of responses was as follows:

	Number of Controllers Commenting
Adequate, Sufficient, Good, Okay, No Problem	5
Very Adequate	1
Outstanding, Excellent	2
Needed a Little More Training	2

The median rating of training adequacy on a 10-point scale was 7.5. This was the answer to QUESTION 4.

QUESTION 5. WHAT DID YOU FIND WAS THE MOST DIFFICULT FOR YOU TO ACCOMPLISH DURING THE LAST THREE RUNS? Controllers mentioned the following aspects as being difficult to accomplish:

	Number of Controllers Commenting
Conflicts (Determining, Resolving)	4
Offsetting Data Blocks	4
Planning (Instead or Reacting)	4
Keyboard Entries	4
Adjusting To High Traffic Volume	3
Getting Departures Up To Altitude	3
Descending Arrivals	3
Altitude Entries	2

QUESTION 6. WHAT INFLUENCES HOW HARD YOU HAVE TO WORK TO MAINTAIN YOUR PERFORMANCE? The comments given by two or more controllers were as follows:

	Number of Controllers Commenting
Weather	8
Traffic Volume	5
Personalities Of Adjacent Sector Controllers	5
Traffic Complexity (Actions Per Aircraft)	4
Computer (Equipment) Failure	3
Pilots Not Answering/Cooperating	3
Controller Mood	2
Unexpected Occurrences	2
Sector Coordination Problems	2
Holding Outbound Aircraft	2
Altitude Transitions	2

Controllers were asked to go beyond the current experiment for this and draw upon the depth of their experience.

QUESTION 7A. WHAT IS YOUR COMMON STRATEGY FOR DEALING WITH HIGH WORKLOAD? Comments to this question by two or more controllers are cited as follows:

	Number of Controllers Commenting
Give Direct Heading, If Adjacent Sector Accepts	6
Get Help	4
Safety First Then Accommodation	4
Prefer Vertical Separation	3
Spin Aircraft Before Confusing Picture	3
Coordinate To Space Incoming Aircraft	2
Prefer Horizontal Separation	2
Save Time Whenever Possible	

QUESTION 7B. DID YOU CHANGE YOUR REGULAR STRATEGIES IN ANY WAY? Comments elicited by two or more controllers are cited below:

	Number of Controllers Commenting
Adopted Normal Strategies	5
Increased Vectoring	
Reacting Not Planning	
Spinning Incoming Aircraft	
Took More Aircraft Than Would In Real Life	2

QUESTION 8. IS THERE ANYTHING ELSE YOU THINK WE SHOULD KNOW? Comments elicited by two or more controllers are listed below:

	Number of Controllers Commenting
Let It Get Out Of Hand (Got Behind)	3
May Have Forgotten Some Handoffs	2
Heavy Run Needed Two Controllers	2

QUESTION 9. IS IT FEASIBLE TO PREDICT WORKLOAD FROM PARAMETER TYPE INFORMATION? (9 controllers received this question). Comments elicited by two or more controllers are listed below:

	Number of Controllers Commenting
Yes	9
Controllers Use Strips Now	6
Previous Experience In Sector Is a Guide	3
Flow Control Is Attempting This	3

All controller participants felt that workload was a predictable entity. They indicated that they were already doing this using (1) what they knew about the sector from experience, (2) current and predicted weather, and (3) flight strip information.

QUESTION 10. WOULD YOU WANT, OR COULD YOU USE, INFORMATION CONCERNING YOUR WORKLOAD AT YOUR RADAR POSITION? Comments elicited by nine controllers on question 10 are given below:

	Number of Controllers Commenting
No	6
Yes	2
Uncertain	1

Two-thirds of the participants saw no advantage in having computer generated workload predictions down at their radar positions. However, given the sample size of only 9, this would not be significant beyond chance ($P \leq .05$). Also, the workload prediction package known as "workload probe" was conceived of as a management rather than operator tool.

TRAINING RESULTS.

The purpose of the training program was to familiarize participants with the equipment, procedures, and sector layout. It was hoped that they would be performing at or near asymtote when the training was completed so that learning would not confound the results from the data runs.

The training system was designed to provide up to 6 hours of full system simulation. In every case, the participants and research personnel agreed that 4 hours was enough. This conclusion was supported by a series of scatterplots in which the frequencies of systems variables at different magnitudes were plotted against training run order.

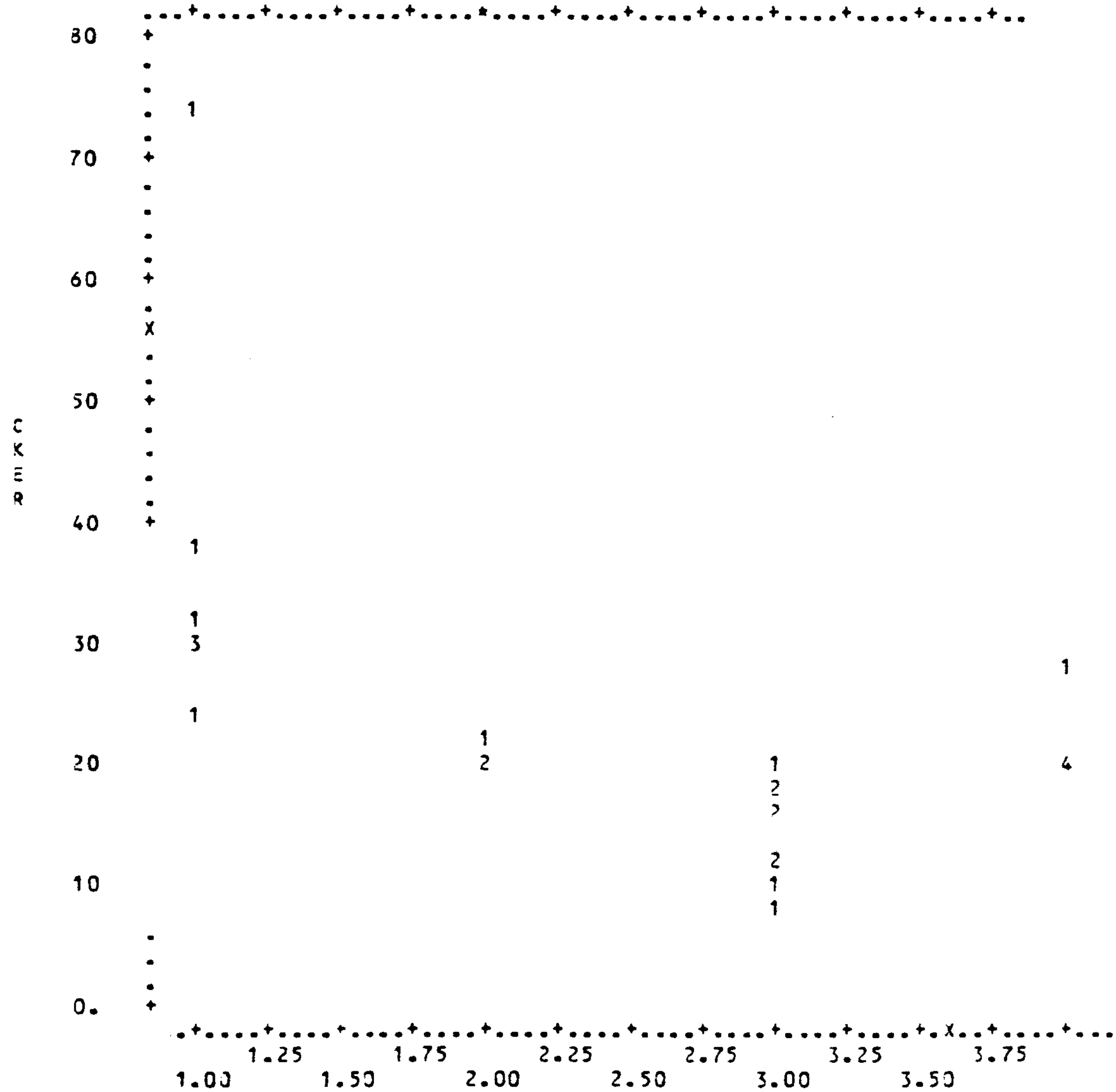
The majority of these plots and correlations indicated little or no systematic change across the four training runs. This could, of course, be interpreted in one of two ways: (1) the controllers already knew what they were doing when they arrived, or (2) the training was not effective. There were a number of relationships which refute the second alternative.

The controllers' familiarization with the data entry keyboard was a problem noted repeatedly in interviews. The number of keyboard entry errors showed a decreasing trend over the four training runs and was inversely correlated with training run order ($r = -.502$) (see figure 5). The duration of inbound handoff delays (where the controller lagged in accepting aircraft) decreased after training ($r = -.624$) (see figure 6). Controllers reported workload (WIAV) using the ATWIT system decreased over the four runs $r = -.570$, (see figure 7).

The mean reported workloads across the training runs were respectively: (1) 8.59, (2) 6.22, (3) 5.65, and (4) 5.51. The primary decrease in workload appears to have occurred between the first and second training runs.

11/09/83   15:33:43   Q6D.FREQ   JOB# 0006   000SOPER GOULD S.E.L. MPX-

PAGE 33   WORKLOAD PROBE - 60 MINUTE DATA SET. 3MDP6D. SAMPLES T AND U ONLY



N= 39
COR=-.5020   ORDER

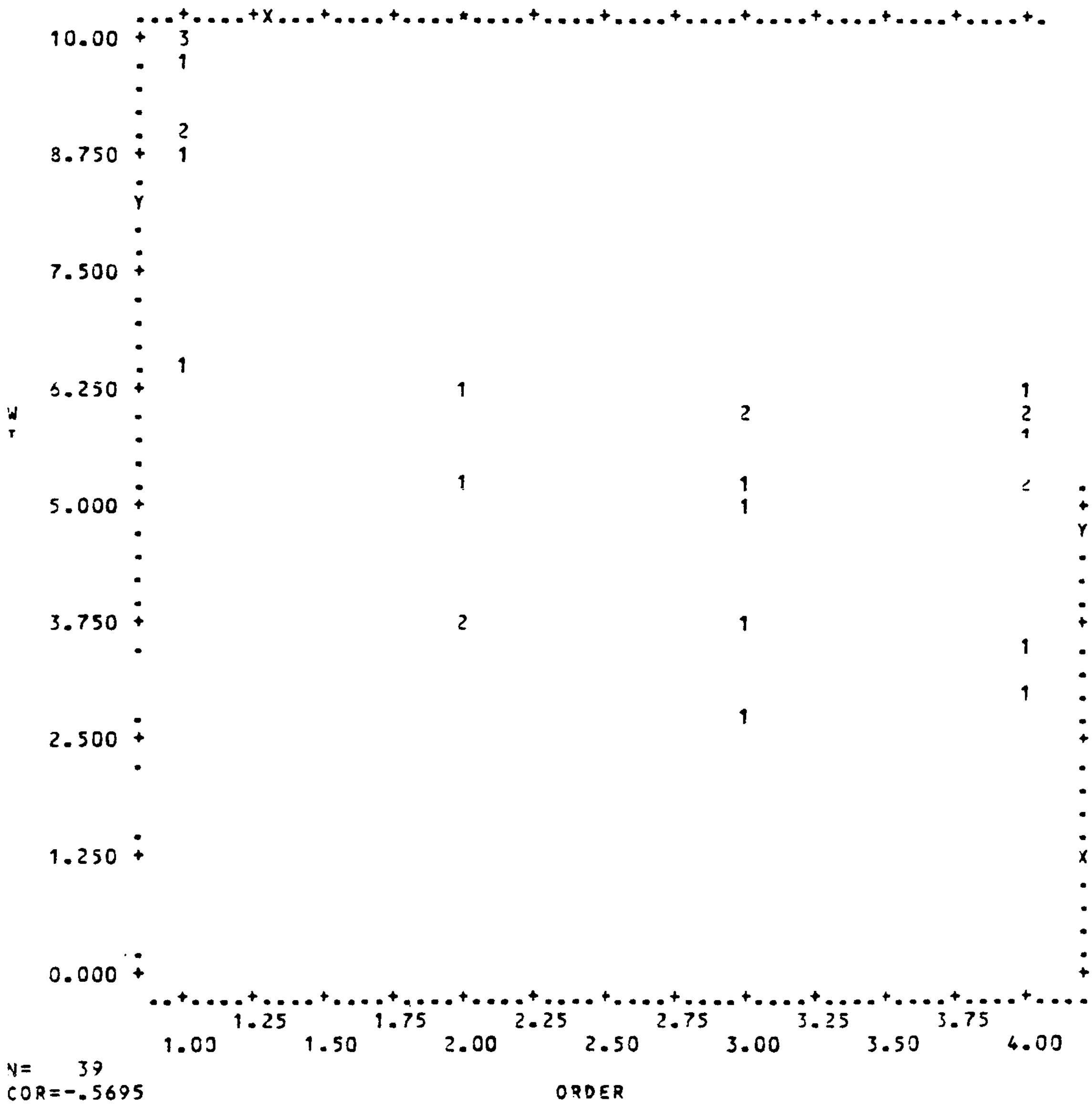
	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	2.5128	1.1441	X=-.04768*Y+ 3.5018	1.0057
Y	20.744	12.045	Y=-5.2345*X+ 34.023	111.47

VARIABLE 64 ORDER VERSUS VARIABLE 37 CKER

FIGURE 5. SCATTERPLOT KEYBOARD ERRORS

UR CATTERPLO HANDOF OU

PAGE 34 WORKLOAD PROBE - 60 MINUTE DATA SET. BMDP6D. SAMPLES T AND U ONLY



N= 39
COR=-.5695

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	2.5128	1.1441	X=-.32796*Y+ 4.6440	.90832
Y	6.4983	1.9869	Y=-.98907*X+ 8.9837	2.7393

VARIABLE 64 ORDER VERSUS VARIABLE 39 WIAV

FIGURE 7. SCATTERPLOT WIAV

The response time to the ATWIT system is another indicator of how busy the controller is in doing his primary tasks. Figure 8 shows a decreasing trend in this response latency (r =-.641) indicating that as participants became more comfortable with the simulation, they provided their workload estimates sooner. The mean response delays (WDAV) in seconds were: (1) 49.77, (2) 34.09,(3) 27.29, and (4) 26.06.

The observers ratings of workload and performance were interesting. The observers estimate of workload OBQ1AV did not show as much of a decreasing trend (r =-.175) as had the participant's ATWIT ratings. Observers did see some increase in performance, however (r =.404). The mean pooled observer effectiveness ratings were (1) 16.75, (2) 17.85, (3) 17.78, and (4) 18.30. The reader might wish to compare these means to those from the 60-minute cumulative data base for the three respective taskloads. The mean observer effectiveness ratings during actual data collection runs were: taskload A - 19.2, taskload B - 18.41, and taskload C - 17.21. The training runs were designed to be most like taskload B. By the fourth hour of training, the mean controller effectiveness rating was almost identical to that later achieved under taskload B.

The evidence appears to support a conclusion that the training package performed as it was designed to perform. Controllers brought a great deal of ability with them to the simulation but still had to learn to operate within the confines of unfamiliar territory. Important indicators, such as the frequency of keyboard errors and the amount of inbound delays, showed improvement. Interview and questionnaire information, as cited in other sections, demonstrated that the controllers became more comfortable and confident throughout the training period. Self-reported workload decreased with experience, and participants were able to make their responses more quickly as the training progressed.

CONCLUSIONS

1. The prediction of air traffic controller workload using a subset of systems variables is feasible.

2. Observer estimates of workload corresponded closely to the cumulative estimates of the controllers themselves.

3. There appeared to be a moderate inverse relationship between workload estimates and observer effectiveness ratings of controllers.

4. Controllers were willing and able to provide real-time workload estimates using ATWIT without any noticeable decrement in performance.

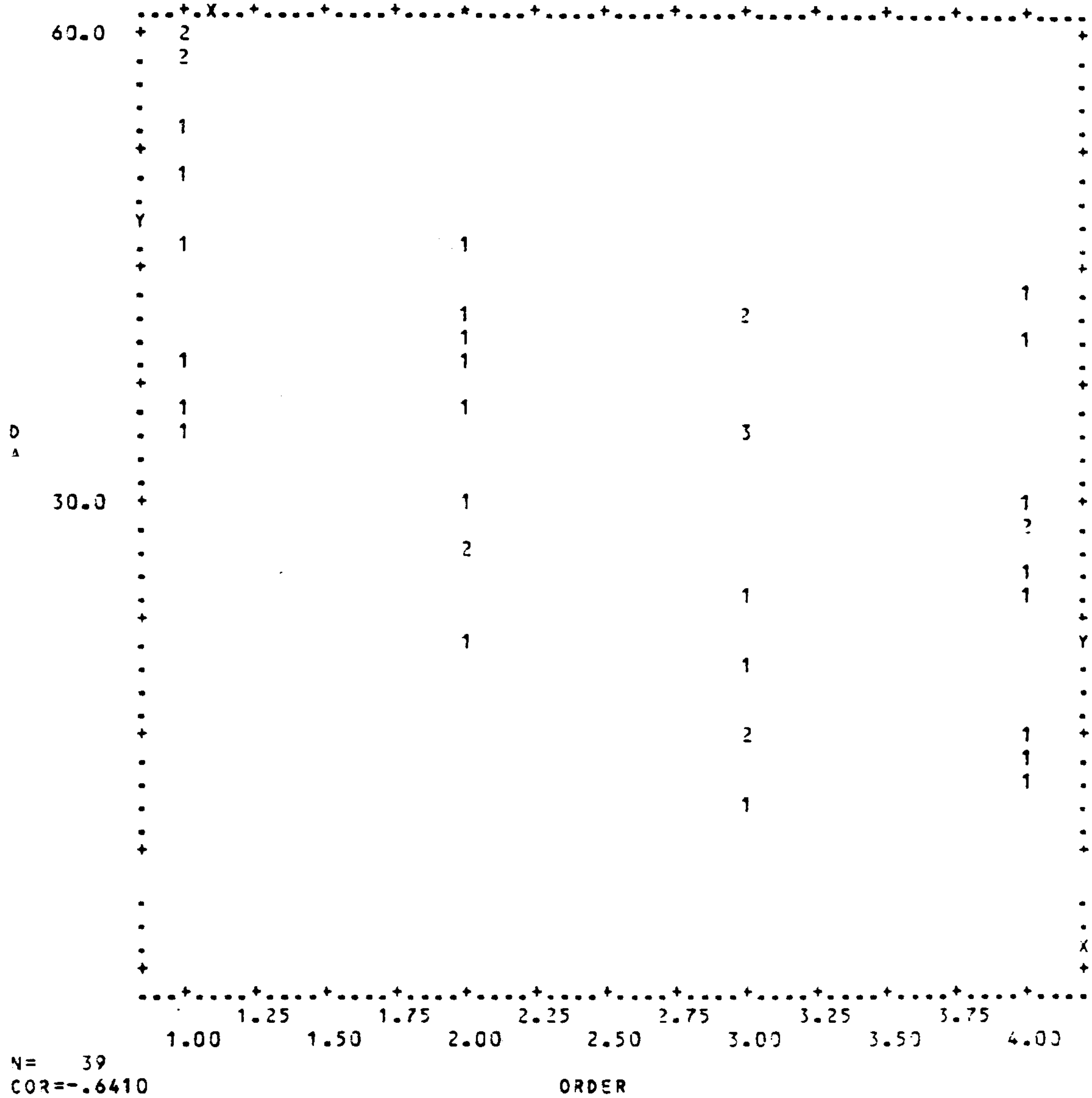
5. Changes in workload overtime during control periods were demonstrated.

6. While some single systems variables could be used to predict controller workload, a multivariate linear combination using regression techniques could do considerably better.

7. Those variables suggested by the workload probe concept could be used to form a viable workload prediction model.

```
11/09/83    15:33:43    Q6D.FREQ    JOB# 0006     000SOPER  GOULD S.E.L. MPX

PAGE  35     WORKLOAD PROBE - 60 MINUTE DATA SET. BMDP6D.  SAMPLES T AND U ONLY
```




```
N=   39
COR=-.6410

    MEAN      ST.DEV.     REGRESSION LINE      RES.MS.
    2.5128    1.1441    X=-.05240*Y+ 4.3100    .79195
    34.298    13.997    Y=-7.8425*X+ 54.005    118.53

VARIABLE    64   ORDER     VERSUS VARIABLE 42   WDAV
```

FIGURE 8.   SCATTERPLOT WDAV

BIBLIOGRAPHY

Barrer, J.N., Testing of Workload Probe Models, Mitre Memo No: W41-M6071, October, 1982.

Buckley, E.P., DeBaryshe, B.D., Hitchner, N., and Kohn, P., Methods and Measurements in Real-Time Air Traffic Control System Simulation, FAA Technical Center Report (DOT/FAA/CT-83/26), Atlantic City, N.J., April 1983.

Buckley, E.P., O'Conner, W.F., and Beebe, T., A Comparative Analysis of Individual and System Performance Indices for the Air Traffic Control System, FAA Technical Center Report (NA69-40, RD-69-50) Atlantic City, N.J., Sept. 1969.

Cobb, B.B., Nelson, P.L., and Mathew, J.J., The Relationship of Age and ATC Experience to Job Performance Rating of Terminal Area Air Traffic Controllers, FAA Civil Aeromedical Institute Report (FAA AM-73-7) Oklahoma City, April 1973.

Coulouris, C.J., et al, Capacity and Productivity Implications of Air Traffic Control Automation, Stanford Research Institute Report (FAA-RD-74-196) December 1974

Danaher, J.W., Human Error in ATC Operations. Human Factors, 22(s), 1980, 535-545.

Hopkin, V.A., Human Factors in the Ground Control of Aircraft, Agardograph No. 142, Hartford House: London 1970. NTIS AD 706550

Kirchner, J.H., and Laurig, W., The Human Operator in Air Traffic Control, Ergonomics, 14(5), 1971, 549-556.

Linton, M. and Gallo, P.S., The Practical Statistician, Monterey: Brooks-Cole, 1875.

Melton, C.E., Workload and Stress in Air Traffic Controllers in Hartman, B.O. and McKensie, R.E. Eds Survey of Methods to Assess Workload Agardograph No. 246, Hardford House: London 1979.

Moray, N., Subjective Mental Workload, Human Factors, 1982, 24(1), 25-40

Niedringhaus, W.P. and Gisch, A.H., Automated En Route Air Traffic Control Algorithmic Specifications - Sector Workload Probe, Mitre Corporation (MTR-83W152) McLean, Va., September 1983.

Rehmann, J.T., 1982, Cockpit Display of Traffic Information and the Measurement of Pilot Workload, Federal Aviation Administration Technical Center, Atlantic City, (DOT/FAA/EM-81/9).

Roscoe, A.H., Introduction to AGARD Monograph Assessing Pilot Workload, Hartford House: London, Feb. 1978, Pp 3-10 (DDIC AO 51587).

Rosenberg, B., Rehmann, J., and Stein, E.S., The Relationship Between Effort Rating and Performance in a Critical Tracking Task, FAA Technical Center Report (DOT/FAA/EM-81/13), Atlantic City, N.J., October 1982.

Sheridan, T.B., and Simpson, R.W., Toward the Definition and Measurement of the Mental Workload of Transport Pilots, Massachusetts Institute of Technology Final Report, Cambridge, Mass. 1979 (DOT-OS-70055)

Smoller, B., and Schulman, B., Pain Control - The Bethesda Program, Kensington: New York, 1982

Sperandio, J.C., Variation of Operators Strategies and Regulating Effects on Workload. Ergonomics, 14(5), 1971, 571-577

Stein, E.S., and Rosenberg, B., The Measurement of Pilot Workload, FAA Technical Center Report (DOT/FAA/EM-81/14), Atlantic City, N.J., January 1983.

Stein, E.S., The Measurement of Pilot Performance: A Masters-Journeyman Approach, FAA Technical Center Report (DOT/FAA/CT-83/15), Atlantic City, N.J., May 1984.

Swedish, W.J., Proposed Algorithm for AERA Sector Density Probe. Mitre Memo No. W41-M5676, May 1982.

Swedish, W.J., and Niedringhaus, W.P., Workload Probe Letter to Dr. E. Stein, Mitre Memo No: W41-3316, March 1983

APPENDIX A

INTRODUCTORY AND TRAINING MATERIALS

ATCS WORKLOAD PROJECT
PARTICIPANT BACKGROUND QUESTIONNAIRE

PARTICIPANT CODE__________ DATE ______

INSTRUCTIONS: The purpose of this questionnaire is to obtain information concerning your experience and background. This information will be used so that we can describe the participants in this study as a group and then relate the group's characteristics to how you perform and what you tell us during the Air Traffic Control Simulation. This data is being collected by participant code number only, and your name will not be associated with it.

1. DURING YOUR CAREER AS AN ATCS WHAT IS THE TOTAL NUMBER OF YEARS AND MONTHS YOU HAVE ACTIVELY CONTROLLED TRAFFIC?

   ____ YEARS ____ MONTHS

2. DURING HOW MANY MONTHS IN THE PAST TWELVE MONTHS HAVE YOU ACTIVELY CONTROLLED TRAFFIC?

   CIRCLE ONE

   0 1 2 3 4 5 6 7 8 9 10 11 12

3. DURING YOUR MOST RECENT EXPERIENCE AS AN ATCS WHERE DID YOU WORK?

   CHECK ONE

   ___ EN ROUTE ___ TERMINAL ___ OTHER (SPECIFY)________

4. CIRCLE THE NUMBER BELOW WHICH BEST DESCRIBES YOUR AGREEMENT WITH THE FOLLOWING STATEMENT:

   "I FREELY VOLUNTEERED TO PARTICIPATE IN THIS STUDY."

   STRONGLY DISAGREE 1 2 3 4 5 6 7 8 9 10 STRONGLY AGREE

5. CIRCLE THE NUMBER BELOW WHICH YOU FEEL BEST DESCRIBES YOUR CURRENT SKILL AS AN ATCS:

   AVERAGE 1 2 3 4 5 6 7 8 9 10 HIGH

6. CIRCLE THE NUMBER BELOW WHICH BEST DESRIBES YOUR AGREEMENT WITH THE FOLLOWING STATEMENT:

   "CONCERNING THIS EXPERIMENT, I AM LOOKING FORWARD TO PARTICIPATING."

   STRONGLY DISAGREE 1 2 3 4 5 6 7 8 9 10 STRONGLY AGREE

7. PLEASE STATE YOUR AGE IN YEARS_____. (YOU MAY DELETE THIS IF RESPONDING WOULD CAUSE YOU DISCOMFORT.)

PARTICIPANT CODE _______ DATE ______

8. IS THERE ANYTHING ELSE WE (THE EXPERIMENTERS) SHOULD KNOW WHICH MIGHT INFLUENCE YOUR PARTICIPATION IN THIS STUDY?

ATCS WORKLOAD PROJECT

Training/Familiarization Guidelines

Training programs generally proceed along one of two general routes. The usual method in academia (where no operational performance is required) is to provide a fixed number of hours of classroom instruction and then evaluate learning using paper and pencil tests and scaled grades. Industry and the military have moved away from this, realizing that test grades may or may not correlate with what a person is actually able to do at the end of training. This has led to the development of performance based criteria for many applied training programs especially where costs must be balanced against the rate of training successes.

It is very important that the training/familiarization portions of the ATCS Workload Project proceed smoothly and have specific goals. The instructional systems design or ISD model may be of some help to us here. This model specifies that prior to developing a program of instruction, a training needs analysis should be accomplished. This amounts to a definition of what we want to accomplish with the training. For this project, we can expect personnel with some diversity of experience and relative currency. This type of individual variability will only confound the results of the current experiment. All participants must be equally familiar with the equipment, procedures, and the airspace sector which we are using. They should be able to perform (a term which must be defined) to approximately the same level, given similar conditions.

The ISD model specifies that once training needs are identified, the trainer must specify three levels of information: tasks, conditions, and standards. An example follows:

Condition: Given control of a low density airspace sector as defined by no weather, approximately five aircraft in sector and no other special circumstances.

Task: The controller will plan, coordinate, communicate, and make necessary keyboard entries.

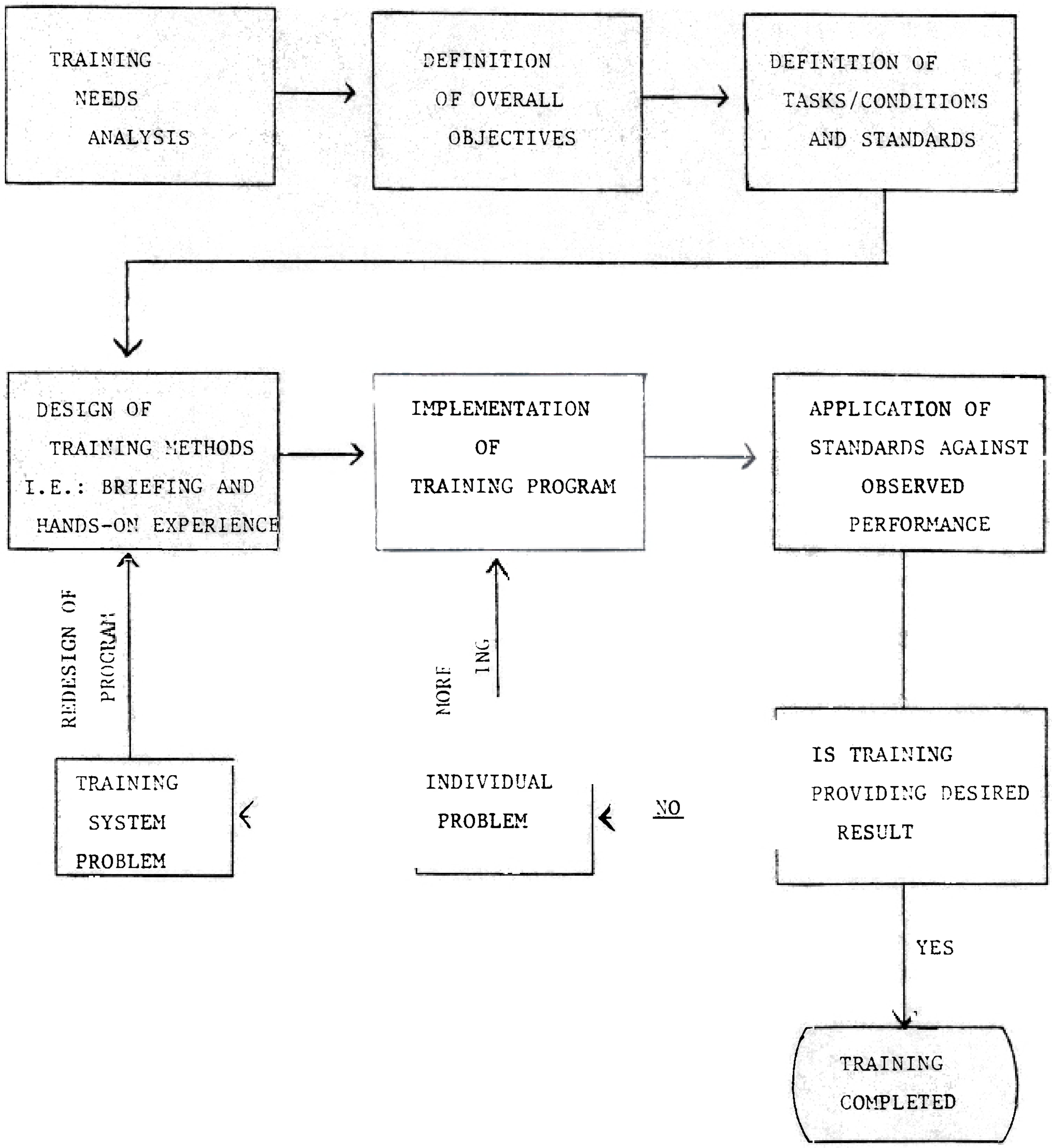
Standard: The controller will maintain separations of ___ miles between all pairs of aircraft. The controller will minimize delays, will transmit clearances and enter keyboard data correctly, etc

The purpose of the ISD model is twofold. First, it emphasizes an economization of resources by specifying exactly what behaviors are desired from trainees. These behaviors must be observable and/or measurable so that training progress can be determined. The second purpose focuses on the standards of performance. Since they are observable, they can generate feedback both to the individual trainee and to the training system. Knowledge of results in specific terms can be beneficial to both

Workload and performance are directly related in most person-machine systems. An operator who is not fully trained will be working with a performance handicap. This individual, assuming that he/she is motivated, will have to work harder to produce a poorer performance than the operator who has the skill and experience. This is why our training program must be adequate. We need a one-to two-page description of the tasks/conditions/standards which specifies what we will be doing in training and how we will know that the trainees have reached our goals. As part of this approach, we have to be prepared to accept a small percentage of training failures. There is no point in placing an individual in the experiment if he or she cannot perform to standard or if the individual feels very uncomfortable with the system. A secondary goal of the training, which is not directly observable, is the operator's feeling that he/she is completely familiar with the system. The decision concerning training failures will be made by consensus between the senior trainer, Mr. George Kupp, the other resident ATCSS, and the project psychologist. Hopefully, this will turn out to be a low probability problem.

A graphical description of the ISD model follows on the next page

A GRAPHICAL APPROXIMATION TO THE
ISD MODEL



Workload Probe

Keyboard Entries

Implied Handoff - 2 digits - Slew Enter.

Implied Tag Offset - digit Slew, Enter.

90° Tag Offset Key 9

Altitude Alt. Key, ALT x x x, Slew Enter
Alt. Key, ALT x x x, Space, A/C ID or CID, Keyboard Enter.

Vector Line - Inh/Sel Vec
(Protected 2 min. Flight Path)

Inhibit/Select - IDent, Mode C Alt, Speed

Data Block

Line 1 ACID

Line 2 Assigned Alt. ↑↓ or ± Mode C or assigned "C" when Mode C ± 200 ft. of Assigned.

Line 3 CID S (SPEED) x x x
or
H - Rec. Cont. No.
O -

SECTOR/	LI Y	HO #	FR QUENCY
CASANOVA	OW	02	343 282 2
F AT ROCK	INT	2	12545/ 26 4
GORDONSV LL	H GH	32	128 5/ 35 9
JORDONS V	LOW	3	1 4 / 3
NOTT NGHA	INT		1 2 / 277 4
R CHMOND	LOW	20	124 5 7 7
DULLES	APC	07	0 4 5
R C MOND	APC	08	9 9
WASH N	APC	09	35 2 3

WORKLOAD PROBE PROJECT

Procedures Briefing

I Map - 1978 version of Brooke Sector

Sector Boundaries - As outlined on map

Sector Altitudes - Exclusive of Approach Controls

0-to-infinity

Traffic Flow - Arrivals; departures and overflights

DCA - 100 ft. @ IRONSIDES IRONS)

RIC - 100 ft. @ MONTPELIER (PELER)

Overflight landing @ PHL 170 and BLO

landing @ NY METRO (LGA EWR JFK) 260 & BLO

II. Ground Rules

a) Flights over GVE RIC above FL 260 will be descending to FL 260

(b) Departures off DCA will be climbing to requested altitude.

c) If H/O not accepted by subject controller at boundary of sector - A/C will be spun by GHOST controller.

d) H/O given will be accepted by GHOST ten (10) miles from H/O point.

e) Anything goes with prior coordination

Subject controller controls the problem. If, at any time, subject feels overloaded he may coordinate with GHOST and slow down or meter traffic.

(g) Strips - keep first bay full

h) Operating techniques are individual's personal preference

(i Aircraft may be cleared direct to OMNIs, VORTACs, etc. and then resume navigation.

III. Observers

(a) Merely recording data.

(b) Will not answer questions once problem starts except during training runs.

(c) At least 4 training runs.

AIR TRAFFIC WORKLOAD INPUT TECHNIQUE

ATWIT

SCALE INSTRUCTIONS

One purpose of this research is to obtain an accurate evaluation of controller workload. By workload, we mean all the physical and mental effort that you must exert to do your job. This includes maintaining the "picture," planning, coordinating, decisionmaking, communicating, and whatever else is required to maintain a safe and expeditious traffic flow.

The way you will tell us how hard you are working is by pushing the buttons numbered from 1 to 10 on the response box mounted on the shelf in front of you. I will review what these buttons mean in terms of your workload. At the low end of the scale (1 or 2), your workload is low - you can accomplish everything easily. As the numbers increase, your workload is getting higher. Numbers 3, 4, and 5 represent increasing levels of moderate workload where the chance of error is still low but steadily increasing. Numbers 6, 7, and 8 reflect relatively high workload where there is some chance of making mistakes. At the high end of the scale are numbers 9 and 10, which represent a very high workload, where it is likely that you will have to leave some tasks incompleted.

All controllers, no matter how proficient and experienced, will be exposed at one time or another to all levels of workload. It does not detract from a controller's professionalism when he indicates that he is working very hard or that he is hardly working. Feel free to use the entire scale and tell us honestly how hard you are working! You will notice the clock at the bottom of your display begin blinking at the beginning of each minute that you control traffic. Please push the workload response button of your choice as soon as possible, then the blinking clock will stop. We realize that this requirement may be somewhat annoying at first, but please give it a chance for the purposes of this project (delete this sentence after the first reading). Thank you again for your cooperation, and remember that this data is being collected without any information which could later be used to identify you. Your privacy is protected.

OBSERVER EVALUATION

PARTICIPANT NO. ___________

RUN NO. EVALUATOR

TIME BLOCK

---

1. CIRCLE THE NUMBER BELOW DESCRIBES HOW HARD THE CONTROLLER WAS WORKING DURING THE PAST 15 MINUTES.

1	2	3	4	5	6	7	8	9	10
VERY EASY									VERY HARD

2. CIRCLE THE NUMBER BELOW WHICH BEST DESCRIBES HOW <u>BUSY</u> THE CONTROLLER APPEARED TO BE DURING THE PAST 15 MINUTES.

1	2	3	4	5	6	7	8	9	10
INTERMITTENT									CONTINUOUS

2A COUNT/TALLY THE NUMBER OF TIMES THE CONTROLLER EXHIBITS BEHAVIORS WHICH ARE <u>NOT</u> DIRECTLY RELATED TO HIS ATC DUTIES. IE: TALKS TO OBSERVERS, STARES AWAY FROM SCOPE OR FLIGHT STRIPS ETC.

3. CIRCLE THE NUMBER BELOW WHICH BEST DESCRIBES THIS CONTROLLERS EFFECTIVENESS DURING THE <u>PAST 15 MINUTES</u> IN TERMS OF RADAR TECHNIQUE. CONSIDER: VECTORING, ACCEPTANCE/DELIVERY OF HANDOFFS, COORDINATION AND THE FREQUENCY OF IN-BOUND DELAYS.

1	2	3	4	5	6	7	8	9	10
AVERAGE									EXCELLENT

OBSERVERS NOTES:

APPENDIX B

BIVARIATE RELATIONSHIPS

WORKLOAD PROBE   60 MINUTE DATA SET BMDP8D

COMPLETE CORRELATION MATRIX

		SAMPLE 3	REPLICAT 5	ALT 6	DG 7	SPD	RTE	HOLD 10	CMTR 11	HOIN 13	HOID 14
SAMPLE	3	1.0000									
REPLICAT	5	.0000	1.0000								
ALT	6	.7029	-.0523	1.0000							
HDG	7	.7759	-.0752	.6308	1.0000						
SPD	8	.1645	-.1008	.3541	.1081	1.0000					
RTE	9	.6402	-.1168	.4901	.8649	.0539	1.0000				
HOLD	10	.1739	-.1735	.1215	.1914	.0034	.1709	1.0000			
CMTR	11	.8806	-.0091	.7580	.5743	.1967	.4079	-.0139	1.0000		
HOIN	13	.9366	.0239	.7397	.6864	.1547	.5611	-.0070	.9285	1.0000	
HOID	14	.8663	-.1336	.6146	.6002	.2126	.4387	.3322	.7910	.7247	1.0000
HOUT	15	.9318	-.0091	.7591	.6567	.1932	.5476	.0159	.9469	.9780	.7700
NIAC	16	.9407	.0147	.7392	.7011	.1848	.5843	.0327	.9036	.9844	.7227
NFLT	17	.9563	-.0137	.7338	.7038	.1332	.5844	.0482	.9234	.9865	.7736
DFLT	18	.9320	.0242	.7626	.6778	.1786	.5585	-.0039	.9232	.9925	.7177
DIST	19	.9415	-.0184	.7521	.6878	.1860	.5667	-.0008	.9222	.9904	.7382
FUEL	20	.8718	.0397	.8112	.6297	.2032	.5290	-.0428	.9156	.9638	.6637
DINB	21	.8963	.0537	.7753	.6265	.1822	.4941	-.0666	.9464	.9823	.6891
NG2G	22	.4420	-.2261	.6255	.4567	.5700	.3705	.3713	.3690	.3351	.5658
DG2G	23	.2830	-.3149	.4090	.3019	.3551	.2355	.6337	.1931	.1281	.5089
NG2A	24	.9209	-.0885	.8527	.7869	.2682	.6311	.1497	.8806	.9159	.7779
DG2A	25	.8196	-.1250	.8325	.7415	.3206	.5727	.1911	.7480	.7803	.7256
NAIR	26	.9073	-.0137	.8644	.8122	.2519	.6412	.0708	.8839	.9370	.6990
DAIR	27	.9149	-.0213	.8119	.7546	.2089	.5796	.0373	.8937	.9463	.7104
NDLY	28	.5135	-.1112	.3117	.5096	.0325	.4383	.4320	.3715	.3658	.5532
DDLY	29	.5139	-.1959	.2516	.5201	.0630	.5368	.5900	.2877	.3729	.4928
NSCF	30	.6380	-.1168	.6275	.4710	.1895	.3949	-.0673	.7134	.7458	.5065
DSCF	31	.5216	-.2412	.5487	.3061	.2006	.2458	-.0156	.6648	.6722	.4017
CKEN	36	.6856	.0532	.6545	.6527	.0057	.5415	-.0460	.7373	.7510	.5225
CKER	37	.6931	-.1386	.4396	.3253	.1591	.2561	.2384	.6384	.5962	.7518
WIAV	39	.8953	-.0744	.6883	.7201	.2151	.6073	.1920	.7992	.8331	.8258
WDAV	42	.6913	-.1080	.3340	.3782	.1561	.3160	.2420	.6016	.6110	.7137
CMAV	44	.9259	.0431	.7545	.6698	.1632	.5433	.0715	.9173	.9653	.7569
OBQ1AV	49	.9530	-.0868	.7891	.7299	.2402	.5953	.1684	.9071	.9283	.8652
OBQ2AV	50	.9182	-.0902	.8095	.6974	.2481	.5458	.1522	.8927	.9062	.8332
OBQ4AV	51	-.5683	.2804	-.3582	-.4637	-.0649	-.4587	-.3873	-.4108	-.4103	.6707
PRQ1	52	.8945	-.1070	.7411	.6822	.1866	.5398	.2350	.8293	.8757	.7901
PRQ2	53	.8925	-.0729	.7585	.6552	.1576	.5154	.2010	.8536	.8893	.8067
PRQ3	54	.7758	-.0396	.6367	.5752	.1209	.4608	.1953	.7176	.7677	.6460
PRQ4	55	.8346	-.1720	.7039	.7325	.1892	.6520	.2144	.7068	.7849	.6969
PRQ5		.5498	-.1188	.4351	.5342	.1304	.4656	.1337	.4566	.4942	.4669

TABLE B-1   BIVARIATE CORRELATIONS   60 MINUTE DATA (1 of

WORKLOAD PROBE - 0 MINUTE DATA SET. BMDP8D

		HOUT 15	NIAC	NFLT 17	DFLT 18	IST 19	FUEL 0	DINB	NG2G 22	DG2G 23	NG2A 24
HOUT	15	1.0000									
NIAC	16	.9650	1.0000								
NFLT	17	.9786	.9845	1.0000							
DFLT	18	.9776	.9908	.9834	1.0000						
DIST	19	.9763	.9897	.9868	.9969	1.0000					
FUEL	20	.9674	.9594	.9483	.9760	.9645	1.0000				
DINB	21	.9689	.9679	.9601	.9830	.9801	.9778	1.0000			
NG2G	22	.3899	.3639	.3913	.3514	.3614	.3835	.3147	1.0000		
DG2G	23	.1936	.1724	.1952	.1407	.1563	.1547	.0345	.7909	1.0000	
NG2A	24	.9189	.9193	.9292	.9188	.9250	.9036	.8972	.6073	.3890	1.0000
DG2A	25	.7795	.7951	.8024	.7843	.7926	.7631	.7525	.7101	.5213	.9164
NAIR	26	.9208	.9406	.9313	.9409	.9376	.9280	.9298	.4794	.2580	.9615
DAIR	27	.9257	.9467	.9381	.9469	.9491	.9217	.9361	.4301	.2225	.9479
NDLY	28	.3939	.4148	.4436	.3874	.4038	.3415	.3228	.4627	.4599	.4925
DDLY	29	.3752	.4214	.4324	.3893	.4045	.3260	.3013	.4343	.4752	.4613
N5CF	30	.6963	.7471	.7130	.7580	.7595	.7459	.7706	.2200	.0649	.6573
D5CF	31	.6395	.6644	.6343	.6803	.6816	.6864	.7046	.2258	.0791	.6085
CKEN	36	.7692	.7321	.7330	.7502	.7310	.7808	.7661	.1548	.0520	.6846
CKER	37	.6205	.6008	.6447	.6003	.6185	.5584	.5823	.3749	.2403	.6092
WIAV	39	.8438	.8372	.8666	.8312	.8447	.7915	.7943	.4956	.3517	.8789
WDAV	42	.6364	.6200	.6512	.6044	.6203	.5367	.5618	.2734	.2927	.5496
CMAV	44	.9541	.9601	.9562	.9710	.9651	.9403	.9703	.3642	.1315	.9053
OBQ1AV	49	.9395	.9310	.9482	.9266	.9327	.8980	.9038	.5228	.3541	.9381
OBQ2AV	50	.9232	.9138	.9242	.9122	.9154	.8959	.8951	.5456	.3581	.9285
OBQ4AV	51	-.4436	-.4338	-.4844	-.3996	-.4262	-.3428	-.3226	-.4851	-.5585	-.4902
PRQ1	52	.8724	.8832	.8882	.8767	.8840	.8340	.8562	.4603	.3138	.8901
PRQ2	53	.8964	.8836	.8993	.8838	.8861	.8584	.8639	.4366	.3030	.8858
PRQ3	54	.7612	.7816	.8076	.7718	.7801	.7353	.7380	.3968	.2548	.8167
PRQ4	55	.7823	.7951	.8144	.7784	.7947	.7435	.7310	.4823	.3501	.8560
PRQ5	56	.5066	.5095	.5382	.4912	.5086	.4744	.4452	.3777	.2625	.6026

TABLE B-1 BIVARIATE CORRELATIONS 60 MINUTE DATA (2 of 3)

WORKLOAD PROBE 60 MINUTE DATA SET. MDP8D

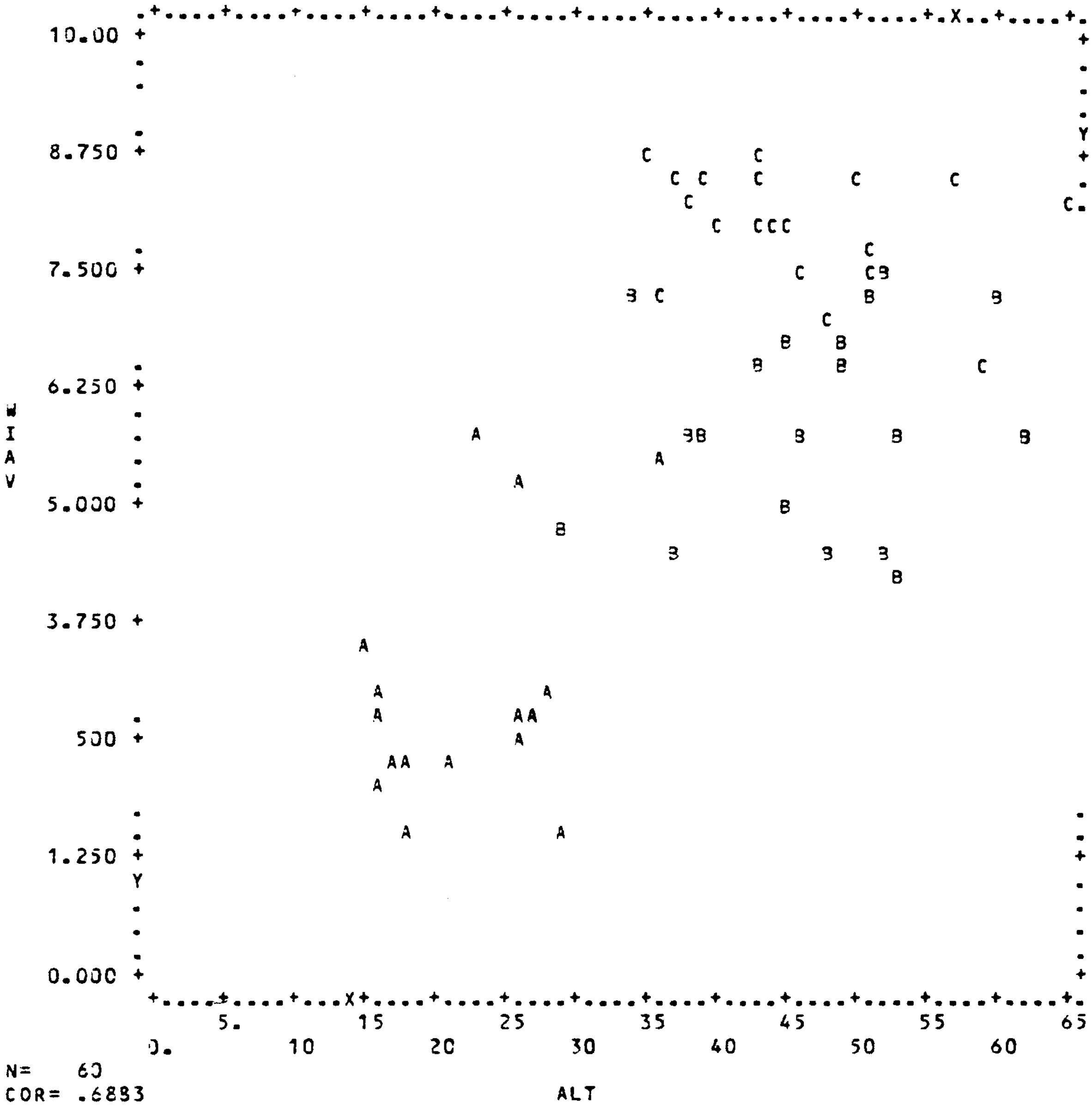
		DG2A 25	NAIR 26	DAIR 27	NDLY 28	DDLY 29	NSCF 30	DSCF 31	KEN 36	CKER 37	WIAV 39
DG2A	25	1.0000									
NAIR	26	.8661	1.0000								
DAIR	27	.8574	.9735	1.0000							
NDLY	28	.4727	.4004	.3531	1.0000						
DDLY	29	.4595	.3802	.3540	.8605	1.0000					
NSCF	30	.5388	.7092	.7256	.2179	.2356	1.0000				
DSCF	31	.5075	.6302	.6722	.1517	.1899	.8101	1.0000			
CKEN	36	.5215	.7664	.7359	.1826	.1637	.5623	.4346	1.0000		
CKER	37	.5163	.5225	.5442	.4265	.3717	.3847	.3384	.2615	1.0000	
WIAV	39	.8400	.8292	.8424	.4950	.4986	.5997	.5142	.6580	.6533	1.0000
WDAV	42	.5646	.5157	.5577	.3670	.4662	.3906	.3305	.3588	.5833	.7364
CMAV	44	.7766	.9149	.9162	.4299	.4108	.7574	.6737	.7284	.6225	.8357
OBQ1AV	49	.8580	.9186	.9185	.4908	.4833	.6903	.5902	.7193	.6941	.9165
OBQ2AV	50	.8674	.9081	.9120	.4661	.4562	.6830	.6044	.7152	.6517	.9025
OBQ4AV	51	-.5550	-.3999	-.3786	-.6537	-.6221	-.2243	-.1278	-.2418	-.5151	-.5663
PRQ1	52	.8290	.8641	.8768	.5014	.5082	.6746	.6440	.6182	.6694	.8662
PRQ2	53	.8274	.8622	.8773	.4387	.4436	.6256	.5816	.7018	.6605	.8846
PRQ3	54	.7764	.7621	.7854	.4903	.4864	.5169	.4944	.5578	.6163	.8126
PRQ4	55	.8263	.8086	.7995	.4589	.4938	.5592	.4860	.5908	.5559	.8665
PRQ5	56	.5847	.5220	.4988	.4833	.4322	.2491	.2110	.3056	.4612	.6101

		WDAV 42	CMAV 44	OBQ1AV 49	OBQ2AV 50	OBQ4AV 51	PRQ1 52	PRQ2 53	PRQ3 54	PRQ4 55	PRQ5 56
WDAV	42	1.0000									
CMAV	44	.6103	1.0000								
OBQ1AV	49	.6967	.9276	1.0000							
OBQ2AV	50	.6809	.9120	.9827	1.0000						
OBQ4AV	51	-.5059	-.4163	-.5846	-.5402	1.0000					
PRQ1	52	.6084	.9175	.9270	.9124	-.5435	1.0000				
PRQ2	53	.6510	.9025	.9421	.9334	-.5419	.9539	1.0000			
PRQ3	54	.5236	.7637	.8106	.8033	-.5324	.8286	.8469	1.0000		
PRQ4	55	.5273	.7726	.8457	.8211	-.6477	.8488	.8352	.8043	1.0000	
PRQ5	56	.2243	.4801	.5511	.5169	-.5410	.5893	.5648	.6125	.7742	.0000

UMBER OF INTEGER WORDS OF STORAGE USED IN PRECEDING PROBLEM 2693
PU TIME USED 56.552 SECONDS

TABLE B-1. BIVARIATE CORRELATIONS 60 MINUTE DATA (3 of 3)

PAGE 6 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



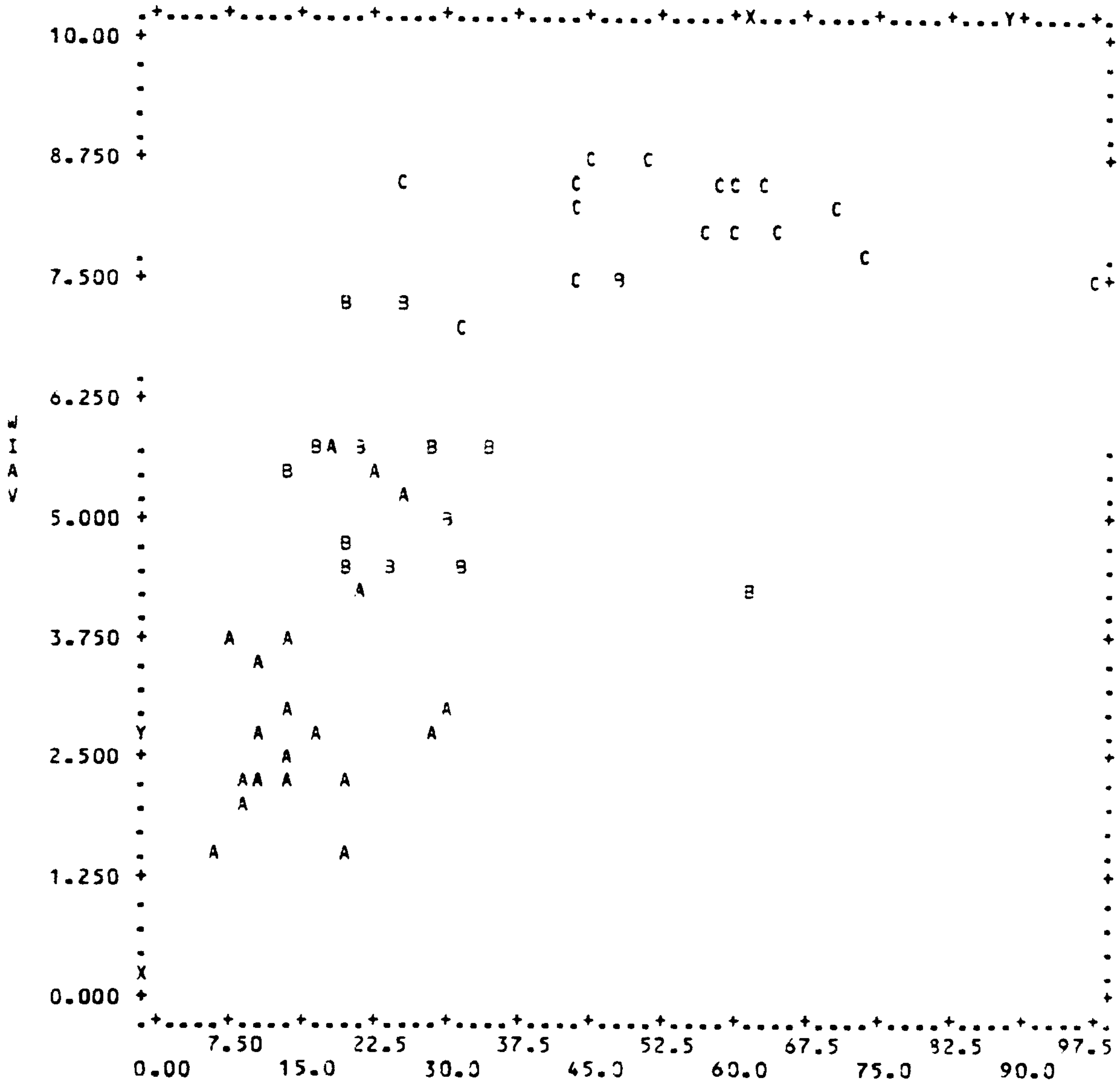
N= 60
COR= .6883

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	38.217	13.173	X= 4.0606*Y+ 15.131	92.963
Y	5.6729	2.2338	Y= .11667*X+ 1.2140	2.6711

VARIABLE	6	ALT	VERSUS VARIABLE	39	WIAV	FOR GROUP A.SAMPLE	SYMBOL=A
VARIABLE	6	ALT	VERSUS VARIABLE	39	WIAV	FOR GROUP B.SAMPLE	SYMBOL=B
VARIABLE	6	ALT	VERSUS VARIABLE	39	WIAV	FOR GROUP C.SAMPLE	SYMBOL=C

10/18/83 14:20:43 Q6DRUN01 JOB# 0017 OOOSOPER GOULD S.E.L. MPX-32 2

PAGE 7 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

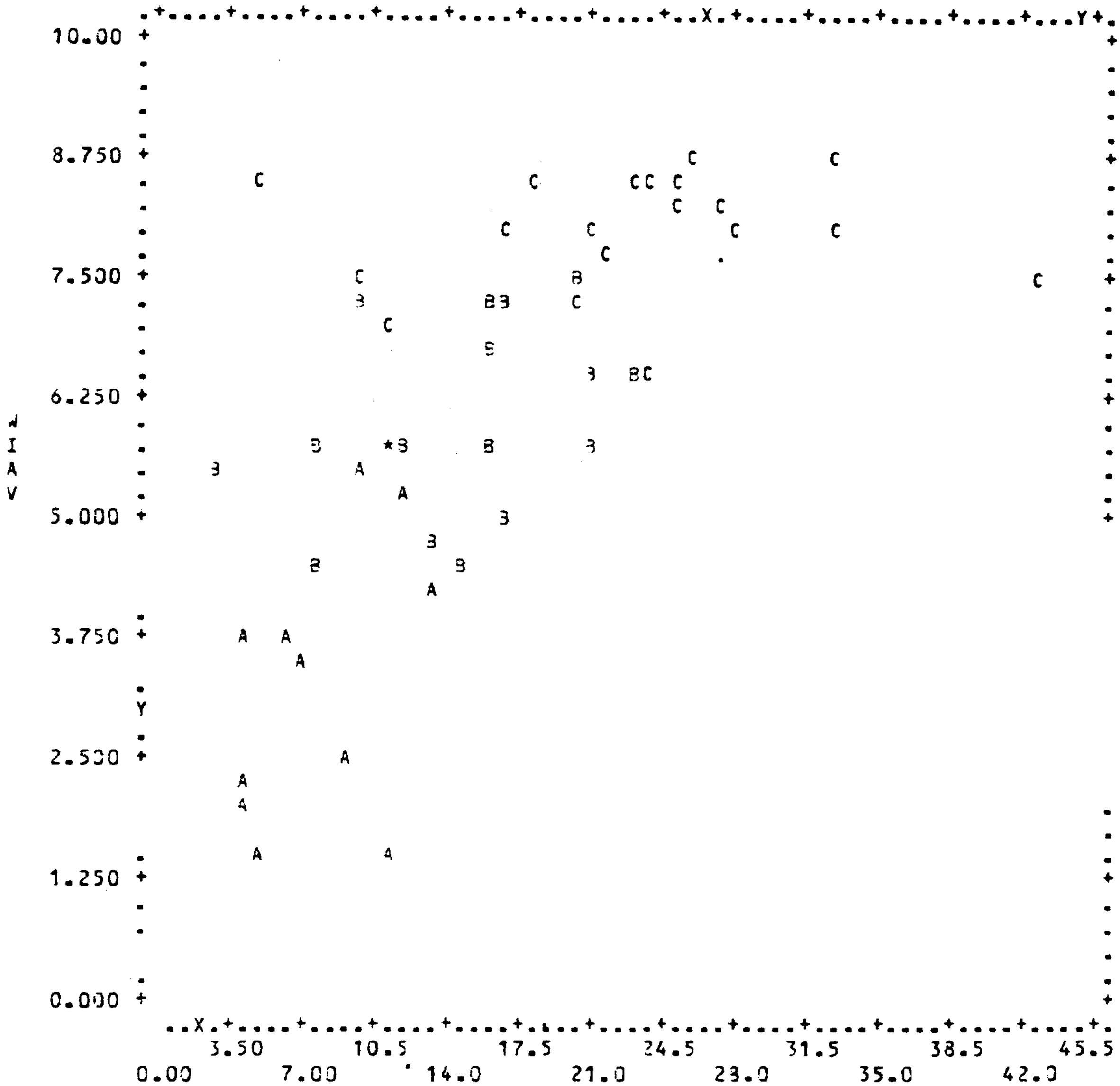


N= 60
COR= .7201

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	32.767	19.658	X= 6.3369*Y-3.1820	189.29
Y	5.6729	2.2338	Y= .08182*X+ 2.9919	2.4441

VARIABLE 7 HDG VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 7 HDG VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 7 HDG VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 8 WORKLOAD PROBE 60 MIN. CUM. DATA. BMDP6D



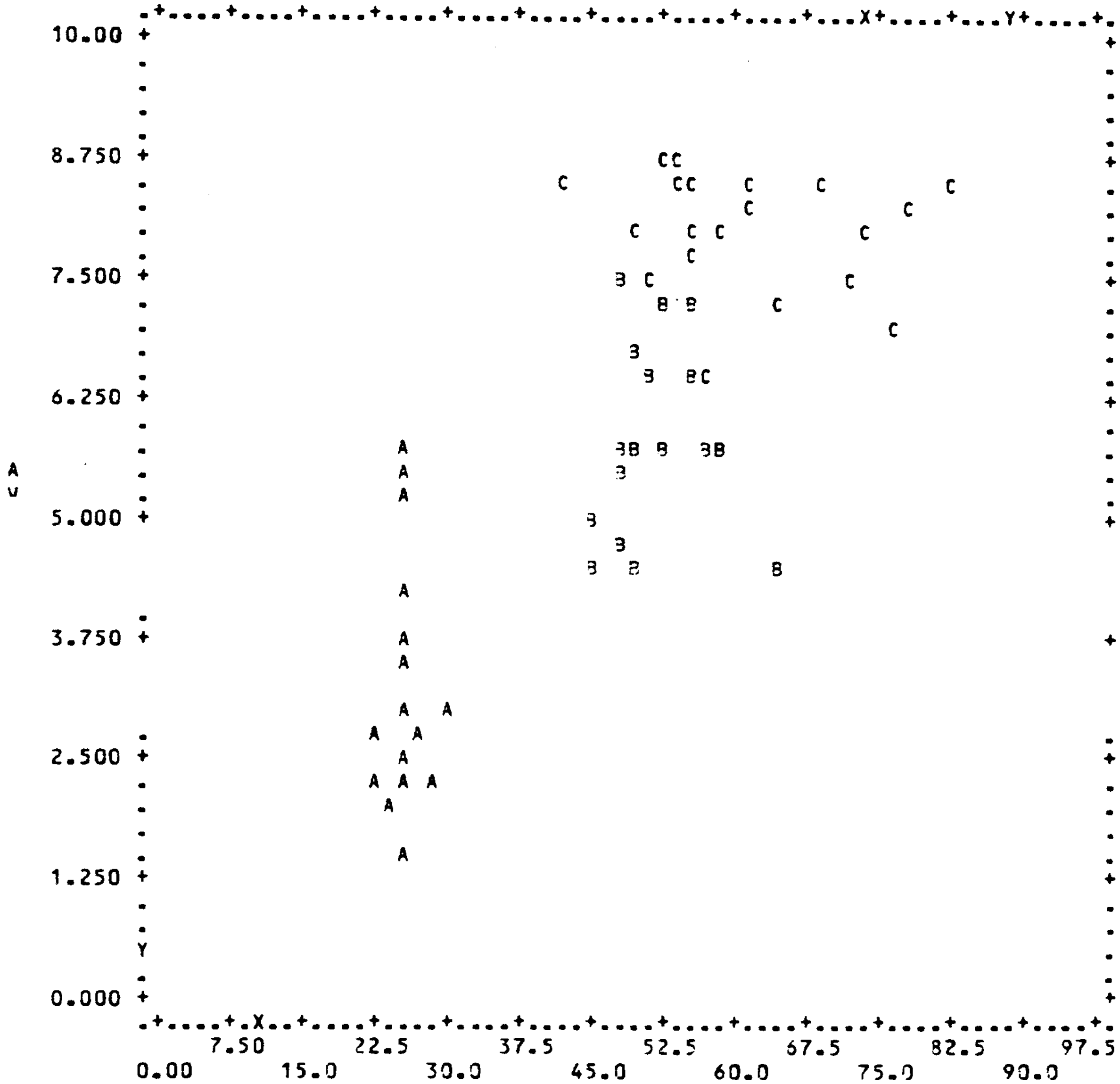
N= 60
COR= .6073

RTE

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	15.933	8.4833	X= 2.3078*Y+ 2.8913	46.258
Y	5.6729	2.2333	Y= .15983*X+ 3.1133	3.2037

VARIABLE 9 RTE VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 9 RTE VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 9 RTE VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 9 WORKLOAD PROBE - 50 MIN. CUM. DATA. BMDP6D

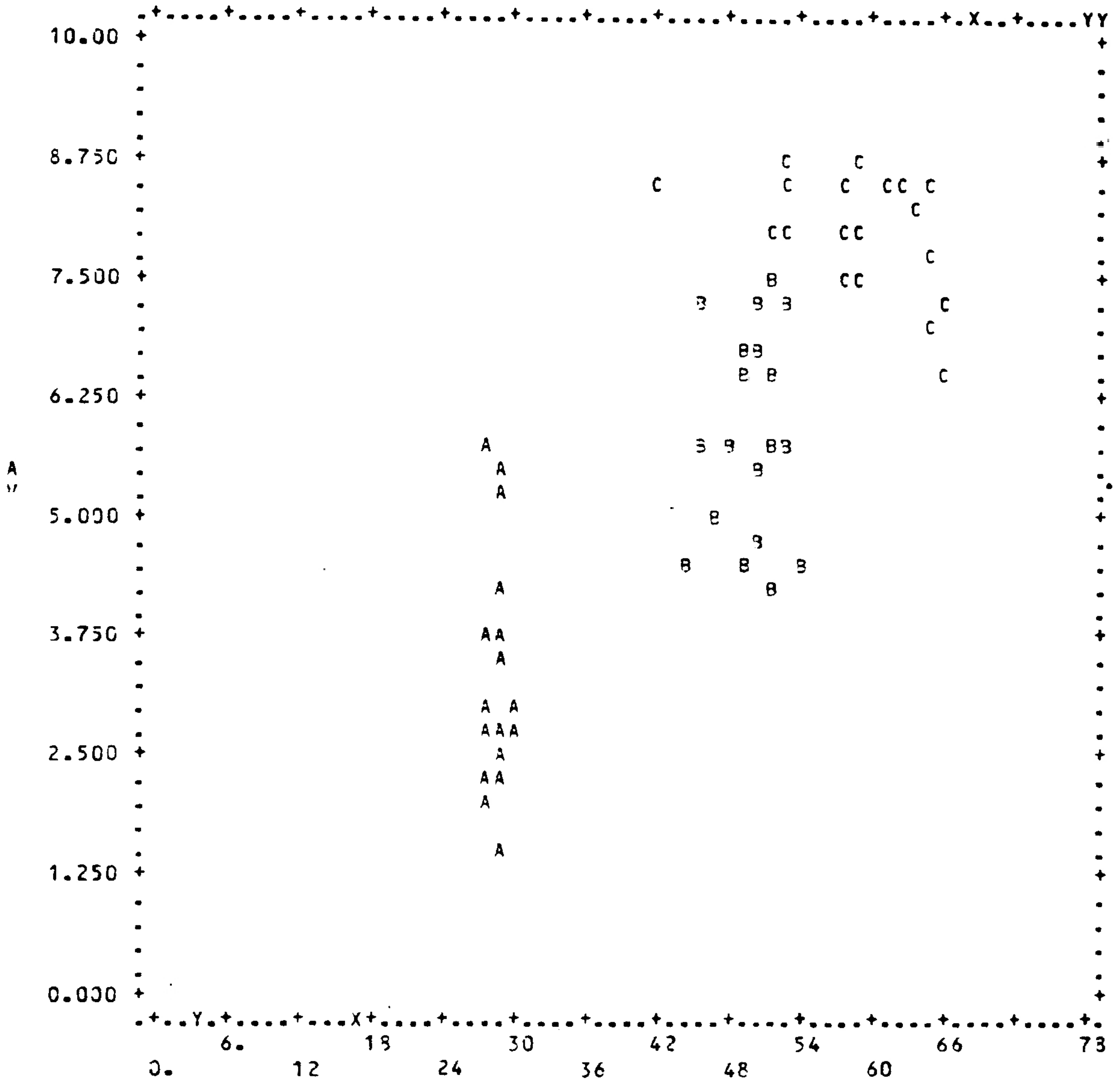


N= 60
COR= .7992

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
46.050	16.668	X= 5.9631*Y+ 12.222	102.11
5.6729	2.2338	Y= .10711*X+ .74064	1.8340

VARIABLE 11 CMTR VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 11 CMTR VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 11 CMTR VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 10 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

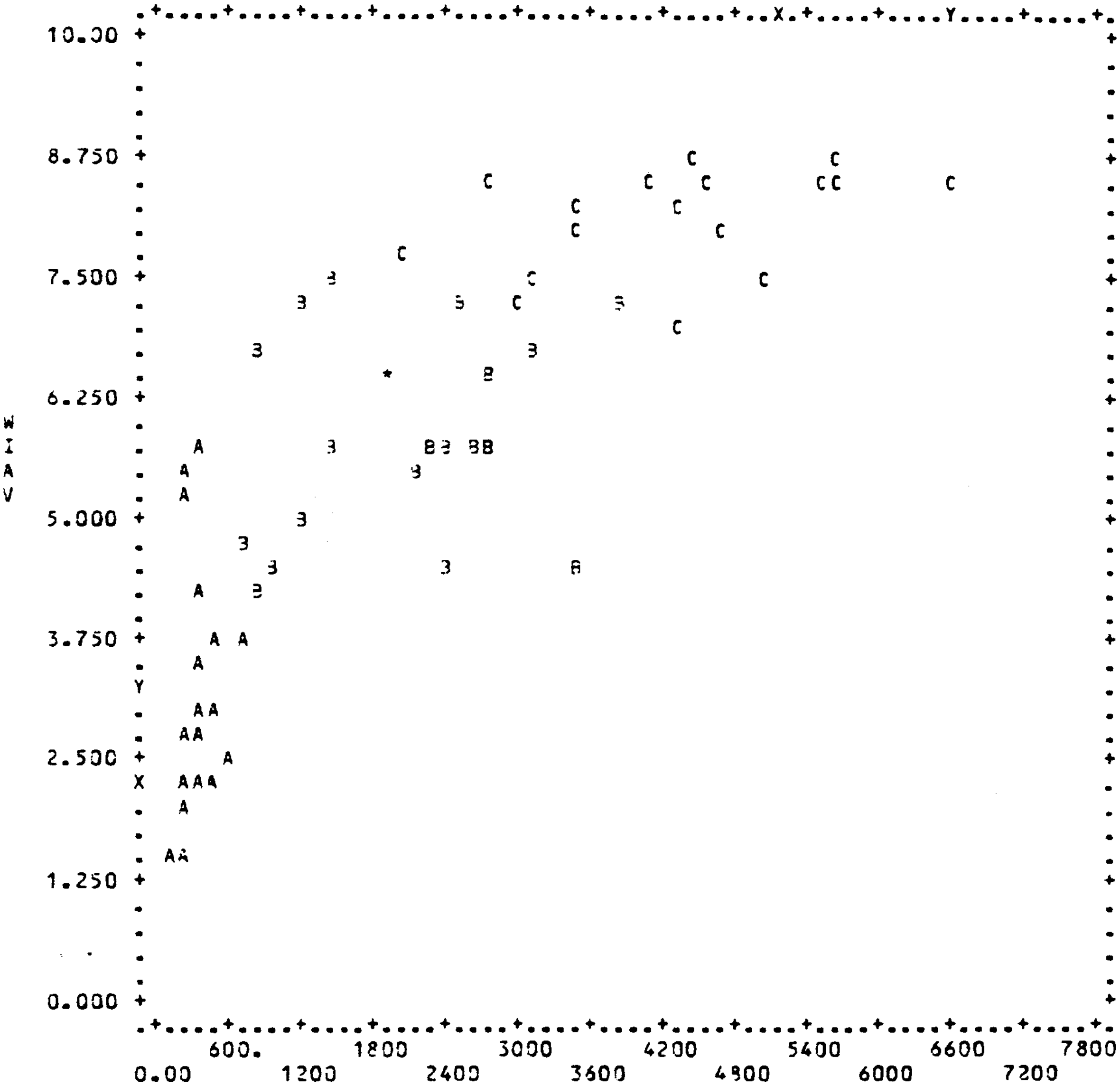


N= 60
COR= .8330 HOIN

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	45.717	13.341	X= 4.9750*Y+ 17.494	55.403
Y	5.6729	2.2338	Y= .13949*X-.70414	1.5534

VARIABLE 13 HOIN VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 13 HOIN VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 13 HOIN VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 11 WORKLOAD PROBE - 60 MIN. CUM. DATA. EMDP6D

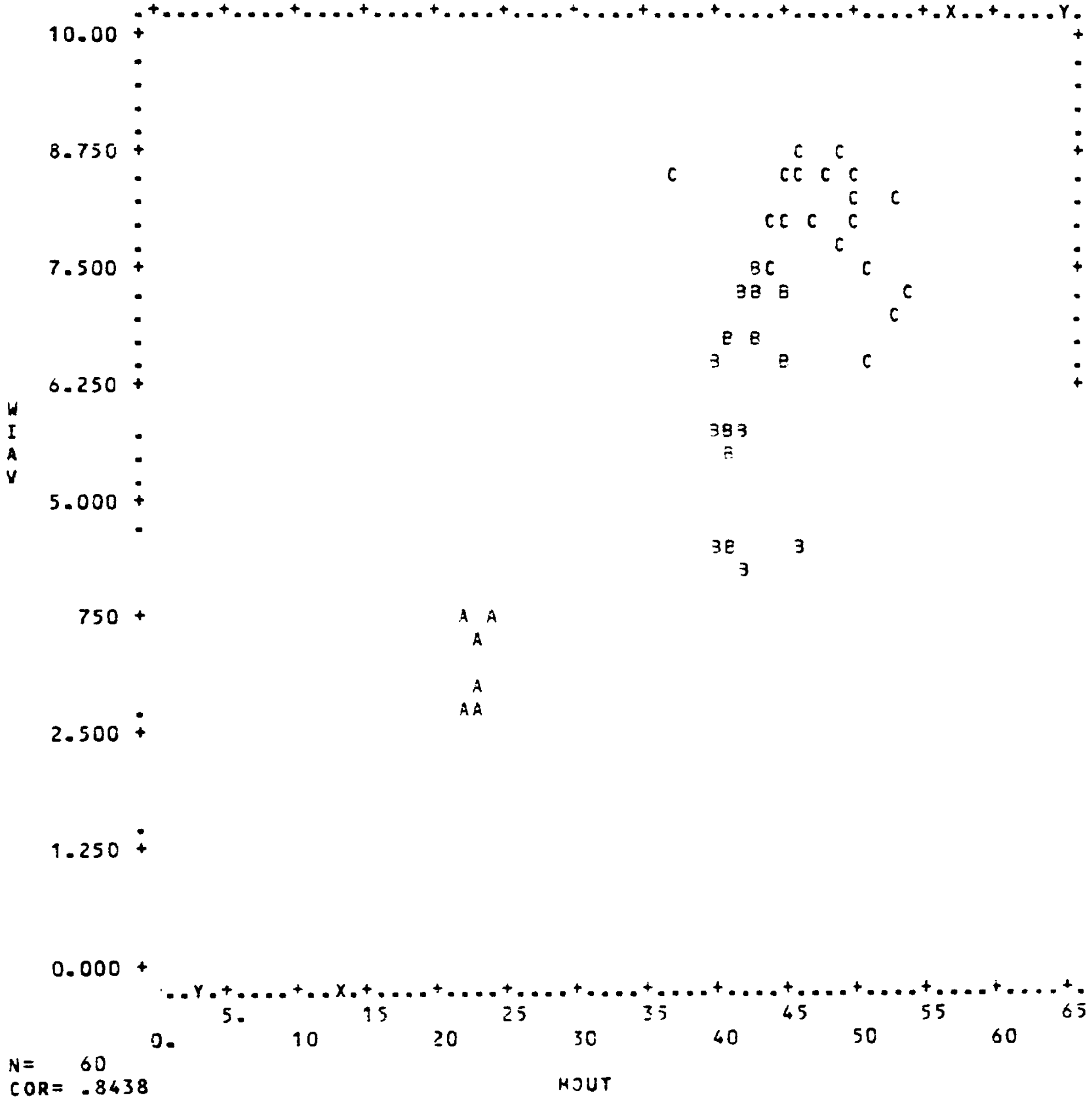


N= 60
COR= .3258

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	2163.9	1773.9	X= 655.30*Y-1556.4	1013E3
Y	5.6729	2.2333	Y= .00104*X+ 3.4226	1.6141

VARIABLE 14 HOID VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 14 HOID VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 14 HOID VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 12 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

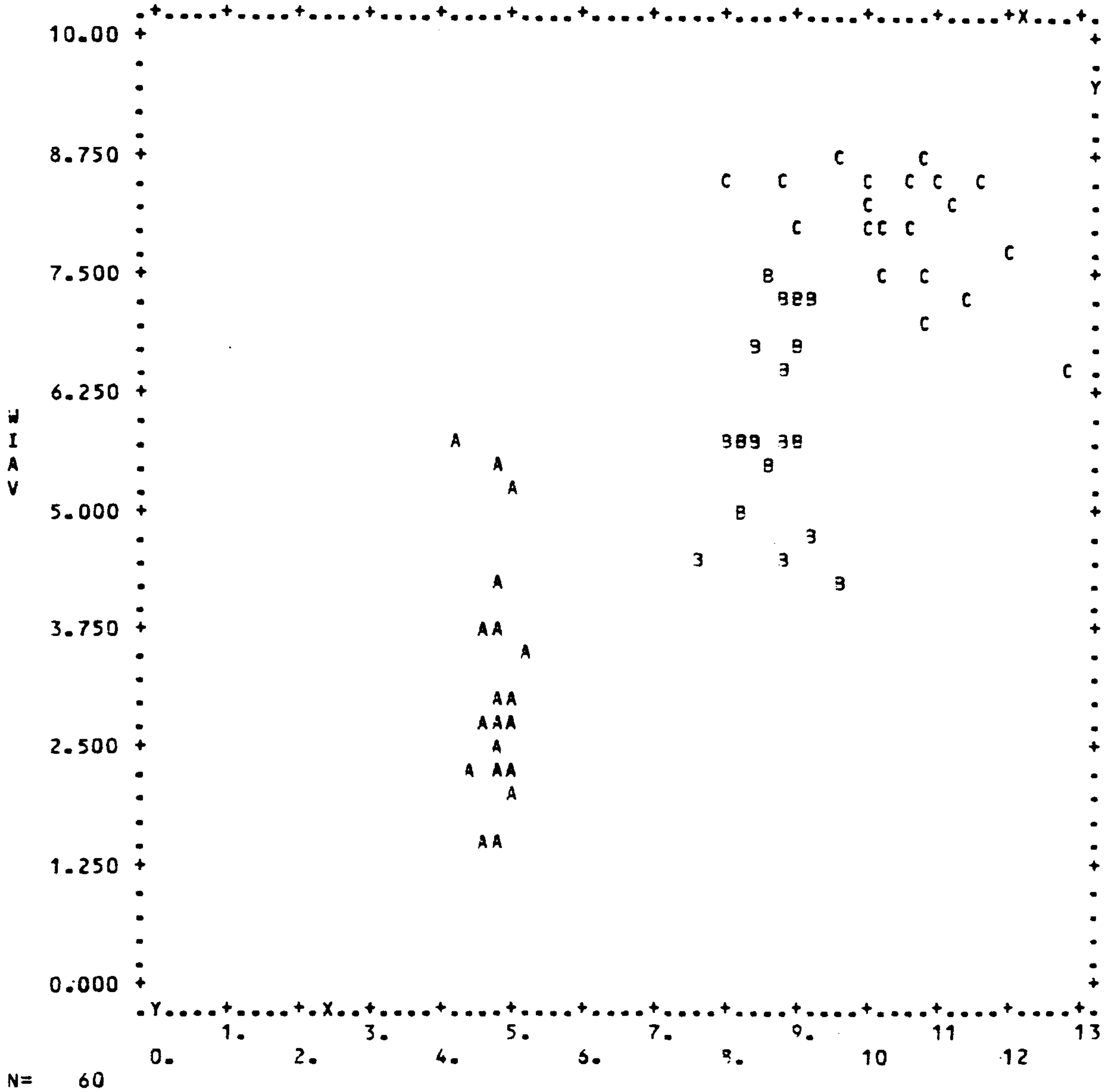


N= 60
COR= .8438

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
37.633	11.068	X= 4.1808*Y+ 13.916	35.396
5.6729	2.2338	Y= .17029*X-.73572	1.4621

VARIABLE 15 HOUT VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 15 HOUT VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 15 HOUT VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 13 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

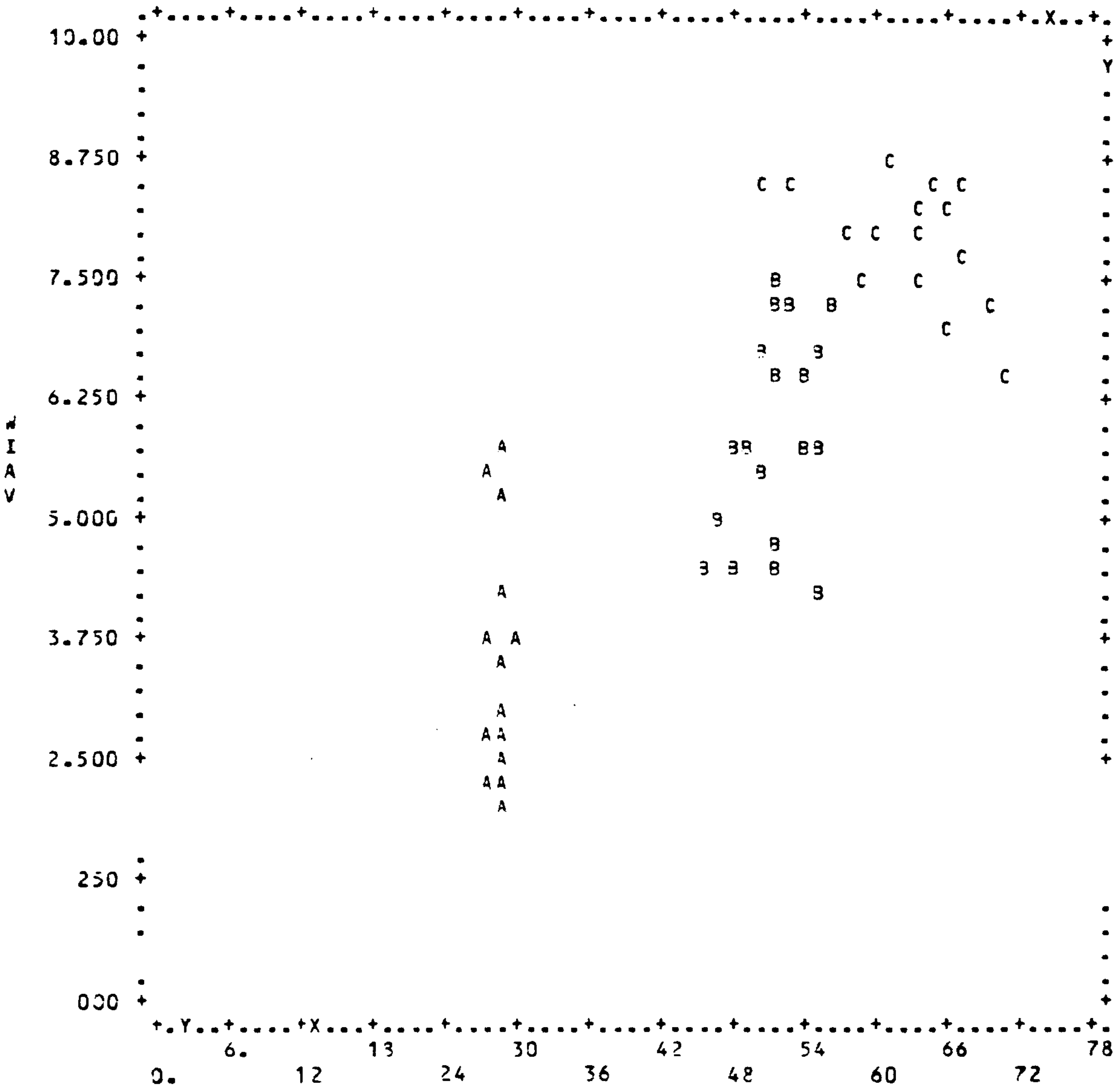


N= 60
COR= .8372

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	7.9810	2.4941	X= .93473*Y+ 2.6784	1.8931
Y	5.6729	2.2338	Y= .74978*X-.31109	1.5185

VARIABLE 16 NIAC VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 16 NIAC VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 16 NIAC VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 14 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

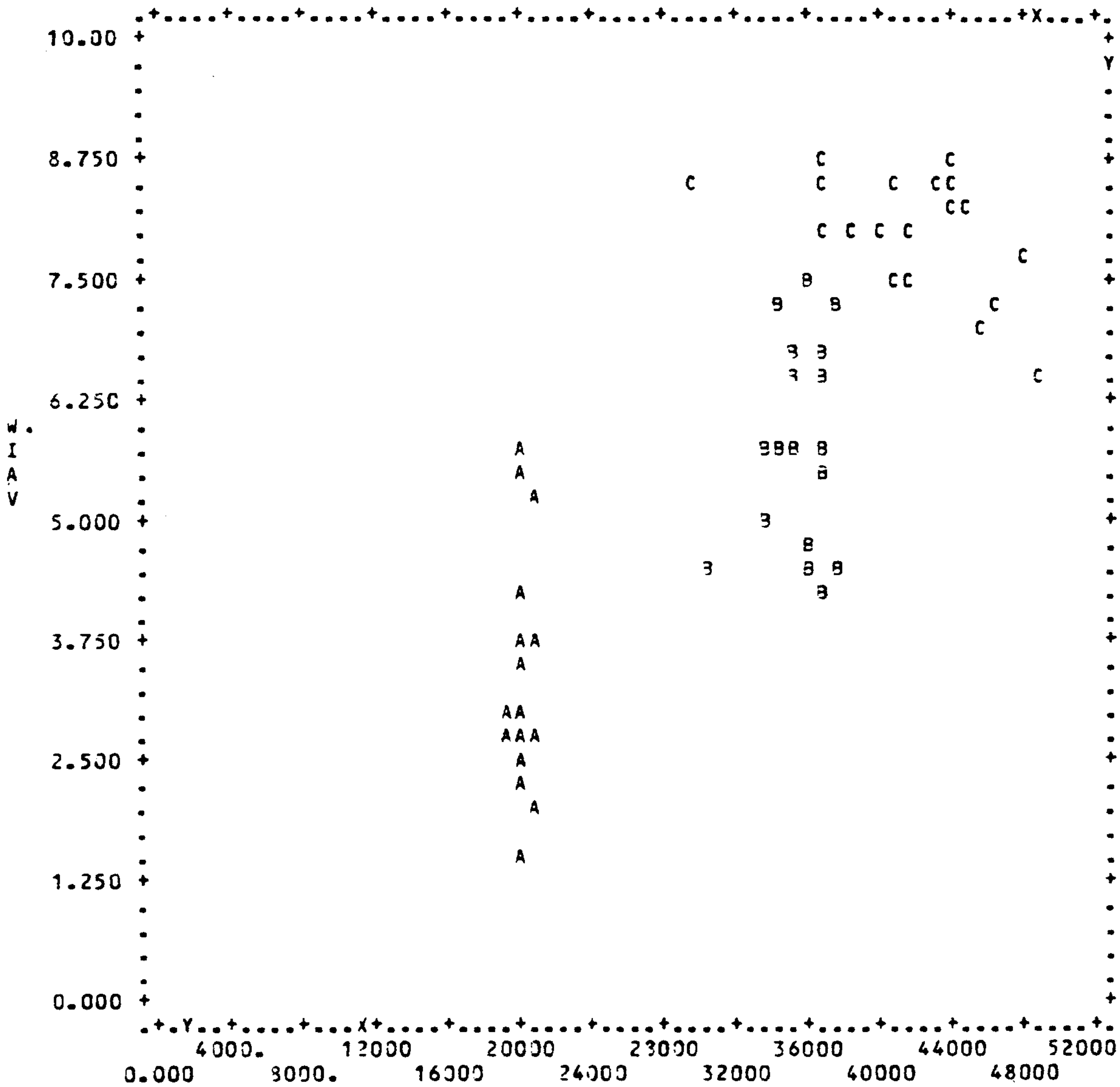


N= 60
COR= .8666

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	47.667	14.745	X= 5.7203*Y+ 15.216	55.067
Y	5.6729	2.2338	Y= .13129*X-.58519	1.2639

VARIABLE 17 NFLT VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 17 NFLT VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 17 NFLT VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 15 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



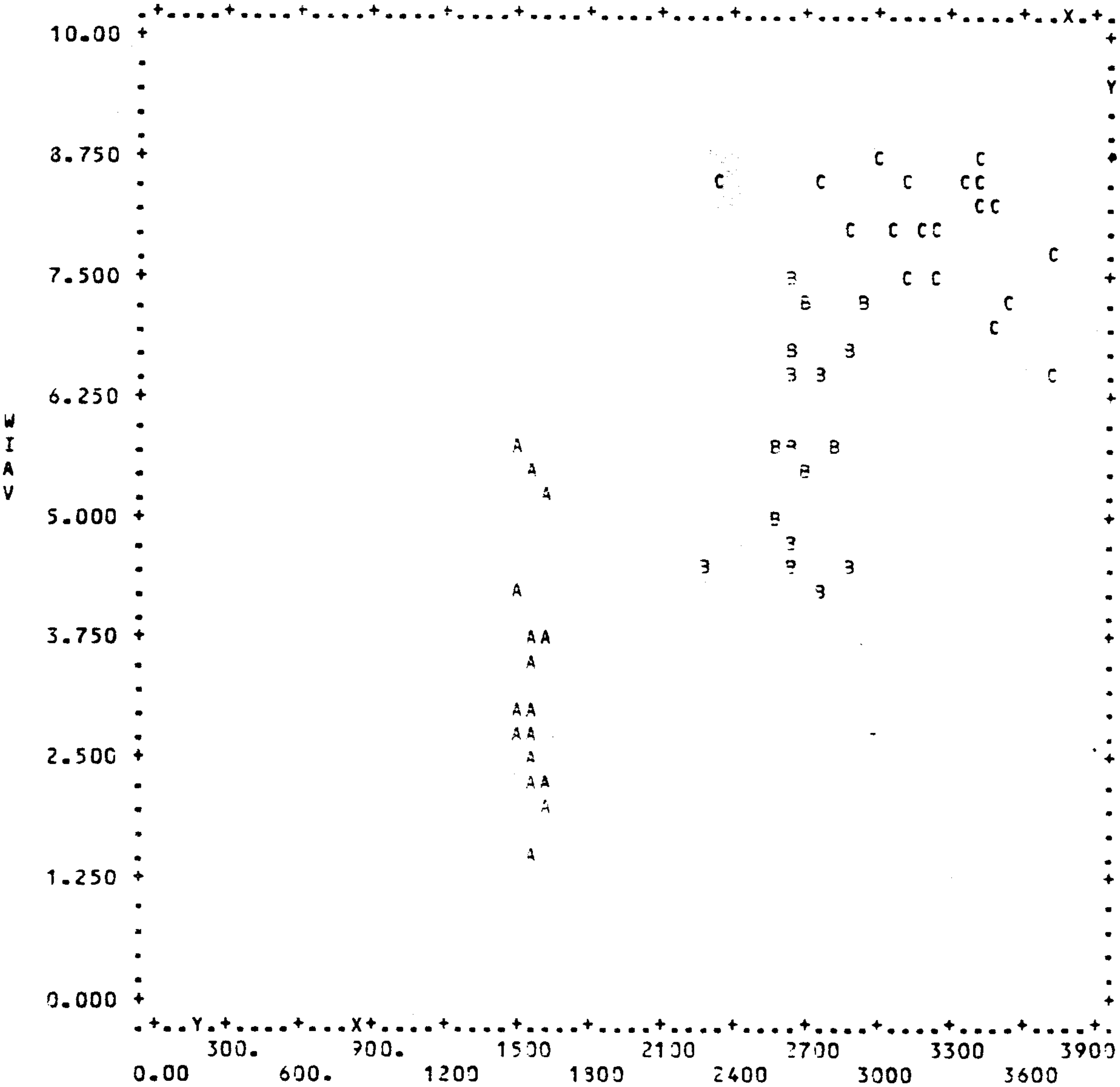
N= 60
COR= .3312

DFLT

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	32377.	9570.6	X= 3551.2*Y+ 12175.	2880E4
Y	5.6729	2.2338	Y= 194E-6*X-.60840	1.5691

VARIABLE	18	DFLT	VERSUS VARIABLE	39	WIAV	FOR GROUP A.SAMPLE	SYMBOL=A
VARIABLE	18	DFLT	VERSUS VARIABLE	39	WIAV	FOR GROUP B.SAMPLE	SYMBOL=B
VARIABLE	18	DFLT	VERSUS VARIABLE	39	WIAV	FOR GROUP C.SAMPLE	SYMBOL=C

PAGE 16 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



N= 60
COR= .8447

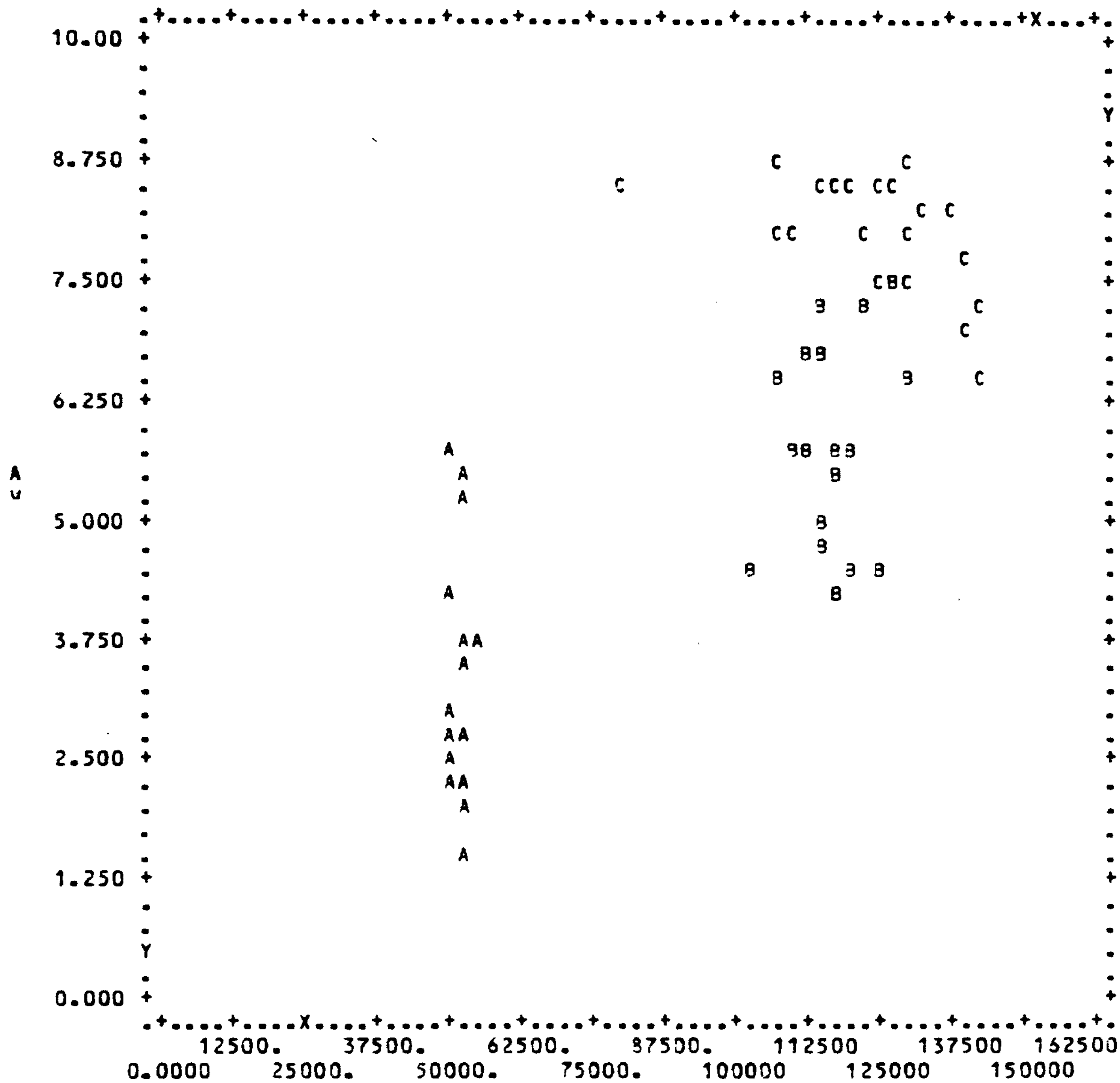
DIST

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	2489.7	731.95	X= 276.77*Y+ 919.58	156153
Y	5.6729	2.2338	Y= .00258*X-.74509	1.4544

VARIABLE 19 DIST VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 19 DIST VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 19 DIST VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

10/18/83 14:20:48 Q6DRUN01 JOB# 0017 OOOSOPER GOULD S.E.L. MPX-32 2

PAGE 17 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



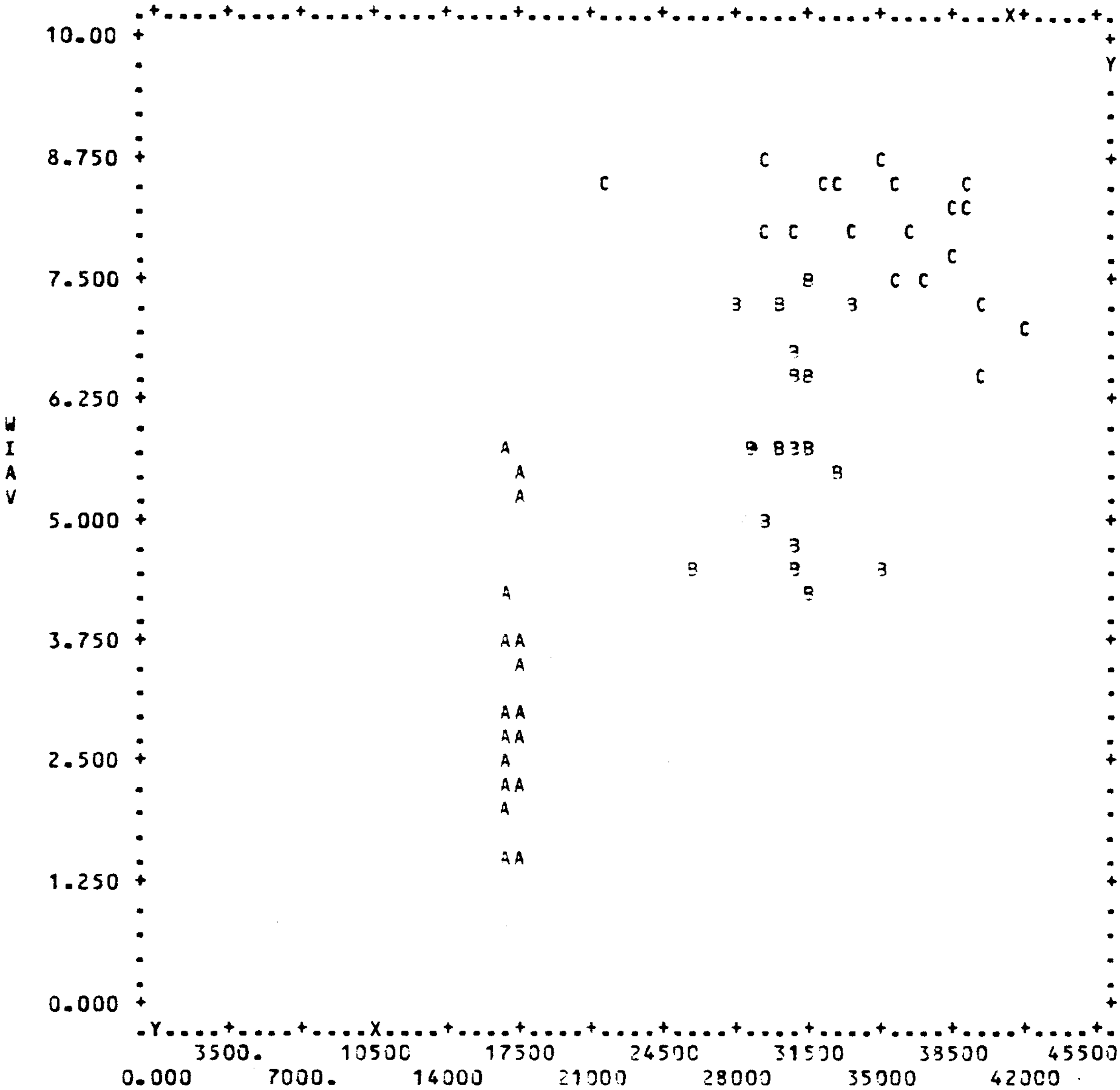
N= 60
COR= .7915

FUEL

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	97507.	34243.	X= 12133.*Y+ 28679.	4456E5
Y	5.6729	2.2338	Y= 516E-7*X+ .63853	1.8962

VARIABLE 20 FUEL VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 20 FUEL VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 20 FUEL VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 18 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

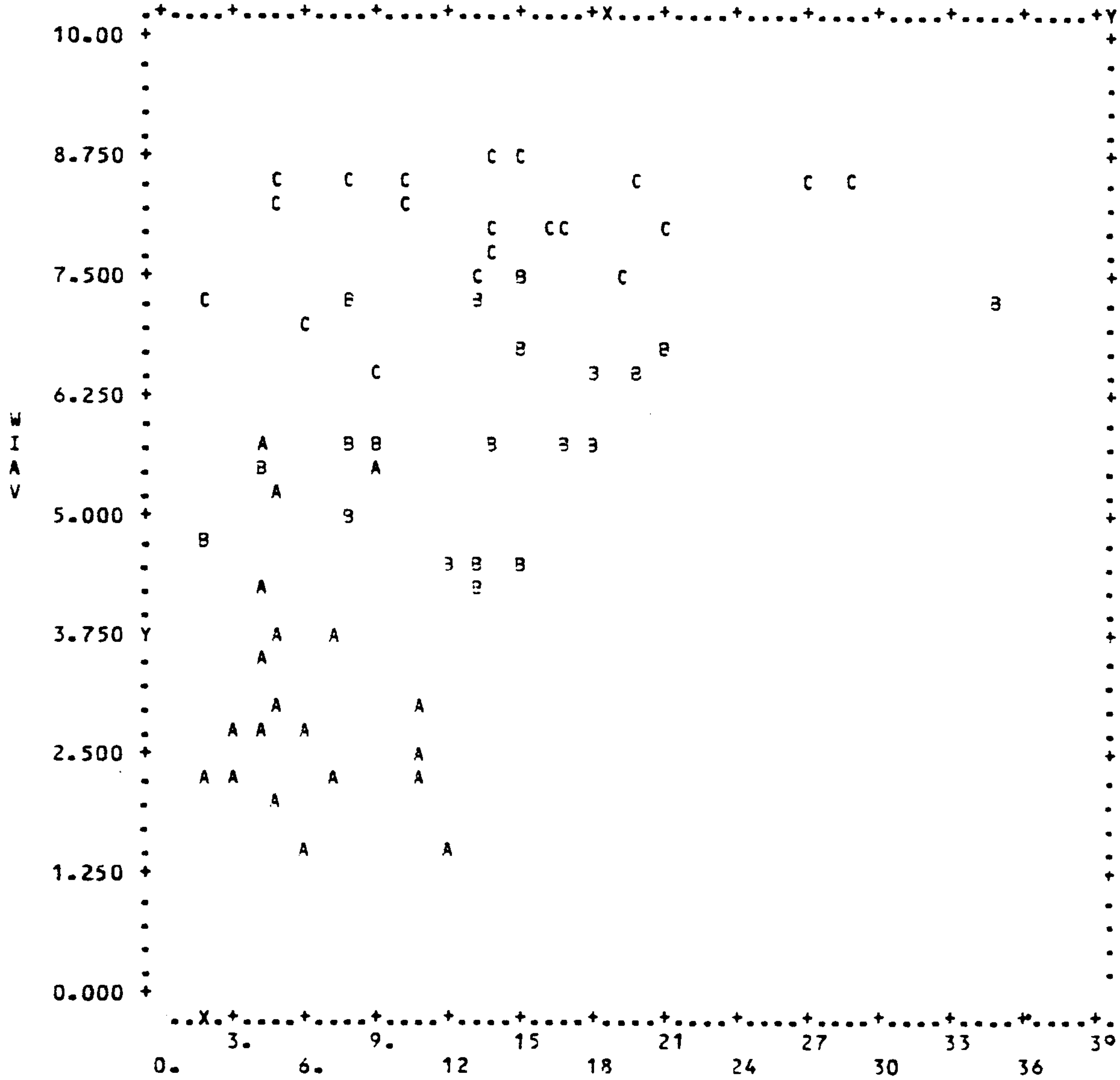


N= 60
COR= .7943 DIN8

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	27691.	8276.3	X= 2943.0*Y+ 10996.	2571E4
Y	5.6729	2.2338	Y= 214E-6*X-.26404	1.8731

VARIABLE 21 DIN8 VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 21 DIN8 VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 21 DIN8 VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 19 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

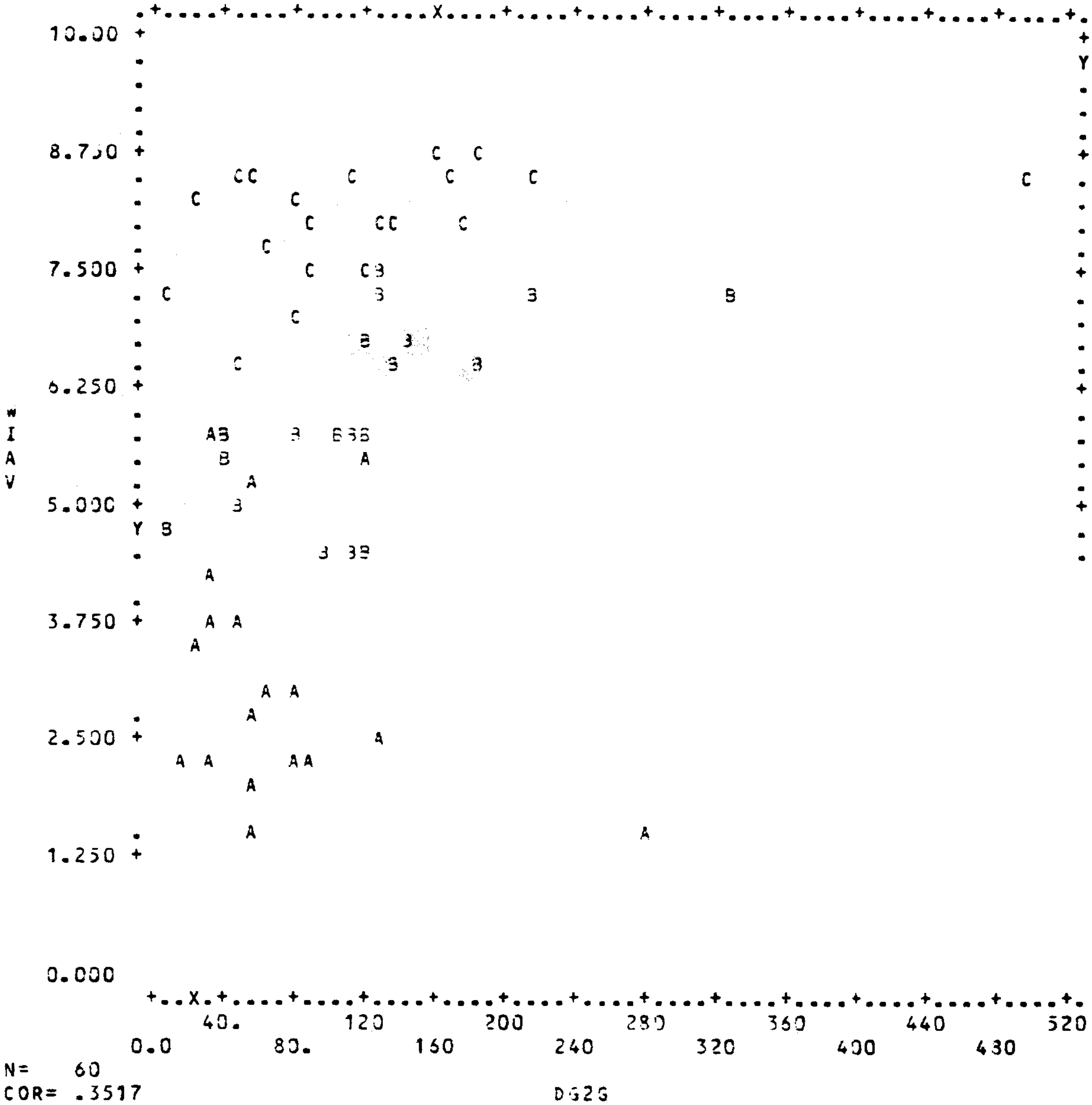


N= 60
COR= .4956

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	11.267	6.9863	X= 1.5499*Y+ 2.4742	37.457
Y	5.6729	2.2338	Y= .15845*X+ 3.8877	3.8294

VARIABLE 22 NG2G VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 22 NG2G VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 22 NG2G VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 20 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



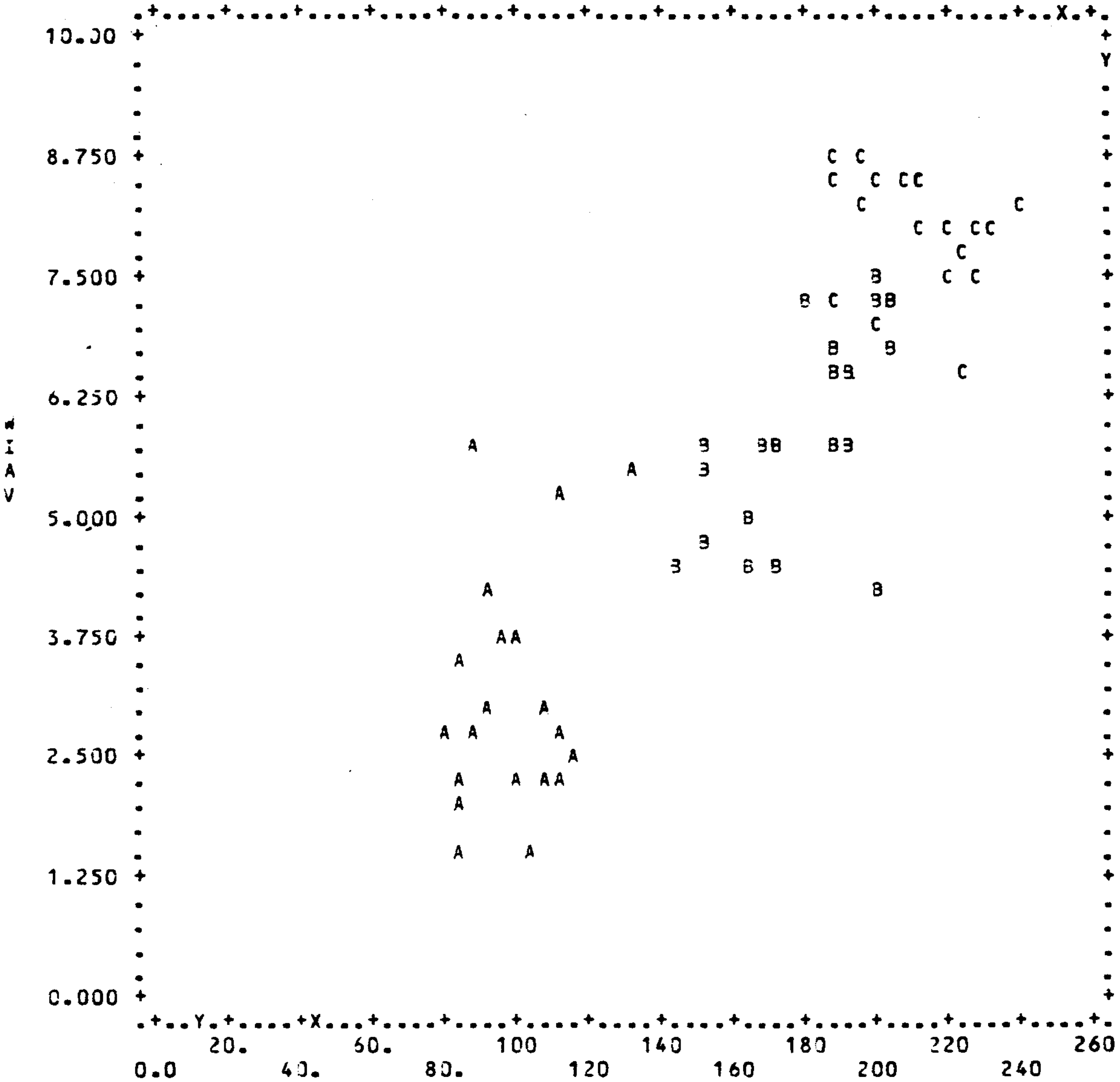
N= 60
COR= .3517

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	102.68	83.199	X= 13.100*Y+ 28.366	6170.4
Y	5.6729	2.2333	Y= .00944*X+ 4.7032	4.4480

VARIABLE 23 DG2G VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 23 DG2G VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 23 DG2G VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

10/18/83 14:20:43 Q6DRUN01 JOB# 0017 000SOPER GOULD S.E.L. MPX-32 2

PAGE 21 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



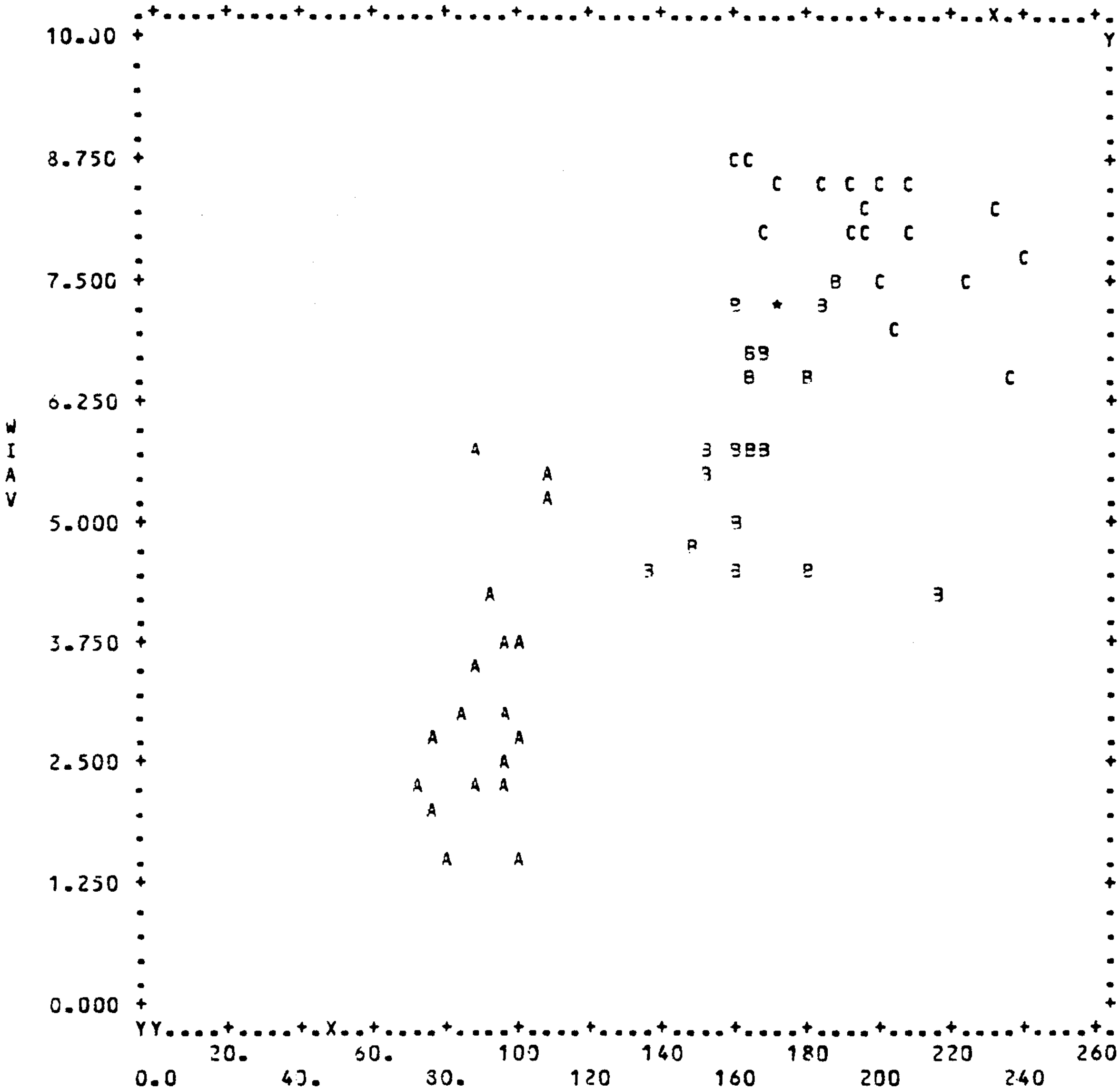
N= 60
COR= .8789

NG2A

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	162.32	50.317	X= 19.797*Y+ 50.009	586.03
Y	5.6729	2.2333	Y= .03902*X-.66044	1.1550

VARIABLE 24 NG2A VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 24 NG2A VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 24 NG2A VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 23 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

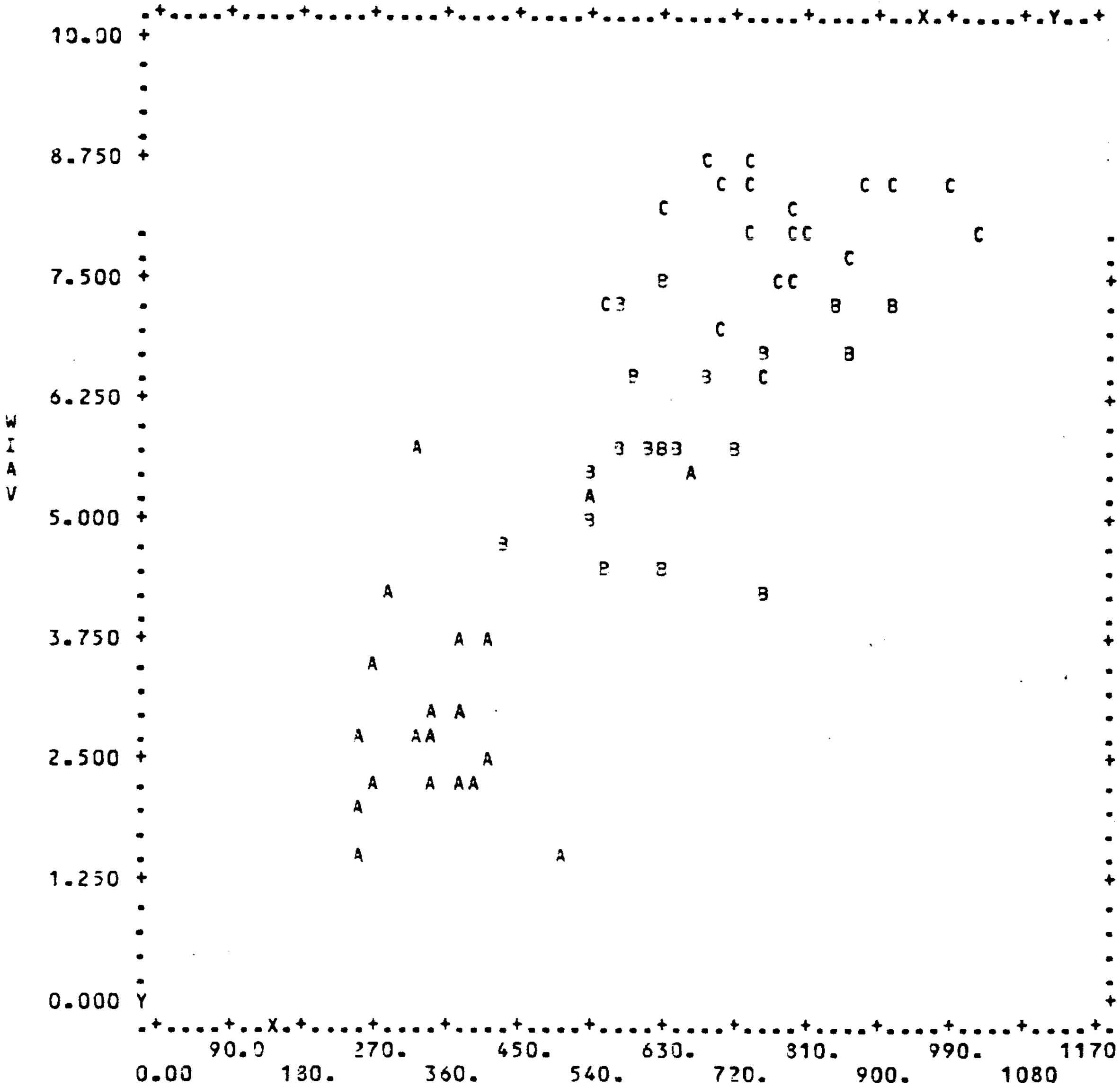


N= 60
COR= .8292

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	151.25	47.690	X= 17.704*Y+ 50.819	722.65
Y	5.6729	2.2338	Y= .03884*X-.20194	1.5855

VARIABLE 26 NAIR VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 26 NAIR VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 26 NAIR VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 22 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



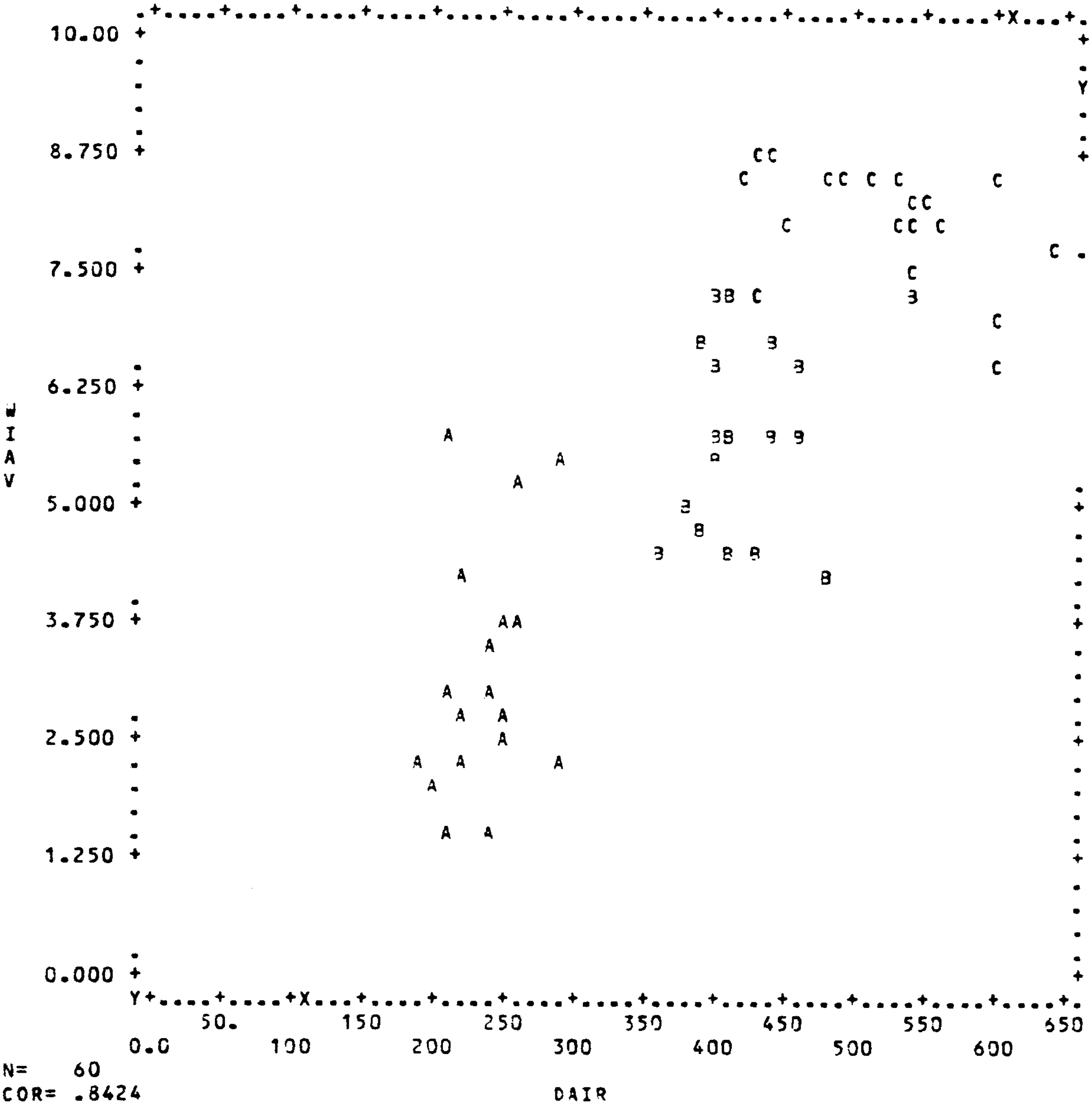
N= 63
COR= .8400

DG2A

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
600.30	207.86	X= 78.161*Y+ 157.40	12939.
5.6729	2.2338	Y= .00903*X+ .24932	1.4945

VARIABLE 25 DG2A VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 25 DG2A VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 25 DG2A VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 24 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

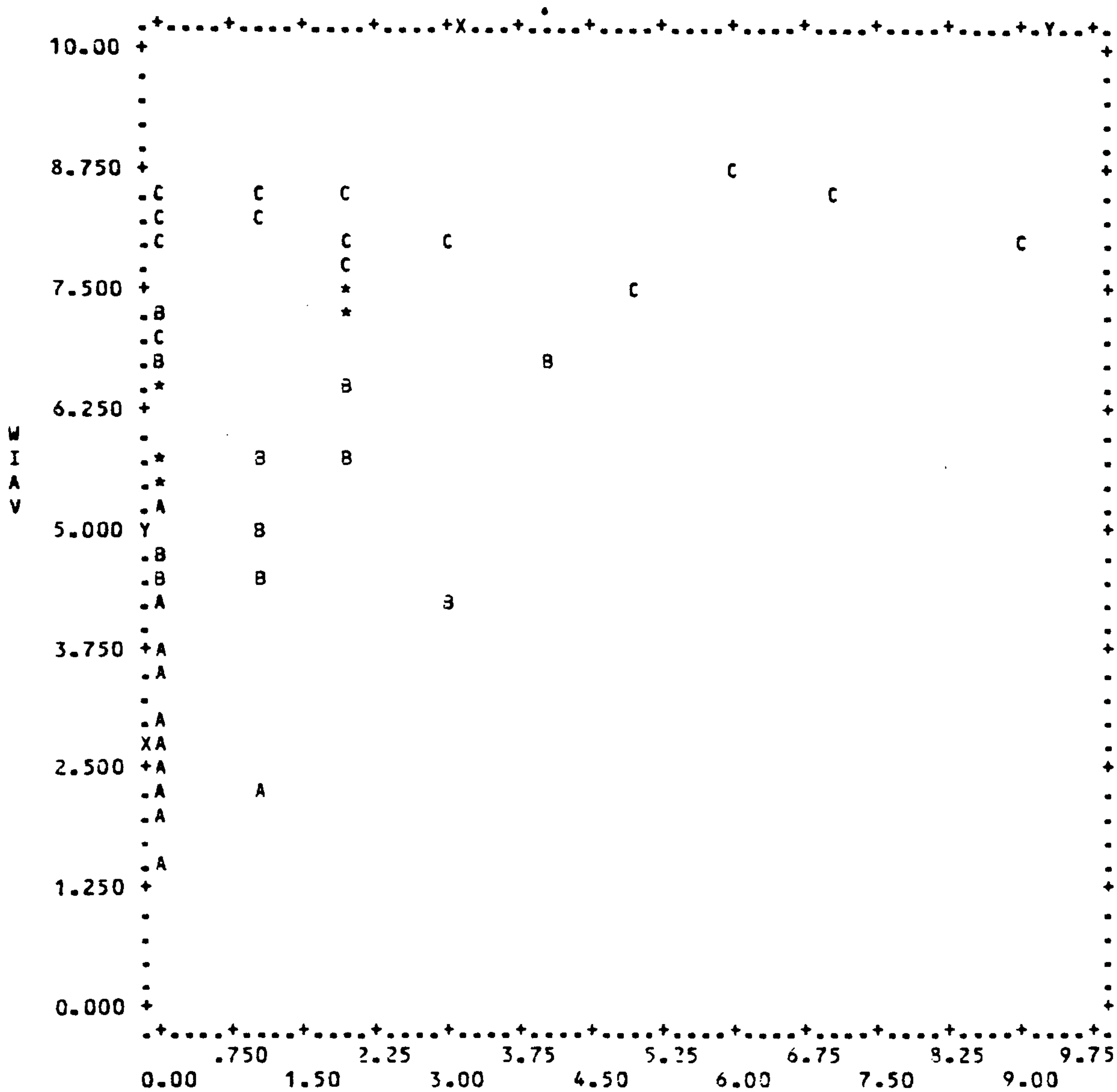


N= 60
COR= .8424

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
392.53	127.70	X= 48.158*Y+ 119.34	4816.6
5.6729	2.2338	Y= .01474*X-.11137	1.4738

VARIABLE 27 DAIR VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 27 DAIR VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 27 DAIR VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 25 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

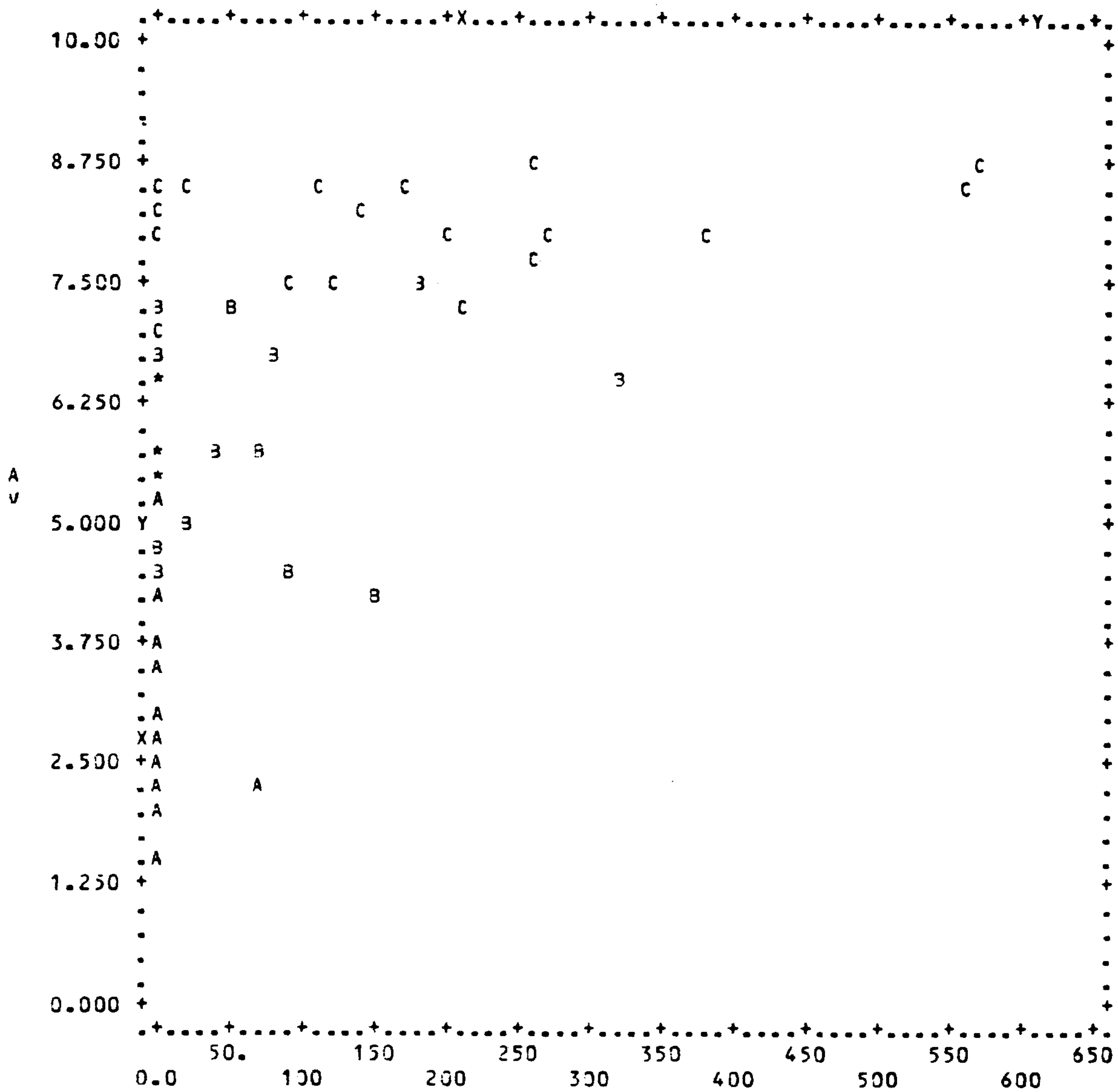


N= 60
COR= .4950

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	1.1500	1.9643	X= .43523*Y-1.3191	2.9635
Y	5.6729	2.2338	Y= .56286*X+ 5.0256	3.8325

VARIABLE 28 NDLY VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 28 NDLY VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 28 NDLY VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 26 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

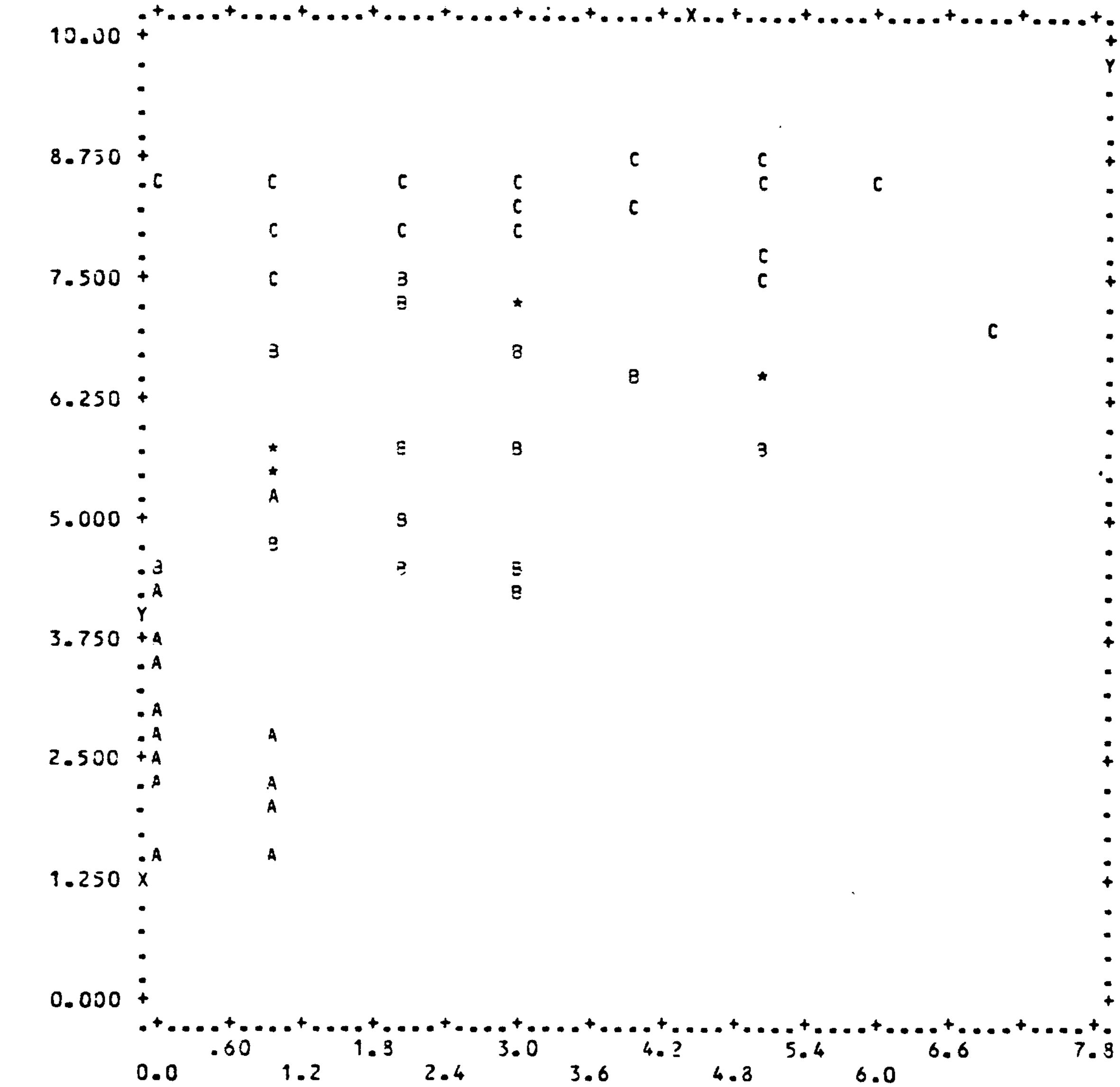


N= 60
COR= .4986

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	73.550	131.21	X= 29.238*Y-92.600	13160.
Y	5.6729	2.2333	Y= .00849*X+ 5.0436	3.8140

VARIABLE 29 DDLY VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 29 DDLY VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 29 DDLY VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 27 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

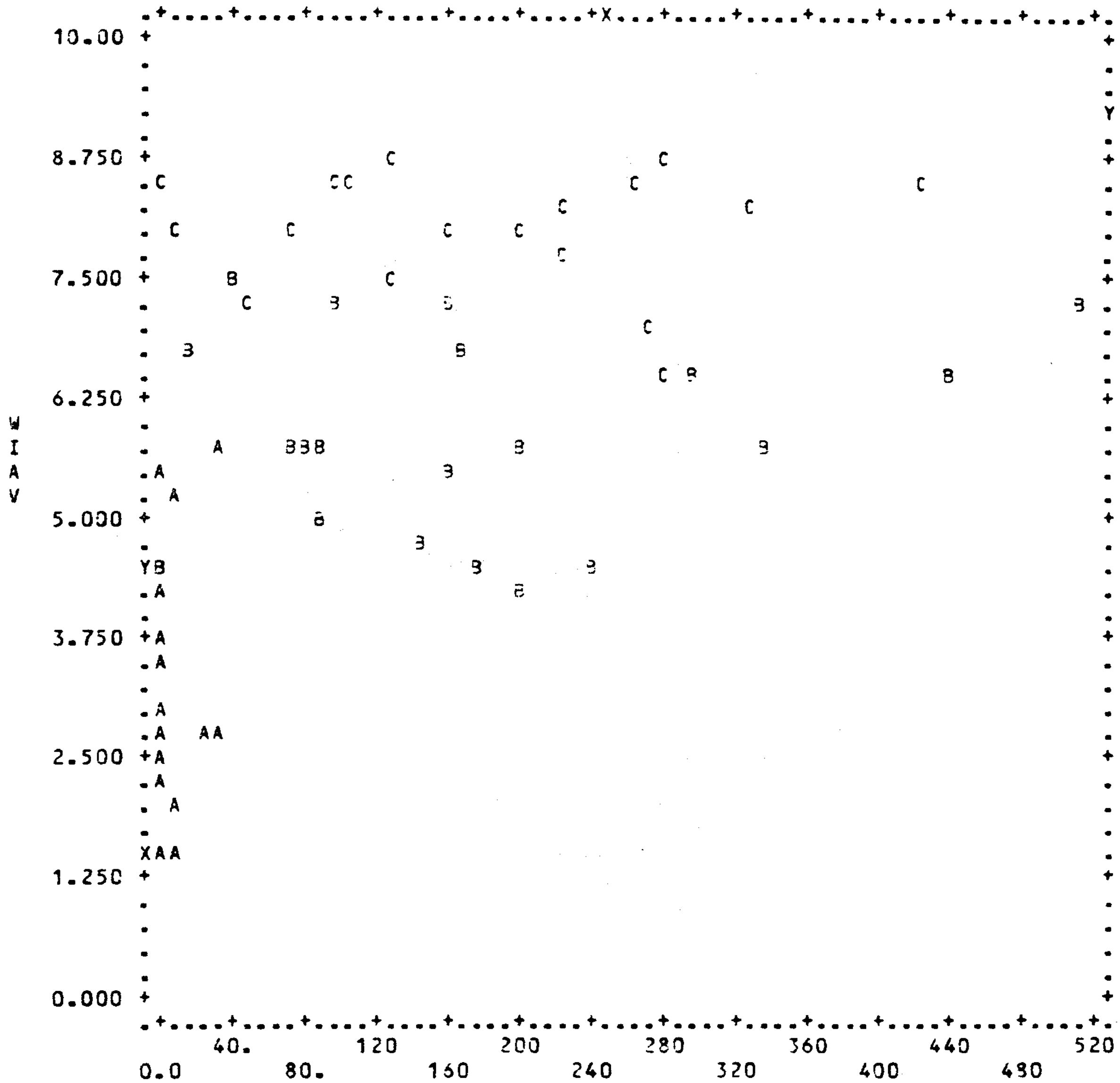


N= 60
COR= .5997 NSCF

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	2.0833	1.8712	X= .50235*Y-.76647	2.2808
Y	5.6729	2.2338	Y= .71591*X+ 4.1814	3.2505

VARIABLE 30 NSCF VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 30 NSCF VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 30 NSCF VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 28 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



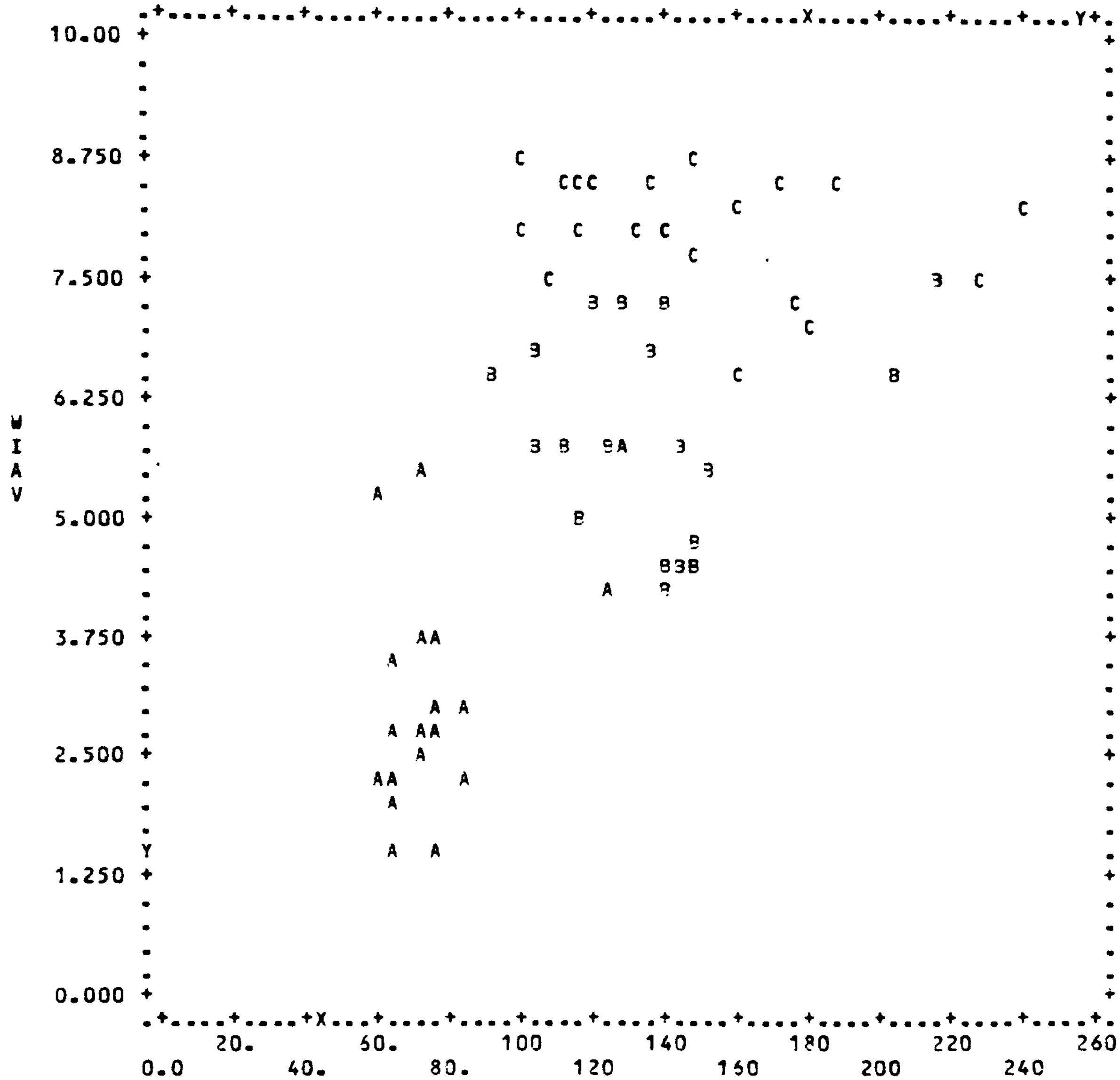
N= 60
COR= .5142

D5CF

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
116.28	128.68	X= 29.620*Y-51.746	12392.
5.6729	2.2338	Y= .00893*X+ 4.6351	3.7341

VARIABLE 31 D5CF VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 31 D5CF VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 31 D5CF VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 29 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



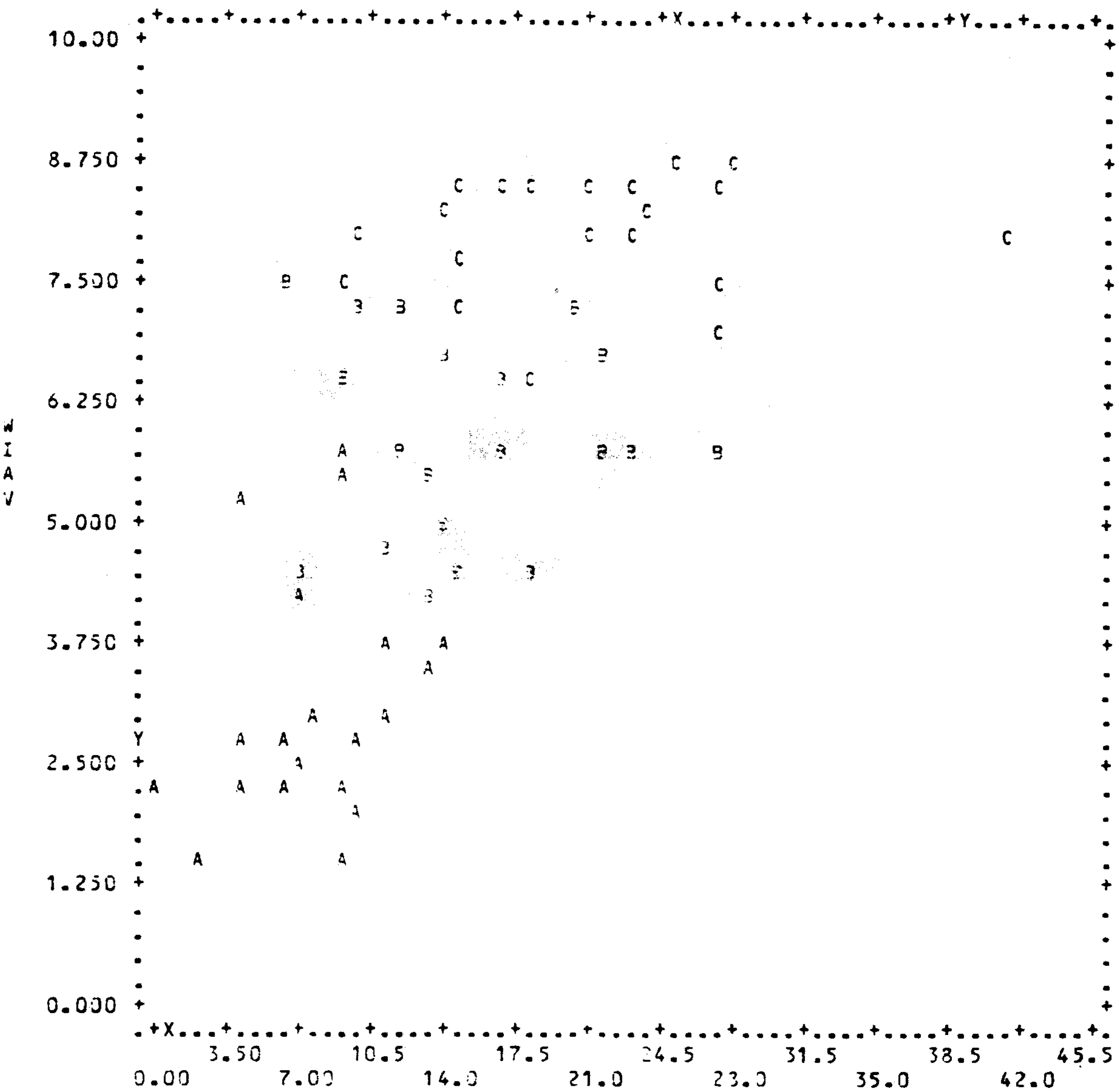
N= 60
COR= .6580

CKEN

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	119.73	44.255	X= 13.035*Y+ 45.736	1129.8
Y	5.6729	2.2338	Y= .03321*X+ 1.6964	2.8785

VARIABLE 36 CKEN VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 36 CKEN VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 36 CKEN VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 30 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



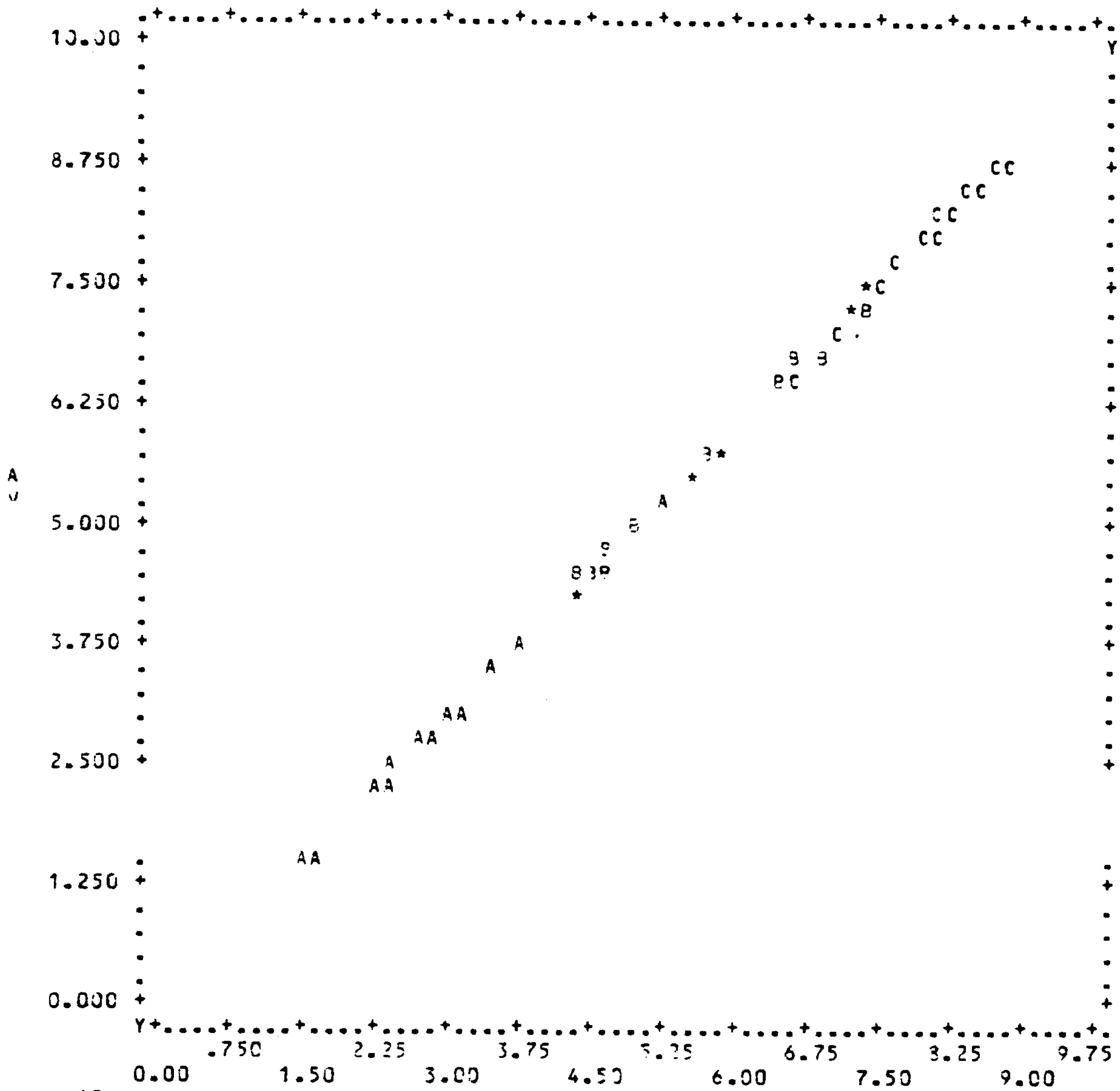
N= 60
COR= .6533

CKER

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	14.550	7.8706	X= 2.3017*Y+ 1.4924	36.122
Y	5.5729	2.2338	Y= .13541*X+ 2.9752	2.9097

VARIABLE 37 CKER VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 37 CKER VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 37 CKER VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 31 WORKLOAD PROBE - 50 MIN. CUM. DATA. BMDP6D

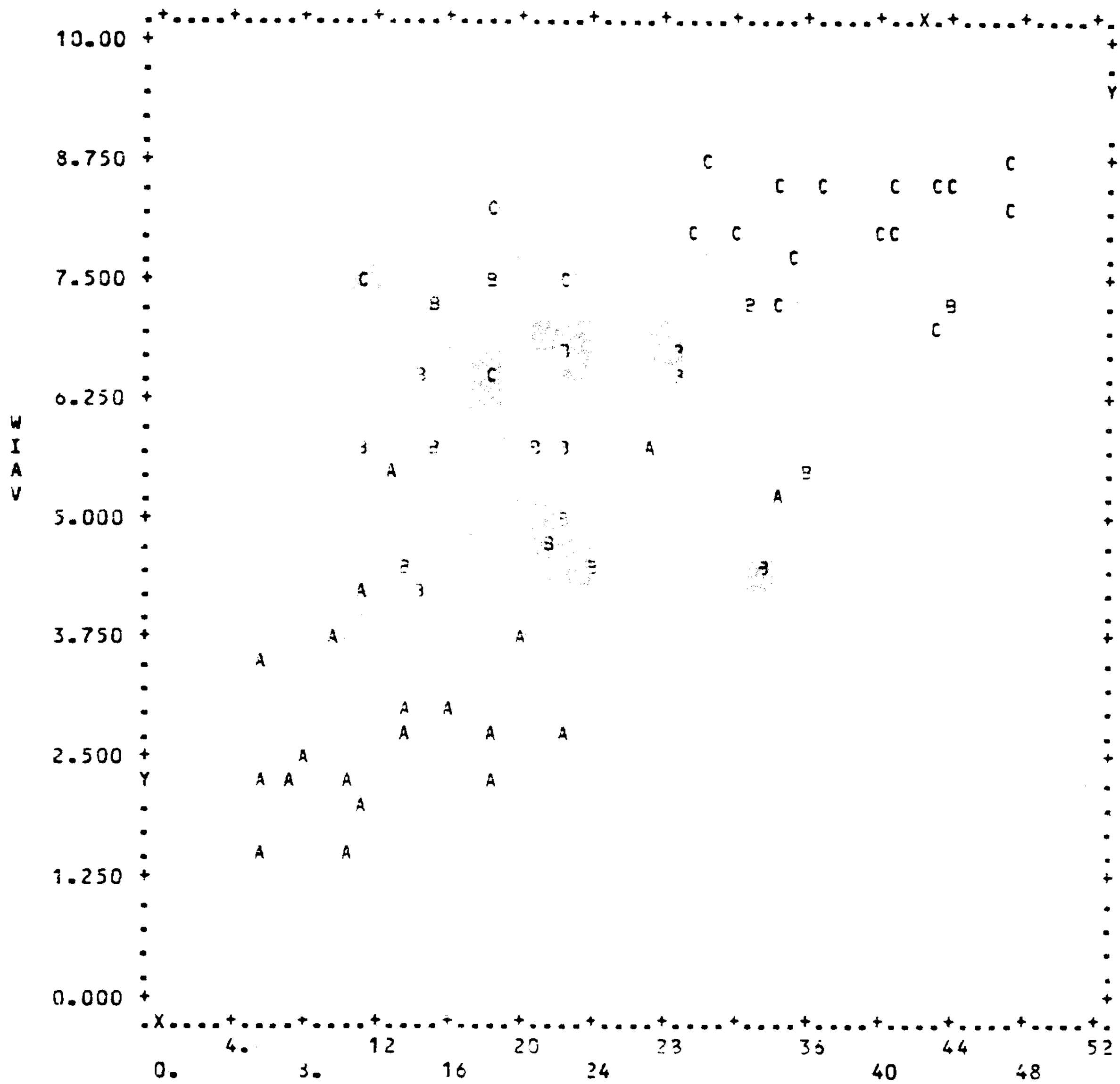


N= 60
COR= 1.000

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	5.6729	2.2339	X= 1.0000*Y+ 0.0000	0.0000
Y	5.6729	2.2339	Y= 1.0000*X+ 0.0000	0.0000

VARIABLE 39 WIAV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 39 WIAV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 39 WIAV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 32 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



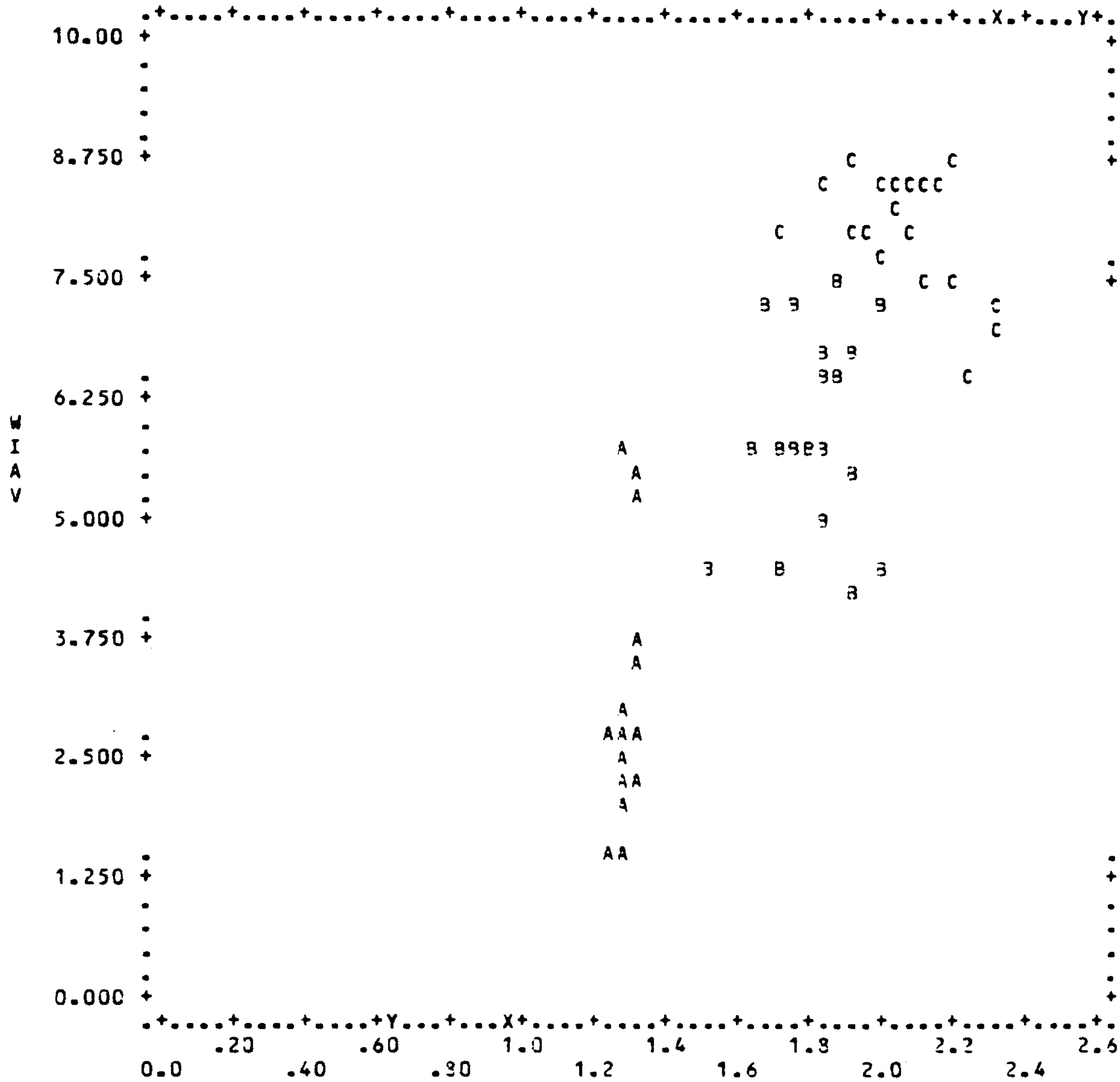
N= 60
COR= .7364

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	23.955	12.201	X= 4.0222*Y+ 1.1370	69.310
Y	5.6729	2.2338	Y= .13433*X+ 2.4432	2.3233

VARIABLE 42 WDAV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 42 WDAV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 42 WDAV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

10/18/83 14:20:43 Q6DRUN01 JOB# 0017 OOOSOPER GOULD S.E.L. MPX-32 2.

PAGE 33 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

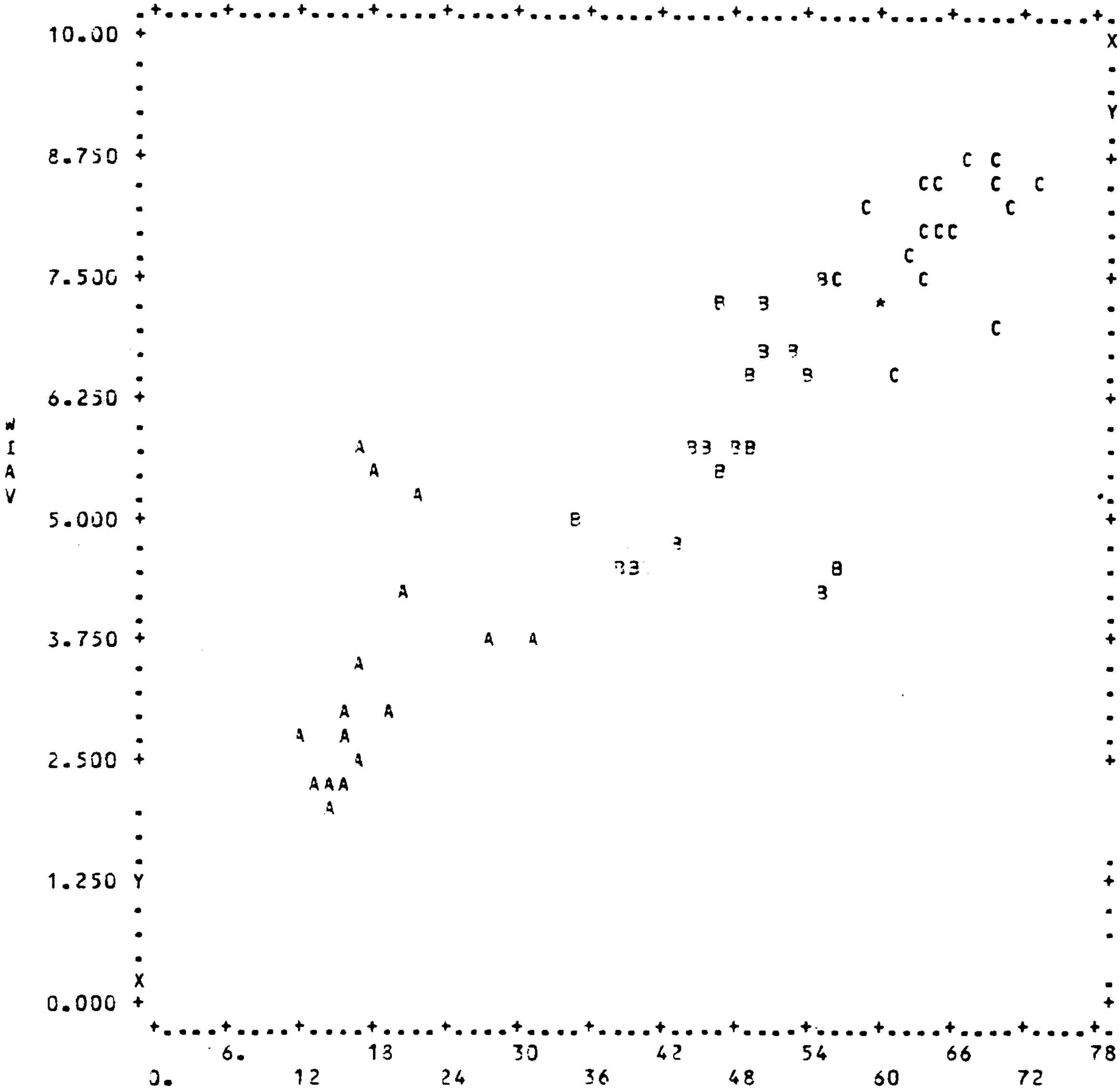


N= 60
COR= .8357 CMAV

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	1.7251	.33963	X= .12705*Y+ 1.0043	.03540
Y	5.6729	2.2333	Y= 5.4963*X-3.8088	1.5313

VARIABLE 44 CMAV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 44 CMAV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 44 CMAV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 34   WORKLOAD PROBE   60 MIN. CUM. DATA. BMDP6D

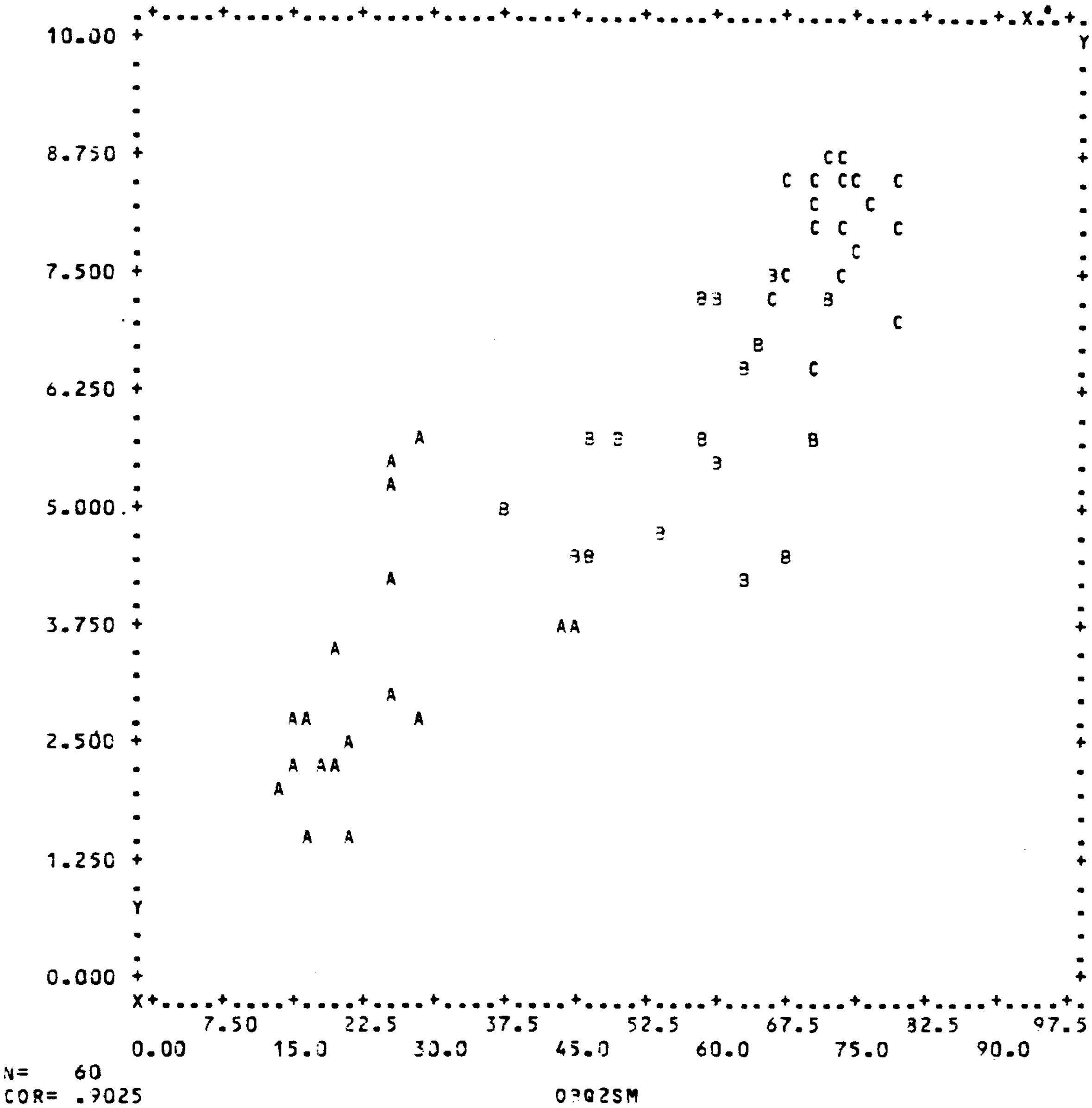


N= 60
COR= .9165

03Q1SM

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	43.433	20.519	X= 3.4133*Y-4.3230	63.566
Y	5.6729	2.2338	Y= .09977*X+ 1.3395	.81263

```
VARIABLE   46  03Q1SM   VERSUS VARIABLE   39  WIAV    FOR GROUP A.SAMPLE  SYMBOL=A
VARIABLE   46  03Q1SM   VERSUS VARIABLE   39  WIAV    FOR GROUP B.SAMPLE  SYMBOL=B
VARIABLE   46  03Q1SM   VERSUS VARIABLE   39  WIAV    FOR GROUP C.SAMPLE  SYMBOL=C
```

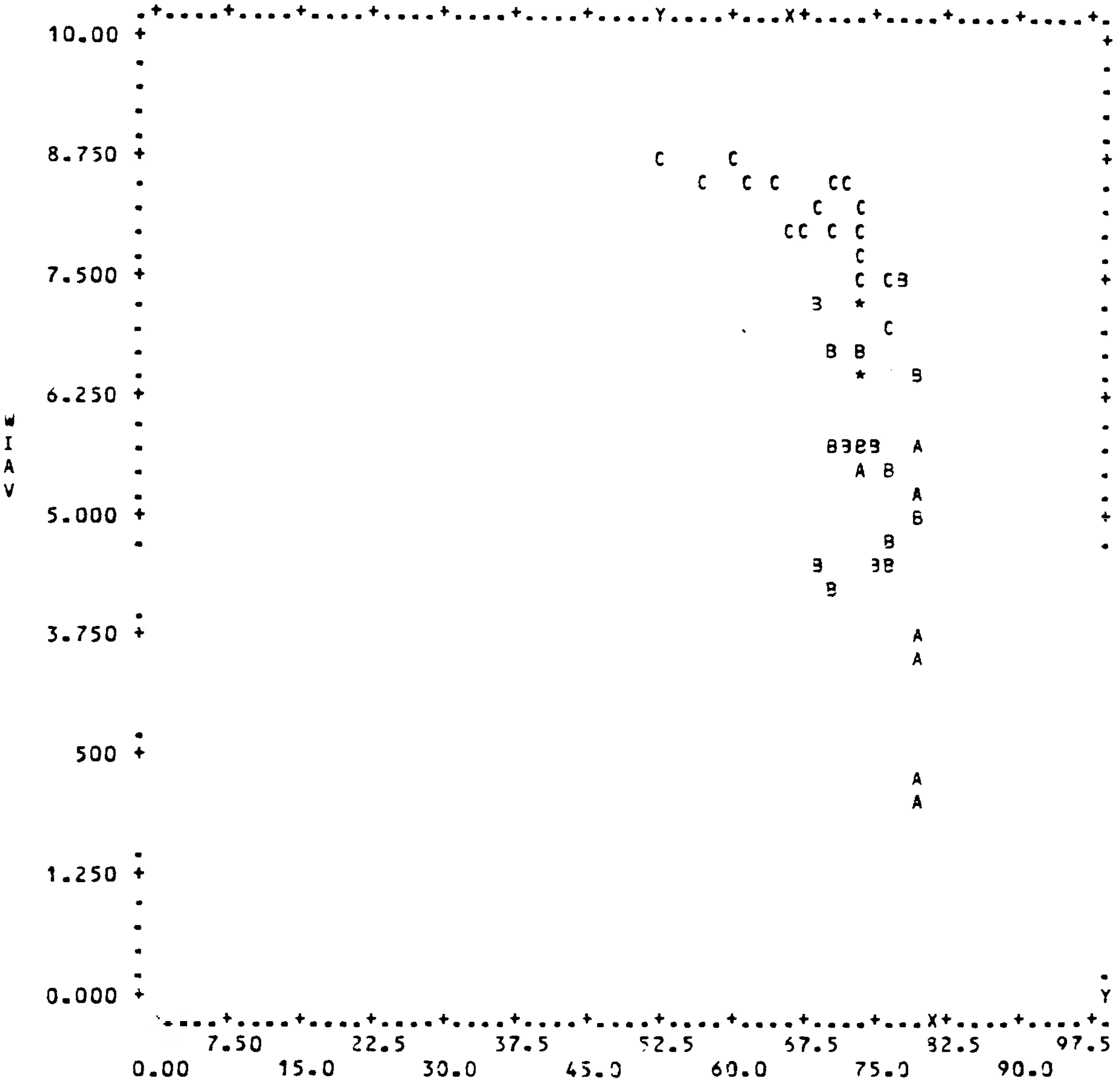


N= 60
COR= .9025

MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
51.733	22.352	X= 9.0304*Y+ .50461	94.301
5.6729	2.2338	Y= .09019*X+ 1.0071	.94183

```
VARIABLE   47  OBQ2SM   VERSUS VARIABLE   39  WIAV    FOR GROUP A.SAMPLE  SYMBOL=A
VARIABLE   47  OBQ2SM   VERSUS VARIABLE   39  WIAV    FOR GROUP B.SAMPLE  SYMBOL=B
VARIABLE   47  OBQ2SM   VERSUS VARIABLE   39  WIAV    FOR GROUP C.SAMPLE  SYMBOL=C
```

PAGE 36 WORKLOAD PROBE 60 MIN. CUM. DATA. BMDP6D



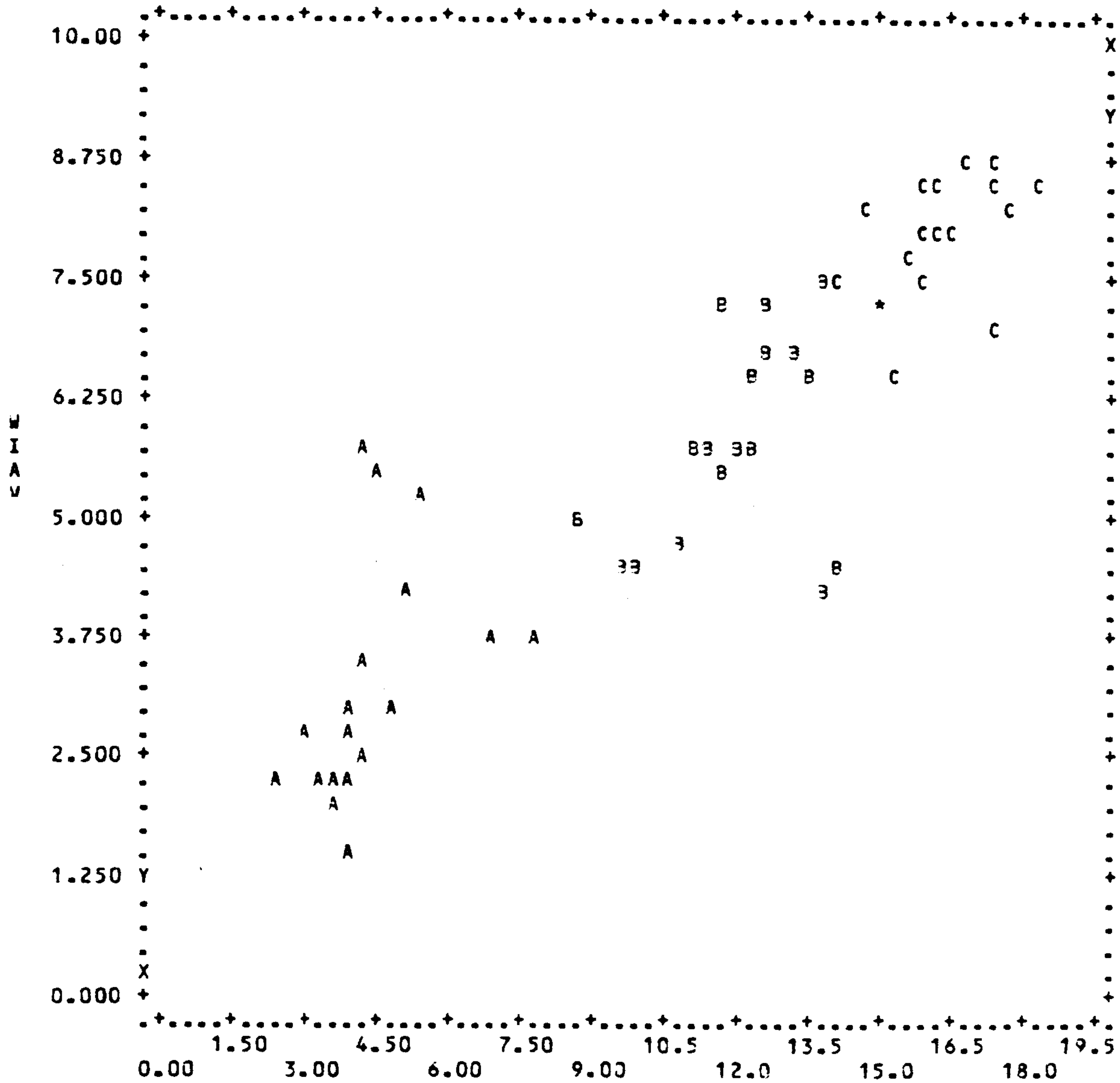
N= 60
COR=-.5663

0BQ4SM

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	73.100	5.7540	X=-1.4586*Y+ 81.375	22.880
Y	5.6729	2.2338	Y=-.21984*X+ 21.743	3.4483

VARIABLE 48 0BQ4SM VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 48 0BQ4SM VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 48 0BQ4SM VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 37 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



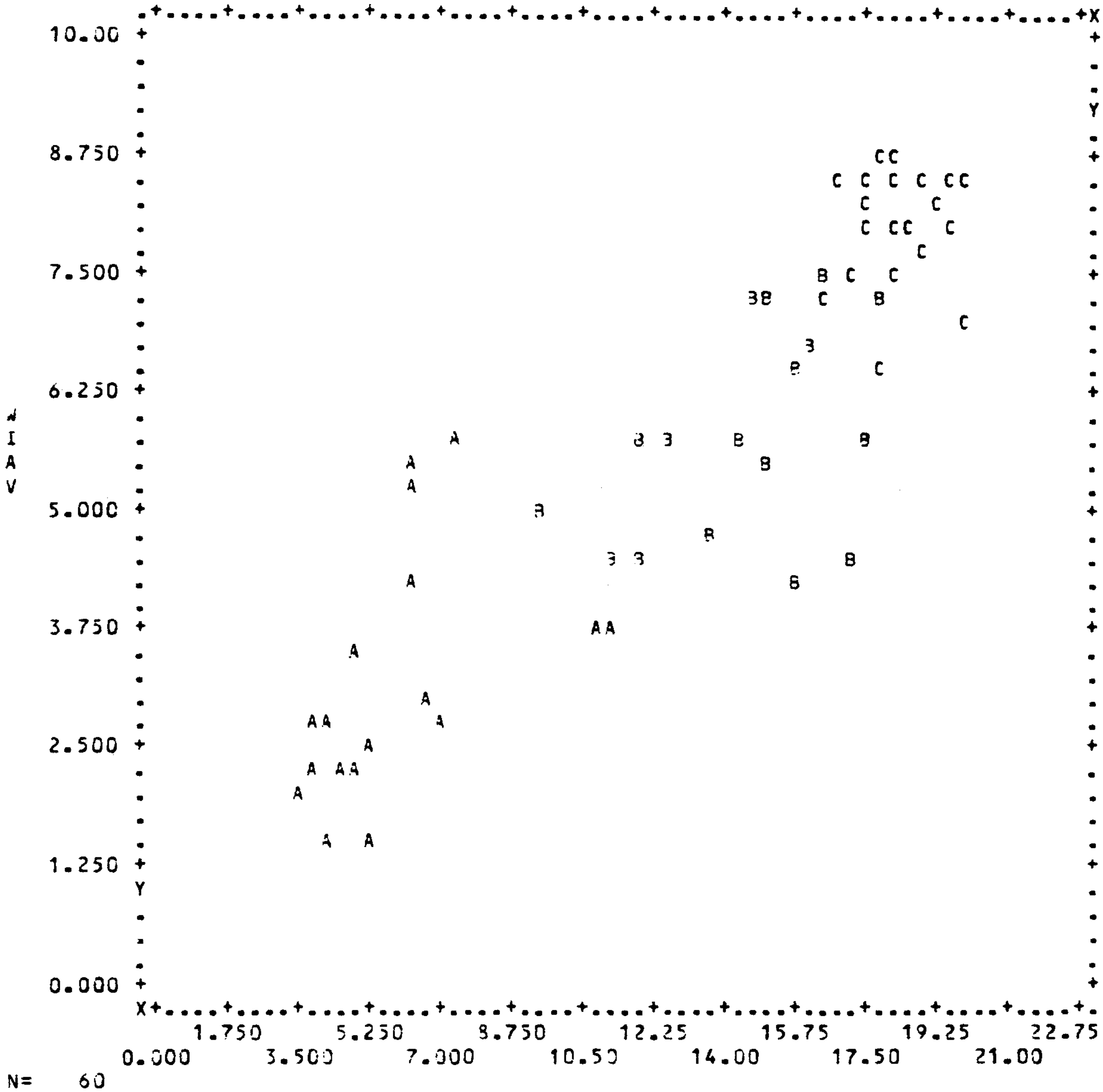
N= 60
COR= .9165

OBQ1AV

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	10.858	5.1298	X= 2.1046*Y-1.0807	4.2354
Y	5.6729	2.2338	Y= .39909*X+ 1.3395	.31264

VARIABLE 49 OBQ1AV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 49 OBQ1AV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 49 OBQ1AV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 38 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

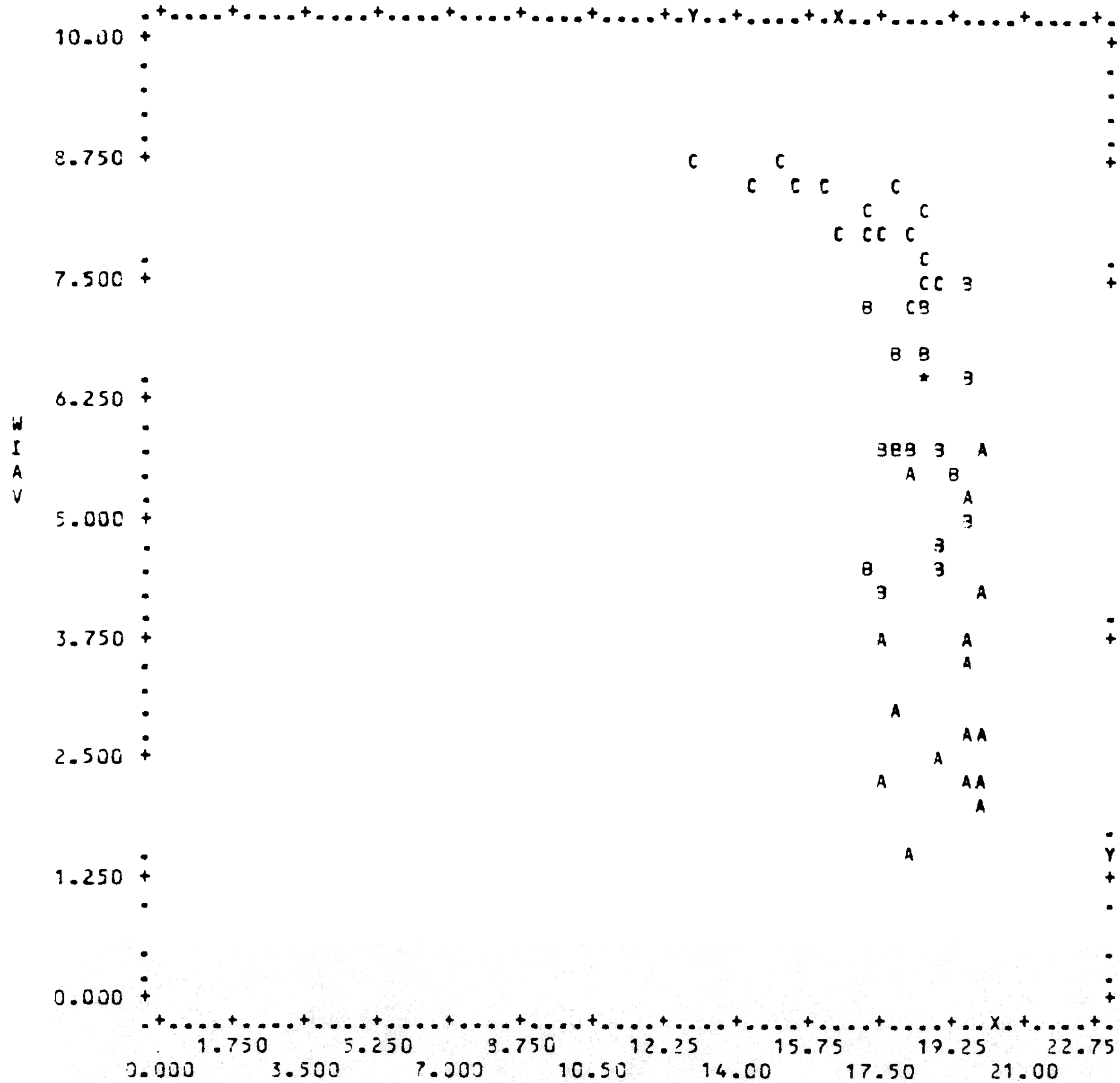


N= 60
COR= .9025

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	12.933	5.5331	X= 2.2576*Y+ .12616	5.8939
Y	5.6729	2.2333	Y= .36076*X+ 1.0071	.94183

VARIABLE 50 03Q2AV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 50 03Q2AV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 50 03Q2AV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 39 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

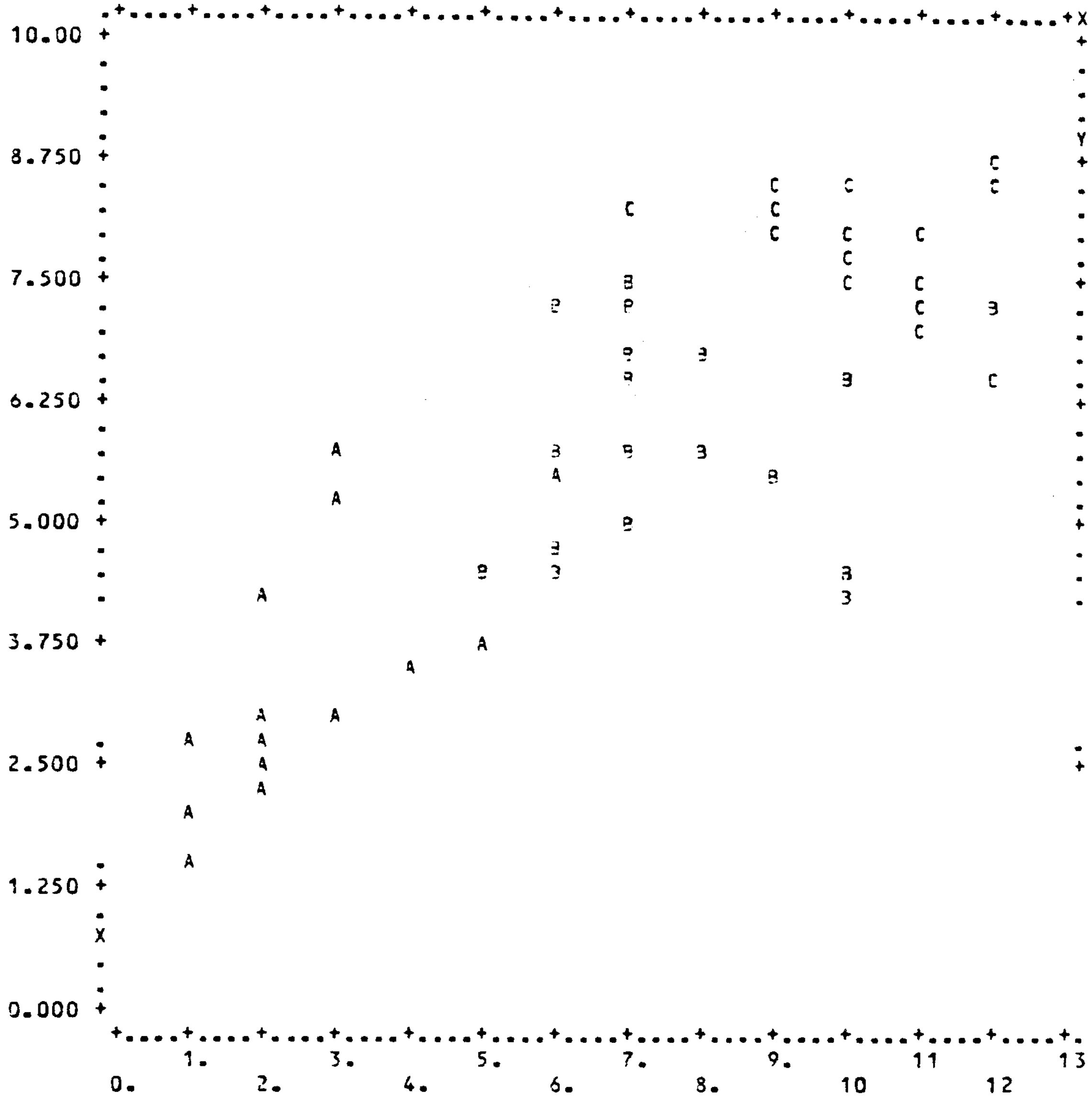


N= 60
COR=-.5663 03Q4AV

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	18.275	1.4385	X=-.36466*Y+ 20.344	1.4300
Y	5.6729	2.2338	Y=-.87935*X+ 21.743	3.4483

VARIABLE 51 03Q4AV VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 51 03Q4AV VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 51 05Q4AV VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

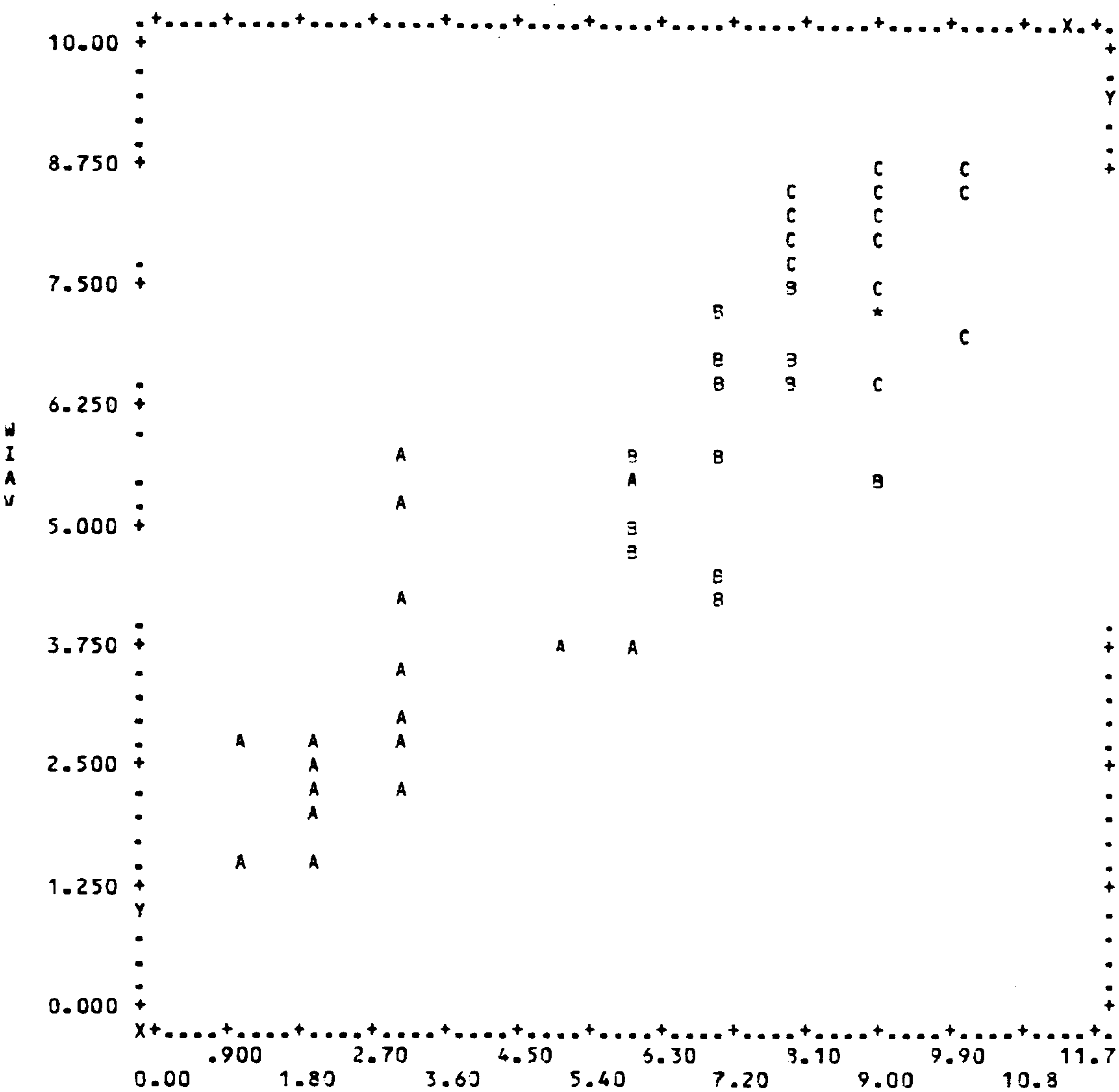
WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



N= 60
COR= .8662

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	6.7833	3.6130	X= 1.4011*Y-1.1648	3.3149
Y	5.6729	2.2338	Y= .53557*X+ 2.0400	1.2671

VARIABLE 52 PRQ1 VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 52 PRQ1 VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 52 PRQ1 VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

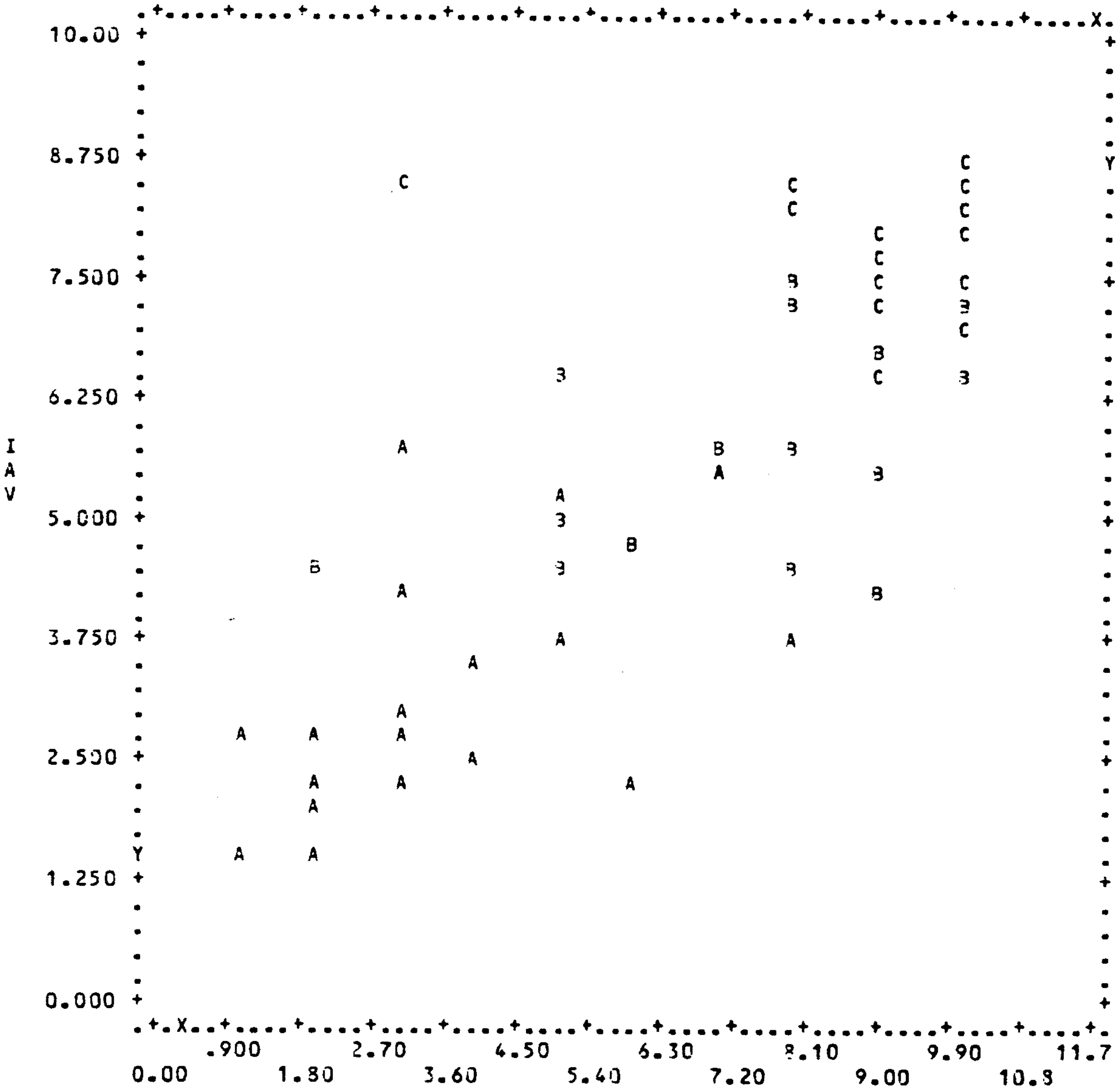


N= 60
COR= .8846

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	6.3667	2.7677	X= 1.0960*Y+ .14940	1.6951
Y	5.6729	2.2338	Y= .71395*X+ 1.1275	1.1043

VARIABLE 53 PRQ2 VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 53 PRQ2 VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 53 PRQ2 VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 42 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

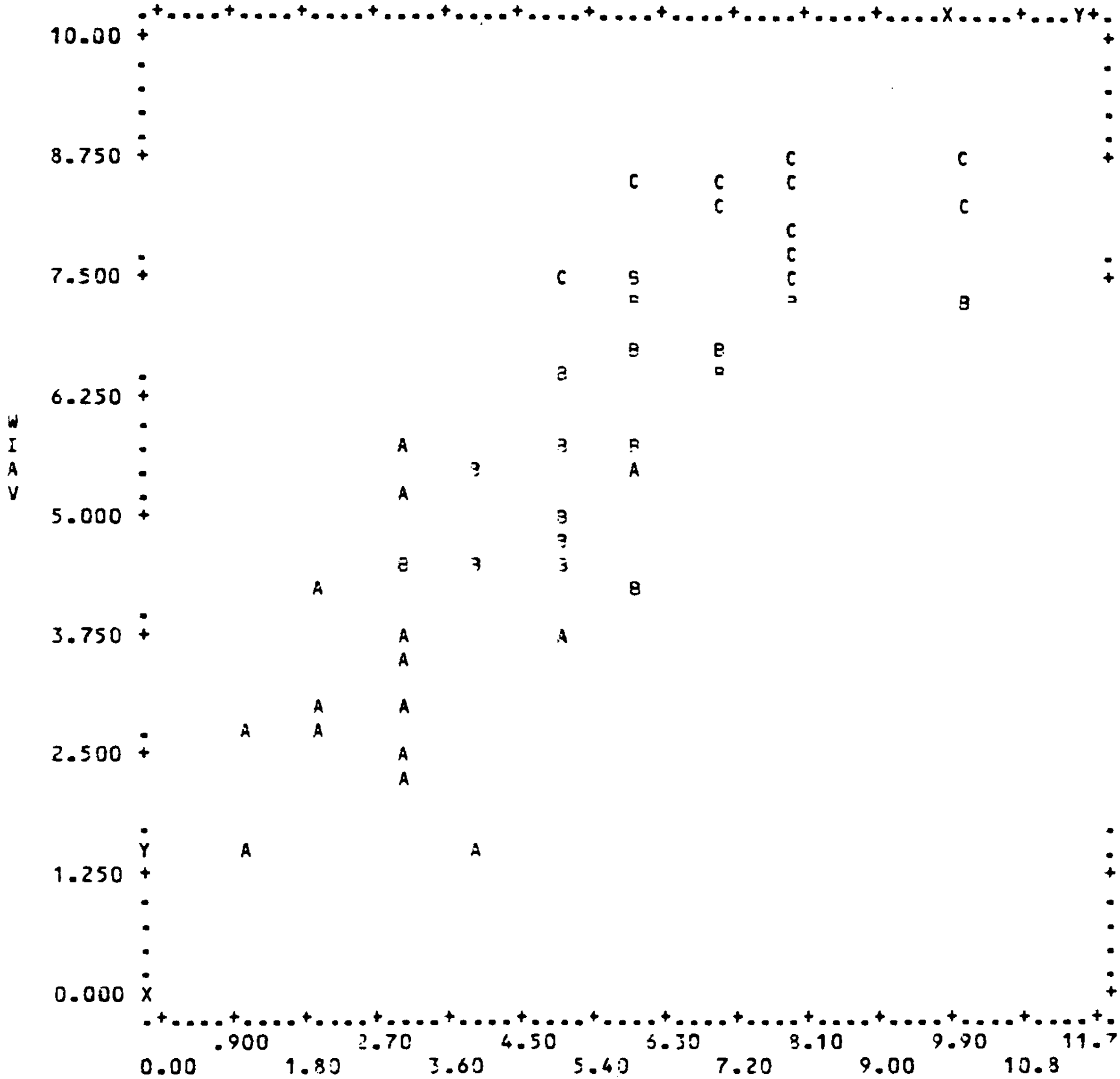


N= 60
COR= .8126

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	6.6833	2.9716	X= 1.0810*Y+ .55090	3.0509
Y	5.6729	2.2338	Y= .61037*X+ 1.5903	1.7241

VARIABLE 54 PRQ3 VERSU ARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 54 PRQ3 VERSU ARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 54 PRQ3 VERSU ARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 43 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D

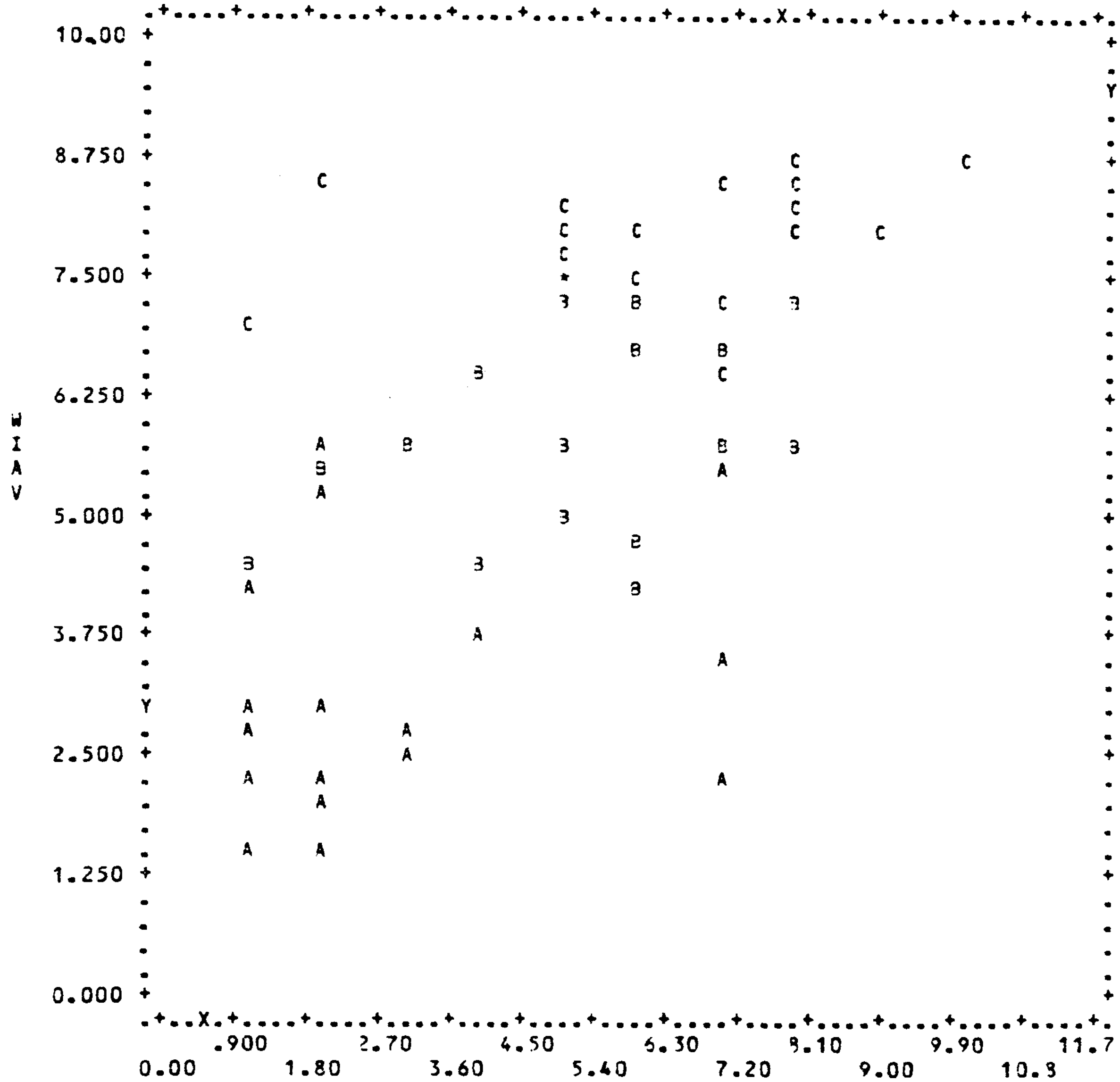


N= 60
COR= .8665

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	5.4333	2.5403	X= .98536*Y-.15655	1.6359
Y	5.6729	2.2338	Y= .76194*X+ 1.5330	1.2650

VARIABLE 55 PRQ4 VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 55 PRQ4 VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 55 PRQ4 VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

PAGE 44 WORKLOAD PROBE - 60 MIN. CUM. DATA. BMDP6D



N= 60
COR= .6101

	MEAN	ST.DEV.	REGRESSION LINE	RES.MS.
X	4.6000	2.5458	X= .69535*Y+ .65531	4.1388
Y	5.6729	2.2338	Y= .53535*X+ 3.2103	3.1864

VARIABLE 56 PRQ5 VERSUS VARIABLE 39 WIAV FOR GROUP A.SAMPLE SYMBOL=A
VARIABLE 56 PRQ5 VERSUS VARIABLE 39 WIAV FOR GROUP B.SAMPLE SYMBOL=B
VARIABLE 56 PRQ5 VERSUS VARIABLE 39 WIAV FOR GROUP C.SAMPLE SYMBOL=C

APPENDIX C

QUESTIONNAIRES/FORMS

PARTICIPANT CODE: ____________ DATE:

SIMULATION CODE: ____________

AIR TRAFFIC CONTROLLER SPECIALIST WORKLOAD QUESTIONNAIRE

INSTRUCTIONS

PLEASE COMPLETE THE FOLLOWING QUESTIONS AS SOON AS YOU HAVE BEEN RELIEVED FROM YOU RADAR POSITION. YOUR RESPONSES SHOULD CONCERN ONLY THE WORK WHICH YOU HAVE JUST COMPLETED.

ALL CONTROLLERS, NO MATTER HOW SKILLED AND EXPERIENCED, HAVE AT ONE TIME OR ANOTHER, BEEN THROUGH ALL THE LEVELS OF WORK-LOAD AND ACTIVITY EXPRESSED IN THIS QUESTIONNAIRE. IT DOES NOT DETRACT FROM YOUR PROFESSIONALISM IF FOR ANY PERIOD YOU ANSWER YOU WERE WORKING VERY HARD, OR IF YOU WERE HARDLY WORKING. FEEL FREE TO USE THE ENTIRE NUMERICAL SCALE FOR EACH QUESTION. BE AS HONEST AND AS ACCURATE AS YOU CAN. YOUR NAME IS NOT RECORDED ON THIS FORM, AND NO ATTEMPT WILL BE MADE TO ASSOCIATE YOUR RESPONSES WITH YOU AS AN INDIVI-DUAL. DATA COLLECTED WILL BE USED FOR RESEARCH PURPOSES ONLY.

ATCS WORKLOAD QUESTIONNAIRE

PARTICIPANT CODE: ____________ DATE: ____________

SIMULATION CODE: ____________

1. CHOOSE THE ONE NUMBER BELOW WHICH BEST DESCRIBES HOW HARD YOU WERE WORKING DURING THIS PERIOD:

DESCRIPTION OF WORKLOAD CATAGORY	RATING (CIRCLE ONE) ○
VERY LOW WORKLOAD - ALL TASKS WERE ACCOMPLISHED EASILY AND QUICKLY	1 2 3
MODERATE WORKLOAD - THE CHANCES FOR ERROR OR OMISSION WERE LOW	4 5 6
RELATIVELY HIGH WORKLOAD - THE CHANCES FOR SOME ERROR OR OMISSION WERE RELATIVELY HIGH	7 8 9
VERY HIGH WORKLOAD - IT WAS NOT POSSIBLE TO ACCOMPLISH ALL TASKS PROPERLY	10 11 12

2. WHAT FRACTION OF THE TIME WERE YOU BUSY DURING THE PERIOD YOU WERE CONTROLLING?

(CIRCLE ONE) ○

SELDOM HAD MUCH TO DO 1 2 3 4 5 6 7 8 9 10 FULLY OCCUPIED AT ALL TIMES

3. HOW MUCH DID YOU HAVE TO THINK DURING THIS PERIOD?

(CIRCLE ONE) ○

THERE WAS MINIMAL THINKING & PLANNING & CONCENTRATION REQUIRED 1 2 3 4 5 6 7 8 9 10 A GREAT DEAL OF THINKING, PLANNING AND CONCENTRATION WAS NECESSARY

4. DID YOU FIND THIS CONTROL PERIOD STRESSFUL?

(CIRCLE ONE) O

THE EXPERIENCE WAS RELAXING	1	2	3	4	5	6	7	8	9	10	THE EXPERIENCE WAS STRESSFUL

---

5. I AM FEELING TIRED

(CIRCLE ONE) O

STRONGLY DISAGREE	1	2	3	4	5	6	7	8	9	10	STRONGLY AGREE

---

6. COMMENTS:

ATCS WORKLOAD QUESTIONNAIRE

Regarding these runs, please rate the workload contributed by various factors.

WORKLOAD FACTORS	Workload Contribution was (Circle One) Very Low									Very High
1. Number of airplanes handled	1	2	3	4	5	6	7	8	9	10
2. Number of conflicts	1	2	3	4	5	6	7	8	9	10
3. Number of vectors given	1	2	3	4	5	6	7	8	9	10
4. Number of altitude changes	1	2	3	4	5	6	7	8	9	10
5. Number of airspeed reductions	1	2	3	4	5	6	7	8	9	10
6. Using strips without D-man	1	2	3	4	5	6	7	8	9	10
7. Emergencies	1	2	3	4	5	6	7	8	9	10
8. Weather	1	2	3	4	5	6	7	8	9	10
9. Pilot verbal response errors/delays	1	2	3	4	5	6	7	8	9	10
10. Pilot route/alt. errors	1	2	3	4	5	6	7	8	9	10
11. Accepting handoffs	1	2	3	4	5	6	7	8	9	10
12. Giving handoffs	1	2	3	4	5	6	7	8	9	10
13. Housekeeping (moving data blocks, removing strips)	1	2	3	4	5	6	7	8	9	10
14. Using trackball	1	2	3	4	5	6	7	8	9	10
15. Using keypack	1	2	3	4	5	6	7	8	9	10
16. Unfamiliarity with airplanes	1	2	3	4	5	6	7	8	9	10
17. Unfamiliar sector geometry	1	2	3	4	5	6	7	8	9	10
18. Area Restrictions	1	2	3	4	5	6	7	8	9	10
19. Simulation glitches (failures/anomalies)	1	2	3	4	5	6	7	8	9	10
20. Lack of foot pedal comm. switch	1	2	3	4	5	6	7	8	9	10
21. Aircraft flight characteristics (climb, descend, airspeed, turn)	1	2	3	4	5	6	7	8	9	10
22. Coordination with other sectors	1	2	3	4	5	6	7	8	9	10
23. Aircraft/Pilot procedural violations	1	2	3	4	5	6	7	8	9	10
24. Responding to the workload response box	1	2	3	4	5	6	7	8	9	10
25. Console Layout	1	2	3	4	5	6	7	8	9	10

ATCS WORKLOAD PROJECT

INFORMAL INTERVIEW

PARTICIPANT CODE

DATE ___________

RUN NO:

INTERVIEWER

ATCS WORKLOAD PROJECT

INTERVIEW PROTOCOL

(Administer After 3rd and 6th Data Run)

PARTICIPANT CODE DATE

(circle one)

YOU HAVE JUST COMPLETED THE (FIRST, SECOND) OF THREE CONTROL PERIODS IN THE EXPERIMENT.

A HOW REALISTIC WAS THE SIMULATION?

B) COULD YOU ASSIGN A NUMBER FROM 1 (LOW) TO 10 (HIGH) TO THIS REALISM? ______

C) IF THE REALISM WAS PERCEIVED AS LOW - PROBE TO IDENTIFY WHAT ELEMENTS OF THE SIMULATION WERE NOT REALISTIC. ______

2. DID YOU NOTICE ANY DIFFERENCES IN YOUR WORKLOAD ACROSS THE THREE CONTROL PERIODS? ______

A) IF NO, ⟶ STOP

B) IF YES, ⟶ CAN YOU RANK ORDER THE THREE CONTROL PERIODS FROM HIGHEST TO LOWEST WORKLOAD? CALL THEM A, B, C, IN THE ORDER YOU JUST RECEIVED THEM.

C IF YES, ⟶ HOW WERE THE THREE CONTROL PERIODS DIFFERENT? ____________________

3. HOW ADEQUATE DO YOU FEEL WAS THE TRAINING/FAMILIARIZATION YOU RECEIVED BEFORE WE STARTED COLLECTING DATA? ____________________

4 COULD YOU ASSIGN A NUMBER FROM 1 (VERY POOR) TO 10 (VERY GOOD) WHICH DESCRIBES HOW ADEQUATE THE TRAINING WAS? ________

5. WHAT DID YOU FIND WAS THE MOST DIFFICULT FOR YOU TO ACCOMPLISH DURING THE LAST THREE RUNS? (NOTE: If respondent has difficulty -- provide examples, i.e., planning, navigation, identifying conflicts, route changes, vectoring, coordination, etc., ⟶ use examples only if necessary.)

PROBE FOR EXPLANATION! ____________________

6. REFLECT BACK ON YOUR OWN EXPERIENCE BOTH AS AN ACTIVE CONTROLLER AND IN THIS SIMULATION. WE WOULD LIKE TO DRAW ON YOUR EXPERTISE! WHAT DO YOU BELIEVE INFLUENCES HOW HARD YOU HAVE TO WORK IN ORDER TO MAINTAIN YOUR PERFORMANCE? ______________________________

____

____

____

____

____

After the respondent has spoken for awhile -- probe to identify if he/she has a verbalizable internalized performance standard.

If this is after Block 2 -- Probe to identify changes in thinking as a function of three more hours in simulation.

7. EVERY CONTROLLER ESTABLISHES STRATEGIES OR COMMON WAYS OF DEALING WITH TRAFFIC.

WOULD YOU DESCRIBE YOUR APPROACH DURING HIGH WORKLOAD?

____

____

____

____

____

- DID YOU CHANGE YOUR REGULAR STRATEGIES IN ANY WAY IN ORDER TO CONTROL DURING THE PAST THREE PERIODS? ______________

____

____

____

8. IS THERE ANYTHING ELSE YOU THINK WE SHOULD KNOW THAT HAS NOT BEEN ALREADY COVERED? ______________________________

____

____

____

____

____

____

____